

**SVEUČILIŠTE U ZAGREBU
FAKULTET ORGANIZACIJE I INFORMATIKE
VARAŽDIN**

GORAN KLEPAC

**OTKRIVANJE ZAKONITOSTI TEMELJEM
JEDINSTVENOGA MODELA TRANSFORMACIJE VREMENSKE
SERIJE**

DOKTORSKA DISERTACIJA

VARAŽDIN, 2005.

© dr.sc. Goran Klepac
www.goranklepac.com
goran@goranklepac.com

Korištenje i citiranje bilo kojeg dijela teksta dopušteno je uz obavezno navođenje reference.

PODACI O DISERTACIJI

I. AUTOR

Ime i prezime :	Goran Klepac
Datum i mjesto rođenja :	10.06.1972, Zagreb
Naziv fakulteta i datum diplomiranja na VII/1 stupnju :	Ekonomski fakultet u Zagrebu, Sveučilište u Zagrebu, 16.06.1997.
Naziv fakulteta i datum diplomiranja na VII/2 stupnju :	Ekonomski fakultet u Zagrebu, Sveučilište u Zagrebu, 13.03.2000.
Sadašnje zaposlenje:	Raiffeisen Consulting, direktor BI direkcije

II. DISERTACIJA

Naslov:	Otkrivanje zakonitosti temeljem jedinstvenoga modela transformacije vremenske serije
Broj stranica, tablica, grafikona, slika:	211. stranica, 56 tablica, 62 slike, 86 referenci
Znanstveno područje i polje iz kojeg je postignut doktorat znanosti	Područje društvenih znanosti, polje informacijske znanosti
Fakultet na kome je obranjena disertacija	Fakultet organizacije i informatike, Varaždin

III. OCJENA I OBRANA

Datum prijave teme:	22. travnja 2003.
Datum sjednice na kojoj je prihvaćena tema:	22. listopada 2003.
Datum predaje rada:	04. siječanj 2005..
Datum sjednice na kojoj je rad prihvaćen:	15. veljače 2005.
Sastav povjerenstva koje je rad ocjenilo:	Prof. dr.sc. Vesna Dušak Prof. dr.sc. Božidar Kliček Doc. dr.sc. Marijana Zekić - Sušac
Datum obrane rada:	18. ožujka 2005.
Sastav povjerenstva pred kojim je rad obranjen:	Prof. dr.sc. Vesna Dušak – predsjednik Prof. dr.sc. Božidar Kliček – član Doc. dr.sc. Marijana Zekić - Sušac – član
Datum promocije:	

SVEUČILIŠTE U ZAGREBU
FAKULTET ORGANIZACIJE I INFORMATIKE
VARAŽDIN

Voditelj rada: prof.dr.sc. Božidar Kliček

Sadržaj

1. Uvod	1
1.1. Svrha, cilj rada, ciljevi istraživanja	1
1.2. Hipoteze i obrazloženje hipoteza	6
1.3. Metode rada	8
1.4. Očekivani znanstveni doprinos	9
1.5. Vremenske serije i analiza podataka u oblasti istraživanja tržišta	10
1.6. Različiti pristupi problematici analize vremenskih serija u domeni istraživanja tržišta	14
1.7. Karakteristike i ciljevi REF II modela	19
1.8. Prednosti REF II modela	23
2. Temeljni koncept REF II modela	26
2.1. Matematička definicija modela	26
2.1.1. Promjena funkcijskih vrijednosti (prirast funkcije)	26
2.1.2. Površina ispod krivulje	30
2.1.3. Definicija elementarnih uzoraka	33
2.2. Definicija i opis REF metode	33
2.2.1. REF model i otkrivanje sezonskih oscilacija	36
2.2.2. Algoritamska interpretacija primjene REF modela	43
2.2.3. REF model – područje efikasnosti / neefikasnosti	43
2.3. Proračun površine ispod krivulje	44
2.3.1. Algoritamska interpretacija numeričke integracije	46
2.3.2. Područje efikasnosti/neefikasnosti proračuna površine ispod krivulje	47
2.4. Proračun koeficijenta kutnog otklona	48
2.4.1. Algoritamska interpretacija proračuna koeficijenata kutnog otklona	50
2.4.2. Prednosti i nedostaci proračuna koeficijenta kutnog otklona	51
2.5. Shematski prikaz REF II modela	52
2.6. Algoritamska interpretacija REFII modela	52
2.7. Kako analizirati vremenske serije posredstvom REFII modela	55
2.8. Trodimenzionalna interpretacija REFII modela	57
3. Pojemne strukture i njihova definicija na osnovu REFII modela	59
3.1. Pojam jednakosti vremenskih odsječaka	59
3.2. Pojam nejednakosti vremenskih odsječaka	61
3.3. Pojam inverznosti vremenskih odsječaka	62
3.4. Pojam korelacije vremenskih odsječaka	63
3.5. Pojam sličnosti vremenskih odsječaka	66
3.6. Pojemne strukture i fuzzy logika	67
4. Konceptualna primjena modela	70
4.1. REFII model i konstrukcija problemskih rješenja	70
4.2. Otkrivanje sezonskih oscilacija	72
4.2.1. Otkrivanja sezonskih oscilacija na temelju REFII modela	72
4.2.2. Algoritam rješenja	75
4.2.3. Otkrivanje sezonskih oscilacija na empirijskim podacima	77
4.3. Otkrivanje cikličkih oscilacija	82
4.4. Izravno otkrivanje pravila iz vremenske serije	84
4.4.1. Izravno otkrivanje pravila iz vremenskih serija posredstvom REFII modela	84
4.4.2. Model rješenja izravnog otkrivanja pravila iz vremenskih serija	85
4.4.3. Izravno otkrivanje pravila na empirijskim podacima	87
4.5. Otkrivanje epizoda u vremenskim serijama	88
4.6. Teorija “spavača”	93
4.7. Grupiranje vremenskih odsječaka	94

4.7.1. Pojam grupiranja vremenskih odsječaka i REFII model	94
4.7.2. Model grupiranja vremenskih odsječaka.....	95
4.7.3. Rezultati otkrivanja sličnih grupa na empirijskim podacima	100
4.8. Formalna logika, ekspertni sustavi i REFII model	102
5. Metodologija primjene REFII modela	107
5.1. Važnost temeljnog pretprocesiranja izvornih podataka iz vremenskih serija ...	107
5.1.1. Primarna transformacija vremenske serije u REFII model	107
5.1.2. Sekundarna transformacija REFII modela	109
5.1.3. Građenje stabla vremenskih indeksa	110
5.2. Integracija metoda data mininga u vremenske serije	113
5.3. Metodologija integracije različitih koncepcija analize	116
5.4. Konstrukcija nestandardnih algoritamskih rješenja unutar REFII modela.....	119
5.5. Određivanje stupnja pouzdanosti analize	120
5.6. Provođenje analiza	122
5.7. Prikaz i primjena rezultata	125
6. Otkrivanje tržišnih zakonitosti pomoću REF II modela na realnim podacima	127
6.1. Otkrivanje uzoraka iz vremenske serije	127
6.1.1. Izvor i struktura podataka za analizu	127
6.1.2. Vremenska interpolacija, i vremenska granulacija.....	128
6.1.3. Transformacija u REFII notaciju.....	128
6.1.4. Rješenje modela otkrivanja uzoraka	130
6.1.5. Rezultati istraživanja	134
6.2. Otkrivanje mutacija unutar uzoraka	135
6.3. Otkrivanje sezonskih oscilacija	137
6.4. Otkrivanje događajnosti	139
6.5. Otkrivanje motiva	141
6.7. Integracija metoda data mininga sa vremenskim serijama	143
6.7.1. Postupak ekspanzije vremenskih odsječaka	144
6.7.2. Tablična i OLAP analiza.....	146
6.7.3. Primjena stabla odlučivanja na vremensku seriju.....	149
6.7.4. Primjena klasteriranja na vremenskim serijama	151
6.8. Konsolidacija rezultata kroz prizmu tržišnih analiza	153
6.9. Analiza tržišta tekstila pomoću REFII modela	155
7. Ostala područja primjene REFII modela (idejni modeli i konceptualna rješenja) ...	156
7.1. Tekst mining.....	157
7.2. Web mining	160
7.3. Medicina.....	163
7.4. "Profile" analize	167
7.5. Trodimenzionalni prostori i REFII model	169
8. Zaključak.....	170
Sažetak.....	174
Životopis	176
Referentna literatura.....	179
Prilozi	193
Fuzzy definicija jednakosti odsječaka.....	193
Uvodna maska "Time explorera-a"	201
Program za vremensku interpolaciju i vremensku granulaciju	201
Transformacija u REFII model	203
Pomoćni objekti transformacije	205
Modul za otkrivanje uzoraka u vremenskim serijama.....	206
Objekt za otkrivanje sezonskih oscilacija.....	209
Kod za otkrivanje događajnosti	210

Program za traženje motiva	211
----------------------------------	-----

Popis tablica

TABLICA 1.1. : FREKVENCIJA POJAVNOSTI	17
TABLICA 1.2. ORIJENTACIJSKI ODNOSI IZMEĐU TIPRA ANALIZE I ELEMENATA REFII MODELA	22
TABLICA 1.3. TRANSFORMACIJA VREMENSKE SERIJE U OBJEKTE.....	22
TABLICA 2.1. DEFINICIJA REF OZNAKA NA TEMELJU PRIRASTA DISKONTINUIRANE FUNKCIJE	27
TABLICA 2.2. PRIMJER NADOPUNE REF MODELA SA NA TEMELJU KOEFICIJENATA KUTNOG OTKLONA	28
TABLICA 2.3. DEFINICIJA TOČAKA FUNKCIJE PRIPADNOSTI REF LINGVISTIČKE VARIJABLE	28
TABLICA 2.4. DEFINICIJA TOČAKA FUNKCIJE PRIPADNOSTI LINGVISTIČKE VARIJABLE "KUTNI OTKLON"	29
TABLICA 2.5. PRIMJER NADOPUNE REF MODELA SA POVRŠINOM ISPOD KRIVULJE.....	31
TABLICA 2.6. DEFINICIJA TOČAKA FUNKCIJE PRIPADNOSTI LINGVISTIČKE VARIJABLE "POVRŠINA"	32
TABLICA 2.7. PRAVILA KLASIFIKACIJE UZORAKA.....	32
TABLICA 2.8. INDEKSIRANE VARIJABLE.....	39
TABLICA 2.9. VREMENSKE SERIJE ZA REF MODEL.....	39
TABLICA 2.10. VREMENSKE SERIJE U REF NOTACIJI	40
TABLICA 2.11. PRORAČUN GREŠKE MODELA	40
TABLICA 2.12. SEZONSKE OSCILACIJE.....	40
TABLICA 2.13. INTERVALNE VRIJEDNOSTI OTKLONA KUTOVA	50
TABLICA 2.14. IZVORNA VREMENSKA SERIJA	54
TABLICA 2.15. NORMIRANE VRIJEDNOSTI VREMENSKE SERIJE	55
TABLICA 2.16. VREMENSKA SERIJA U REFII NOTACIJI.....	55
TABLICA 2.17. MATRICA TRANSFORMACIJE.....	56
TABLICA 3.1. SUPSTITUCIJA ELEMENATA.....	63
TABLICA 3.2. TABLICA KORELATORNIH ODNOSA	64
TABLICA 3.3. TUMAČENJE KOEFICIJENTA KORELACIJE [HAN, 2001].....	65
TABLICA 4.1. TABLICA VREMENSKIH KONTINUITETA SA PRIPADAJUĆIM KOEFICIJENTIMA POUZDANOSTI	74
TABLICA 4.2. TABLICA PRETPROCESIRANIH VREMENSKIH SERIJA U REFII NOTACIJI ZA PROCJENU SEZONSKIH OSCILACIJA	76
TABLICA 4.3. STRUKTURA SOM REŠETKE	76
TABLICA 4.4. "CRISP " RAZREDI OTKLONA KUTOVA	78
TABLICA 4.5. SEZONSKE OSCILACIJE.....	78
TABLICA 4.6. TABLICA SEZONSKIH UZORAKA.....	79
TABLICA 4.7. TABLICA SEZONSKIH UZORAKA S OBZIROM NA CIKLUSE	80
TABLICA 4.8. TABLICA SEZONSKIH UZORAKA S OBZIROM NA TRENDOVE	81
TABLICA 4.9. TABLICA KLASIFIKACIJE OTKLONA KUTOVA	86
TABLICA 4.10. TABLICA TRANSFORMACIJE U REFII MODEL ZA POTREBE IZRAVNOG OTKRIVANJA PRAVILA	86
TABLICA 4.11. DEFINICIJA LEKSIČKE OZNAKE.....	89
TABLICA 4.12. TABLICA SEKUNDARNE TRANSFORMACIJE	98
TABLICA 4.13. PROCJENA SLIČNOSTI NA TEMELJU DVIJE VREMENSKE SERIJE.....	98
TABLICA 4.14. PROCJENA SLIČNOSTI NA TEMELJU DVIJE VREMENSKE SERIJE.....	99
TABLICA 4.15. TABLICA PRIMARNE TRANSFORMACIJE OBLIKA KRIVULJE DIJELA PODATAKA ZA ANALIZU	100
TABLICA 4.16. TABLICA SEKUNDARNE TRANSFORMACIJE OBLIKA KRIVULJE DIJELA PODATAKA ZA ANALIZU	101
TABLICA 4.17. TABLICA KATEGORIJA SLIČNOSTI.....	101
TABLICA 4.18. BLOKOVI PRAVILA	105
TABLICA 5.1. METODE RUDARENJA PODATAKA S OBZIROM NA MOGUĆNOST PRIMJENE REFII MODELA	114

<i>TABLICA 6.1. STRUKTURA TABLICE ŠTETNIKA.....</i>	<i>127</i>
<i>TABLICA 6.2. RAZREDI KUTNIH OTKLONA.....</i>	<i>129</i>
<i>TABLICA 6.3. ODSJEČAK REZULTATA OBRADE</i>	<i>129</i>
<i>TABLICA 6.4. SEKUNDARNA TRANSFORMACIJA RAZREDA.....</i>	<i>130</i>
<i>TABLICA 6.5. SEZONSKE OSCILACIJE NA TEMELJU TRENDOVA</i>	<i>137</i>
<i>TABLICA 6.6. SEZONSKE OSCILACIJE NA TEMELJU RAZREDA OTKLONA KUTOVA.....</i>	<i>138</i>
<i>TABLICA 6.7. KLASIFIKACIJA NORMIRANE POVRŠINE.....</i>	<i>142</i>
<i>TABLICA 6.8. OTKLONI KUTOVA PREMA SPOLU</i>	<i>146</i>
<i>TABLICA 6.9. OTKLONI KUTOVA PREMA MJESTU REGISTRACIJE.....</i>	<i>147</i>
<i>TABLICA 6.10. OTKLONI KUTOVA PREMA MJESTU REGISTRACIJE(STRUKTURA).....</i>	<i>149</i>
<i>TABLICA 6.11. BROJ SLUČAJEVA U KLASITERIMA</i>	<i>152</i>
<i>TABLICA 6.12. VRIJEDNOSTI CENTROIDA</i>	<i>152</i>
<i>TABLICA 7.1. HIPOTETSKA TABLICA ZA PROCJENU SLIČNOSTI TEKSTOVA</i>	<i>159</i>
<i>TABLICA 7.2. SUSTAV NUMERIČKE TRANSFORMACIJE STRANICA</i>	<i>161</i>

Popis slika

<i>SLIKA 1.1. PRETPROCESIRANJE VREMENSKE SERIJE.....</i>	<i>11</i>
<i>SLIKA 1.2 METODOLOGIJA PROCESUIRANJA VREMENSKE SERIJE DATA MINING ALGORITMIMA POSREDSTVOM REF II ALGORITMA.....</i>	<i>13</i>
<i>SLIKA 1.3. PREDVIĐANJE TRENDOVA PRIMJENOM NEURALNE MREŽE.....</i>	<i>15</i>
<i>SLIKA 1.4. EPIZODE U VREMENSKOJ SERIJI</i>	<i>16</i>
<i>SLIKA 1.5. REPREZENTACIJA VREMENSKE SERIJE POMOĆU VEKTORA.....</i>	<i>17</i>
<i>SLIKA 1.6. KOMPONENTE REFII MODELA</i>	<i>20</i>
<i>SLIKA 1.7. ULOGA KOEFICIJENTA KUTNOG NAGIBA PRAVCA U REFII MODELU.....</i>	<i>21</i>
<i>SLIKA 1.8. ETAPE U OTKRIVANJU ZNANJA POSREDSTVOM REFII MODELA</i>	<i>24</i>
<i>SLIKA 2.1. PRIRAST KONTINUIRANE FUNKCIJE.....</i>	<i>26</i>
<i>SLIKA 2.2. PRIRAST DISKONTINUIRANE FUNKCIJE.....</i>	<i>27</i>
<i>SLIKA 2.3. FUNKCIJA PRIPADNOSTI REF VARIJABLI</i>	<i>28</i>
<i>SLIKA 2.4. FUNKCIJA PRIPADNOSTI VARIJABLI KUTNIH OTKLONA</i>	<i>29</i>
<i>SLIKA 2.5. NUMERIČKA INTEGRACIJA</i>	<i>30</i>
<i>SLIKA 2.6. FUNKCIJA PRIPADNOSTI NORMIRANE VARIJABLE POVRŠINA</i>	<i>32</i>
<i>SLIKA 2.7. FUZZY SUSTAV KLASIFIKACIJE</i>	<i>32</i>
<i>SLIKA 2.8. PRIKAZ VREMENSKE SERIJE KROZ REF MODEL.....</i>	<i>35</i>
<i>SLIKA 2.9. VREMENSKA SERIJA PRODAJE ARTIKLA Z PO DANIMA.....</i>	<i>36</i>
<i>SLIKA 2.10. LOMLJENJE VREMENSKIH SERIJA U MANJE VREMENSKE JEDINICE</i>	<i>37</i>
<i>SLIKA 2.11. UZROCI VELIKE STANDARDNE GREŠKE USPRKOS PRIHVAĆANJU MODELA.....</i>	<i>41</i>
<i>SLIKA 2.12. GRAFIČKA INTERPRETACIJA NUMERIČKE INTEGRACIJE U REFII MODELU.....</i>	<i>47</i>
<i>SLIKA 2.13. MODEL PRORAČUNA NAGIBA PRAVCA I DULJINE KRAKA U VREMENSKOJ SERIJI.....</i>	<i>49</i>
<i>SLIKA 2.14. SHEMA REF II MODELA</i>	<i>52</i>
<i>SLIKA 2.15. REFII I 3D PROSTOR.....</i>	<i>58</i>
<i>SLIKA 3.1. RAČUNANJE KORELATORNIH ODNOSA U VREMENSKIM SERIJAMA.....</i>	<i>64</i>
<i>SLIKA 3.2. FUZZY SUSTAV ZA PROCJENU JEDNAKOSTI ODSJEČAKA.....</i>	<i>67</i>
<i>SLIKA 3.3. DEFINICIJA FUZZY VARIJABLI U SUSTAVU ZA PROCJENU JEDNAKOSTI ODSJEČAKA.....</i>	<i>69</i>
<i>SLIKA 4.1. TEMELJNA ULOGU REFII MODELA U DATA MINING ANALIZAMA VREMENSKIH SERIJA</i>	<i>71</i>
<i>SLIKA 4.2. OTKRIVANJE SEZONSKIH OSCILACIJA NA TEMELJU REFII MODELA.....</i>	<i>73</i>
<i>SLIKA 4.3. CIKLIČKE OSCILACIJE.....</i>	<i>82</i>
<i>SLIKA 4.4. MODEL UVJETNIH VJEROJATNOSTI PROCJENE DOGAĐAJA.....</i>	<i>83</i>
<i>SLIKA 4.5. IDEJNI MODEL RJEŠNJA IZRAVNOG OTKRIVNJA PRAVILA.....</i>	<i>85</i>
<i>SLIKA 4.6. TRANSFORMACIJA ELEMENATA REFII MODELA U DOGAĐAJE</i>	<i>90</i>
<i>SLIKA 4.7. FORMIRANJE EPIZODA U VREMENSKOJ SERIJI</i>	<i>90</i>
<i>SLIKA 4.8. KORELATORNE ANALIZE DOGAĐAJA SA ZANEMARIVANJEM MJERILA VREMENSKE KOMPLEKSNOSTI.....</i>	<i>92</i>
<i>SLIKA 4.9. GRAFIČKA INTERPRETACIJA TEORIJE SPAVAČA PREKO VREMENSKE SERIJE</i>	<i>93</i>
<i>SLIKA 4.10. OTKRIVANJE SLIČNOSTI</i>	<i>97</i>
<i>SLIKA 4.11. LEKSIČKE VARIJABLE I BLOKOV I PRAVILA</i>	<i>105</i>
<i>SLIKA 5.1. SPIRALNI PRISTUP ANALIZE I REFII MODEL.....</i>	<i>108</i>
<i>SLIKA 5.2. METODOLOGIJA PROVOĐENJA SEKUNDARNE IZ PRIMARNE TRANSFORMACIJE.....</i>	<i>110</i>
<i>SLIKA 5.3. PRIMJER METODOLOGIJE IZGRADNJE VREMENSKIH INDEKSA.....</i>	<i>111</i>
<i>SLIKA 5.4. KREIRANJE RESTRUKTURIRANIH INDEKSA</i>	<i>112</i>
<i>SLIKA 5.5. EKSPANZIJA VREMENSKOG ODSJEČKA NETEMPORALNIM ATRIBUTIMA</i>	<i>115</i>
<i>SLIKA 5.6. RJEŠENJE KOMPLEKSNE ANALIZE VREMENSKIH SERIJA TEMELJENE NA REFII MODELU ..</i>	<i>117</i>
<i>SLIKA 5.7. DEFINICIJA SEZONSKIH OSCILACIJA PREKO SLIČNOSTI</i>	<i>118</i>
<i>SLIKA 5.8. SPIRALNI PRISTUP U ODREĐIVANJU ŽELJENOG STUPNJA POUZDANOSTI.....</i>	<i>121</i>

<i>SLIKA 5.9. METODOLOGIJA SUSTAVNOG PRISTUPA ANALIZE VREMENSKIH SERIJA POMOĆU REFII</i>	
<i>MODELA</i>	<i>124</i>
<i>SLIKA 5.10. PRIJEDLOG OBJEDINJAVANJA ZNANJA U FUZZY EKSPERTNOM SUSTAVU</i>	<i>126</i>
<i>SLIKA 6.1. MODEL PRETRAGE UZORAKA U VREMENSKOJ SERIJI</i>	<i>131</i>
<i>SLIKA 6.2. ALGORITAM PRETRAGE UZORKA U VREMENSKOJ SERIJI</i>	<i>132</i>
<i>SLIKA 6.3. OTKRIVENI UZORCI SA KOEFICIJENOM POUZDANOSTI 0.2.....</i>	<i>134</i>
<i>SLIKA 6.4. OTKRIVENI UZORCI SA KOEFICIJENOM POUZDANOSTI 0.1</i>	<i>135</i>
<i>SLIKA 6.5. OTKRIVANJE MUTACIJE UNUTAR UZORAKA PRI KOEFICIJENU POUZDANOSTI 0.1</i>	<i>135</i>
<i>SLIKA 6.6. ALGORITAM ZA OTKRIVANJE MUTACIJE UNUTAR UZORAKA.....</i>	<i>136</i>
<i>SLIKA 6.7. ALGORITAM ZA OTKRIVANJE DOGAĐAJNOSTI.....</i>	<i>140</i>
<i>SLIKA 6.8. OTKRIVANJE DOGAĐAJNOSTI PRI KOEFICIJENU POUZDANOSTI 0.1</i>	<i>141</i>
<i>SLIKA 6.9. STABLO ODLUČIVANJA GRUPE DOBI- TRENDVI VREMENSKIH OTKLONA.....</i>	<i>150</i>
<i>SLIKA 6.10. STABLO ODLUČIVANJA SPOL-TJEDNI TRENDVI.....</i>	<i>151</i>
<i>SLIKA 6.11. KLASITERI KONCENTRACIJE KUTNOG OTKLONA.....</i>	<i>153</i>
<i>SLIKA 7.1. INTEGRACIJA REFII MODELA I PROBABLISTIČKOG KONCEPTA U WEB MININGU</i>	<i>162</i>
<i>SLIKA 7.2. SHEMA SUSTAVA ZA PRIKUPLJANJE PODATAKA SA MEDICINSKIH UREĐAJA ZA POTREBE</i>	
<i>ANALIZA</i>	<i>164</i>
<i>SLIKA 7.3. INTEGRACIJA REFII MODELA I BAYESOVII MREŽA U SVRHU ISTRAŽIVANJA EFIKASNOSTI</i>	
<i>LJEKOVA.....</i>	<i>166</i>

POPIS UPOTRIJEBLJENIH KRATICA

BN	Bayesian Network (Bayesove mreže)
DM	Data Mining (rudarenje podataka)
ES	Expert system (ekspertni sustav)
REF	Akronim od Rise, Equal, Fall. Model nagiba pravaca vremenskih odsječaka, dio REFII modela, ujedno i preteča REFII modela.
REFII	Akronim od Rise, Equal, Fall, gdje II označavava drugu, dorađenu generaciju REF modela. REFII model je jedinstveni model transformacije vremenske serije.
SOM	Self Organizing Maps (samoorganizirajuće mape)

Predgovor

S obzirom na karakter ulaznih podataka, te razvoj metoda rudarenja podataka (eng. Data mining) koje su se povijesno razvijale za skupove podataka koji nisu strukturirani u vremenske serije, problematika kompatibilnosti metoda i metodologija njihove lančane primjene tijekom analize, evolucijski se je razvijala sa samim metodama rudarenja podataka.

Sa druge strane potencijali vremenskih serija bili su dobro poznati i u periodu nepoznavanja termina rudarenja podataka i nepostojanja računala, te su one bile uglavnom predmet izučavanja matematičara.

Kao rezultat tog izučavanja pojavila se čitava paleta metoda i postupaka analize vremenskih serija, da bi ona svoj procvat doživjela u novije vrijeme u eri informatizacije.

Kao što su i netemporalni podaci u vidu metoda rudarenja podataka doživjeli svojevrsnu analitičku renesansu, tako su se i vremenske serije priključile tom trendu sa jednom značajnom razlikom.

Razlika se očitovala, i još uvijek se očituje u činjenici da za nevremenske attribute postoje standardne metode i postupci pretprocesiranja podataka prilagođeni određenim vrstama metoda i analitičkih postupka te je između njih vrlo lako postići kompatibilnost, dok kod vremenskih serija postoji šarenilo u shvaćanju i provedbi kako pretprocesiranja podataka, tako i same analize što rezultira nekompatibilnošću metoda tijekom provedbi analiza.

Djelomični razlog tome proizlazi iz činjenice što je rudarenje vremenskih serija relativno mlado područje, te postoji vrlo živa aktivnost u konstrukciji novih algoritamskih rješenja koja su prvenstveno fokusirana na rješavanje zadanih problema, te se ne vodi dovoljno računa o postizanju kompatibilnosti sa već otkrivenim analitičkim metodama.

Glavni razlog ipak proizlazi iz činjenice, što su vremenske serije specifične strukture podataka, koje osim svog temporalnog karaktera imaju i niz drugih karakteristika koje su nepoznanica za netemporalne strukture podataka, kao što je primjerice redoslijedna zavisnost .

Kao posljedica svih ovih faktora, danas postoji čitav niz nepovezanih, međusobno nekompatibilnih metoda za rudarenje vremenskih serija koje su prvenstveno fokusirane na rješavanje parcijalnih analitičkih problema.

Prilikom kompleksnih analiza koje zahtijevaju ulančavanje metoda rudarenje vremenskih serija postoje problemi u provedbi takvih analiza zbog nekompatibilnosti među metodama.

Cilj rada je ponuditi rješenje ovog problema u vidu jedinstvenog modela transformacije vremenske serije. To je bitno drugačiji pristup spomenutoj problematici nego što je do sada bio slučaj, jer u prvi plan stavlja model transformacije vremenske serije kao polazišnu osnovu daljnjih analize, za razliku od tradicionalnog pristupa u konstrukciji algoritama za rudarenje vremenskih serija.

Na taj način izgrađeni su temelji za razvoj čitave palete različitih, kompatibilnih, međusobno povezanih analitičkih postupaka, kao i primjene tradicionalnih metoda rudarenja podataka na vremenskoj seriji. Rad također nudi i autorska rješenja proizašla iz jedinstvenog modela transformacije koja se odnose na standardnu problematiku rudarenja vremenskih serija, a odnosi se na otkrivanje uzoraka,

otkrivanje događaja, otkrivanje mutacija, procjene sličnosti, klasifikaciju vremenskih uzoraka, izravno otkrivanje pravila iz vremenske serije i dr.

Ovakav pristup analizi omogućava provođenje kompleksnih analiza u okviru istog sustava, te ostavlja odškrinuta vrata za konstrukciju novih algoritamskih postupaka za otkrivanje znanja u vremenskim serijama.

Posebno bih se zahvalio mentoru prof. dr. sc. Božidaru Kličeku na njegovom stručnom vodstvu, korisnim savjetima i podršci tijekom izrade rada, kćeri Lauri i supruzi Antoniji na iskazanom strpljenu i podršci tijekom izrade disertacije.

1. Uvod

1.1. Svrha, cilj rada, ciljevi istraživanja

Određeni tipovi sustava su vremenski promjenjivi, pa spadaju u posebnu kategoriju sustava, koje promatramo i analiziramo u vidu vremenskih serija.

Vremenske serije su oduvijek bile vrlo interesantno područje primjeni iz perspektive rudarenja podataka (eng. data mining). Promatranje vrijednosti nekog obilježja u jedinici vremena može biti vrlo korisno za spoznavanje promatrane pojave. Ovom spoznajom koristi se i klasična statistika, koja nudi metode kao što su skupni indeksi vremenskih serija, srednje vrijednosti vremenskih serija, individualni indeksi, regresijske metode, procjene trendova i slično [Šošić, 1990].

Ove metode mogu biti vrlo korisne kod procjene određenih pokazatelja vezanih uz promatranu problematiku, čak i procjene trendova pojava, no ako cjelokupnu problematiku promatramo kroz perspektivu inteligentne analize podatka dobivamo vrlo skučen instrumentarij za kvalitetnu analizu temporalnih obilježja .

Razvojem koncepcije rudarenja podataka uviđa se važnost vremenskih serija kao područja koje može pružiti relevantne informacije pa se analiza vremenskih serija velikim dijelom oslanja na koncepciju neuralnih mreža u predviđanjima vremenskih trendova [Apostols, 1996], [Taylor, 1996]. Primjena modela neuronskih mreža prvenstveno BPN algoritma najočitija je u predviđanju trendova vrijednosnica, financijskih pokazatelja, prihoda poduzeća, procjenu potražnje i slično [Kliček, 2002] .

Druga grana temeljena na analizama vremenskih serija koristi se metodologijom fuzzy logike, te njene resurse pokušava implementirati u vremenske nizove [Cheng, 1997]. Svaki od opisanih pristupa ima svoje karakteristike koje se aplikativno mogu primijeniti u praksi, ali osnovna karakteristika svakog od ovih pristupa jest nejednoznačnost opisa vremenske serije odnosno određenog segmenta vremenske serije. U posljednje vrijeme kada je riječ o promatranoj problematici pojavljuje se trend razvoja jezika baziranog na analizama vremenskih serija. Radovi koji pobuđuju pažnju i vrlo studiozno pristupanje otkrivanju znanja iz vremenskih serija su [Pratt, 2001], i [Manilla, 1997]. U prvom slučaju riječ je o *data mining* metodologiji pristupu otkrivanja znanja posredstvom algoritma koji u osnovi ima elemente funkcija udaljenosti i koji može ekstrahirati znanje o korespondentnim elementima iz vremenske serije. U drugom slučaju je riječ o definiciji pojma epizoda u vremenskim serijama kojima točke u vremenskom prostoru reprezentiraju događaje. Ovaj koncept bit će izložen u radu kroz prizmu opisanog modela.

Problematika vremenskih serija provlači se i kroz problematiku preprocesiranja podataka [Pyle, 2001]. Postoji niz metoda koje u vremenskim serijama nadomještaju nedostajuće vrijednosti atributa. Od najrasprostranjenijih metoda spominjem metodu linearne interpolacije, koja svoje korijene vuče iz područja matematičke analize, metodu srednje vrijednosti, metodu sličnosti koja distancama pokušava procijeniti vrijednost nedostajućih atributa, te metodu neuronskih mreža.

Tradicionalan pristup analizi vremenskih serija, pa tako i analizi vremenskih serija s ciljem otkrivanja tržišnih zakonitosti, oslanja se na korištenje niza nezavisnih metoda koje rješavaju određene tipove problema. U otkrivanju vremenskih zakonitosti tržišta

možemo koristiti čitav niz metoda za otkrivanje sezonskih oscilacija, cikličkih oscilacija, [Taylor 1996.], [Westhpal 2001.], predviđanja trendova, [Pyle 2001.], metode otkrivanja epizoda unutar vremenskih serija [Manilla 1997.] te metode otkrivanja uzoraka [Han 2000.], [Dougherty 1988.], [Taylor 1996.].

Autori koji obrađuju pretprocesiranje podataka, kao vrlo važnu komponentu rudarenja klasičnih transakcijskih podataka, ne posvećuju dovoljnu pažnju pretprocesiranju vremenske serije [Han, 2000.], [Pyle, 2001.] s ciljem stvaranja jedinstvenoga modela transformacije koji bi omogućio primjenu široke lepeze analitičkih metoda na vremensku seriju kroz jedinstveni sustav analize.

Spomenuti se autori fokusiraju na tradicionalan pristup korištenja niza nepovezanih metoda u analizi vremenskih serija u području otkrivanja tržišnih zakonitosti..

Važnost ovoga problema vidljiva je u radovima [Pratt, 2001.], [Manilla, 1997.], [Bradley, 1997.], [Han, 1998], [Xsniaping, 1998.], koji, nakon nuđenja rješenja za određenu problematiku u okviru svojih radova, kao moguća buduća istraživanja na svojim modelima, često spominju probleme koje je moguće riješiti s već postojećim metodama rudarenja podataka ili pomoću već otkrivenih metoda, navodeći kao glavni argument superiornost svoga modela u određenoj oblasti analize.

Na taj način oni i dalje ne nude odgovor kako postići "kompatibilnost " između različitih kvalitetnih metoda iz ostalih segmenata analize i njihove metode koja nudi prihvatljivije rješenje za usko specijalizirano područje.

Problem se još više produbljuje kada vremensku seriju želimo analizirati primjenom metoda rudarenja podataka, kao što su npr. stabla odlučivanja i klasteriranje, a da ne govorimo o analizi tipa otkrivanja sezonskih oscilacija na razini vremenskoga klastera.

Osnovni problem nemogućnosti povezivanja proizlazi iz činjenice što ne postoji jedinstveni polazišni model analize, niti metodologija koja bi povezala niz različitih koncepcija analize na razini podataka.

Tako Manilla [Manilla, 1997.] u okviru svoga rada o analizi događaja i epizoda unutar vremenskih serija, gdje su događaji definirani kao, primjerice, zvonjava alarma, krađa, bijeg i slično, ukazuje na moguću analizu vremenske serije koja se sastoji od numeričkih vrijednosti te upozorava da treba osmisliti način definicije događaja na temelju numeričkih vrijednosti.

Xsniaping , [Xsniaping, 1998.] u svom radu, koji prikazuje metodu za otkrivanje uzoraka unutar vremenskih serija, kao orijentiranost na buduća istraživanja spominje razvoj modela procjene vjerojatnosti trendova vremenskih odsječaka na temelju vlastitog modela.

Han [Han, 1998.], prikazuje model otkrivanja uzoraka koji se temelji na supstituciji vremenskih vrijednosti alfanumeričkim oznakama i računanju frekvencija pojavnosti, ali ga Prat [Pratt, 2001.] kritizira zbog nepreciznosti modela koja proizlazi iz grube definicije vremenskih odsječaka te nudi svoj model za otkrivanje uzoraka temeljen na minimalnim i maksimalnim vrijednostima. Kao moguća buduća istraživanja, navodi metodologiju otkrivanja cikličkih oscilacija unutar svoga modela, koji je Hanov model

sposoban riješiti, a moguće ga je učiniti preciznijim primjenjujući model transformacije kao polazišnu osnovu analize.

Mnogi autori [Pratt, 2001.], [Han, 1999,2.], [Bettini, 1998.] , [Sheng, 2000.], rješavajući usku problematiku u domeni analize vremenskih serija, ne nude rješenje na koji način njihov model analize metodološki povezati s ostalim uspješnim modelima, što bi uvelike pridonijelo učinkovitosti analize u oblasti marketinških analiza.

Sva navedena rješenja mogu se vrlo uspješno primijeniti u oblasti istraživanja i analize tržišta, pri čemu je nepovezanost metoda glavni problem spomenutoga područja.

Probleme koji se često navode kao predmet budućih istraživanja, kao što je to slučaj kod Pratta i Xsniapinga, moguće je riješiti povezivanjem postojeće metodologije s metodama rudarenja podataka ili već otkrivenim metodama, uz uvjet postojanja jedinstvene koncepcije modela transformacije vremenske serije kao polazišne osnove za analizu.

Ako u obzir uzmemo usku stručnu problematiku (analiza događajnosti u alarmnim sustavima, analiza oscilacija u financijskim podacima,...), koja je bila poticaj za razvoj određenih metoda, tada taj problem ne dolazi toliko do izražaja. No, ako cjelokupnu problematiku promatramo kroz prizmu analize, segmentacije i otkrivanja zakonitosti tržišta na temelju vremenskih serija, tada je ovaj problem sve izraženiji.

Problem nepovezanosti metodologija analize vremenskih serija i njene implikacije na analizu tržišta mogu ilustrirati primjerom, npr. kada želimo znati postoji li određena ciklička pravilnost ponašanja određene skupine klijenata i kolika je vjerojatnost nekog događaja (npr. dizanja "velikog" iznosa gotovine na bankomatu) unutar tjedan dana nakon odvijanja te cikličke pravilnosti.

Na ovom ilustrativnom primjeru vidljiv je čitav niz poteškoća koje proizlaze iz analize, ako je promatramo na tradicionalan način.

Problemi se javljaju kada želimo opisati cikličku pravilnost kao epizodu, što naravno ovisi i o metodi koju smo primijenili za otkrivanje cikličke oscilacije, te kada želimo definirati događaj unutar vremenske serije.

Daljnji je problem kako Bayesovom mrežom procjenjivati vjerojatnost odvijanja događaja nakon pojavljivanja epizode. Odnosno, osnovni se problem javlja u metodologiji inicijalizacije parametara unutar tablice uvjetnih vjerojatnosti.

Ako izaberemo tradicionalan pristup analizi koji nema jedinstven polazišni model, pojavljuje se čitav niz tehničkih problema koje je vrlo teško ili gotovo nemoguće riješiti. Ako taj problem i uspijemo riješiti u okviru klasične analize, stvoreni podatkovno-metodološki mostovi bit će nam od vrlo male ili nikakve koristi kod rješavanja nekog novog kompleksnog problema iz domene vremenskih serija.

Marketinški analitičar se ovdje nalazi pred ozbiljnim problemom na koji način povezati niz metoda, što je posebno izraženo kada na vremensku seriju želimo primijeniti tradicionalne metode rudarenja podataka, od kojih spominjem stabla odlučivanja, klasteriranje, analizu tržišne košarice i slično.

Primjena ovih metoda može biti od velikog značaja kada npr. želimo klasterirati klijente na osnovi vremenskih serija te otkriti koja je skupina klijenata po klasterima prema obilježju dobi sklonija koristiti uslugu tipa X u prvoj polovici tjedna.

Praksa pokazuje da za kvalitetnu tržišnu analizu baziranu na vremenskim serijama nije dovoljno koristiti samo jednu metodu. Ponekad, nakon što otkrijemo uzorak unutar vremenske serije, želimo znati je li to ponavljajući uzorak, prethodi li nekom događaju, želimo znati kako će neki događaj utjecati na taj uzorak, s kojom se vjerojatnošću pojavljuje u određenom mjesecu ili tjednu, učestvuju li klijenti određenih karakteristika više ili manje u tom događaju.

Tradicionalnim pristupom vrlo teško možemo vršiti ovakvu vrstu analiza.

Povezivanjem različitih modela analize spomenuti primjeri, koje je tradicionalnim putem nepovezanih metodologija teško ili nemoguće riješiti, bili bi rješivi unutar koncepcije jedinstvenoga modela transformacije vremenske serije, koji bi za zadatak imao precizan (jednoznačan) opis vremenske serije i koji bi uz to bio temelj analizama.

Povezivanje različitih modela analize u analizama vremenskih serija dovelo bi ne samo do sinergijskog efekta, već bi otvorilo mogućnost konstrukcije novih algoritama koji bi bili u stanju rješavati kompleksnije probleme u oblasti marketinške analize, poput segmentiranja tržišta na temelju vremenskih serija.

Osnovni razlozi nemogućnosti povezivanja različitih koncepcija i metoda u analizama vremenskih serija proizlaze iz prvenstvene usmjerenosti dosadašnjih autora na rješavanje određenoga tipa problema, a ne na polje analize tržišta i potreba koje proizlaze iz toga područja.

Takav pristup zanemaruje važnost modela transformacije vremenske serije, što je rezultiralo spomenutim problemima. Kao ilustraciju važnosti ovog problema citiram tvrdnje iz knjige: Dorian Pyle: "Data preparation for data mining", MKP, 1999 (www.data-miners.com)

"Podaci vremenske serije zahtijevaju više angažmana prilikom pretprocesiranja podataka od strane analitičara nego što je to slučaj sa netemporalnim podacima. Ne postoje potpuno automatizirani postupci i procesi za pretprocesiranje vremenskih serija, te je se analitičar oslanjao na vlastito iskustvo i znanje prilikom izbora optimalnog načina pretprocesiranja. U pretprocesiranju vremenskih serija najviše se koriste metode vizualizacije i različiti modeli pretprocesiranja ovisno o modelu kojim ćemo analizirati vremensku seriju." str. 299.

"U budućnosti, kada kompjuterski sustavi osnaže, bit će moguće automatski pretprocesirati vremenske serije heurističkim i algoritamskim postupcima. Automatsko pretprocesiranje vremenskih serija ovisi o pristizanju mnogo snažnijih i jeftinijih kompjuterskih sustava." str. 348

REFII¹ model konstruiran je s ciljem objedinjavanja različitih koncepcija analize vremenskih serija, tradicionalnih metoda rudarenja podataka, te s ciljem konstrukcije

¹ REF II model (Akronim od Rise, Equal, Fall; pri čemu II označava drugu dorađenu generaciju REF modela)

novih algoritamskih postupaka iz oblasti analize tržišta, kao i pronalaženju rješenja problema automatskog pretprocesiranja vremenskih serija temeljeći se isključivo na novom idejnom rješenju a ne na snažnijem hardverskom okruženju.

Razlog razvoja i primjene REFII modela proizlazi iz empirijskih iskustava temeljenih na analizi tržišnih zakonitosti u području segmentacije tržišta te na procjenama ponašanja tržišnih segmenata.

Važnost problema razjedinjenosti metodoloških postupaka analiza vremenskih serija uočio je Graham Williams u svom radu [Williams, 2002] koji daje pregled data mining metoda u domeni vremenskih serija. U istom radu naveo je izazove budućih istraživanja (eng. challenge questions) gdje bi trebalo riješiti problem generalne teorije analize vremenskih serija koja bi objedinila sadašnja i buduća istraživanja na području data mining analize vremenskih serija.

On ukazuje na dosadašnju razjedinjenost i nekompatibilnost metoda, te na loše rezultate dosadašnjih istraživanja u domeni objedinjavanja različitih koncepcija analize u temporalnom rudarenju podataka.

Kao dio rješenja nudi teorijski koncept nazvan *general hidden distribution based analysis theory for temporal data mining*, te ukazuje na čitav niz trenutno nedostajućih elemenata koje navodi kao predmete budućih istraživanja. Ovo rješenje bi objedinjavalo njegova dosadašnja istraživanja [Williams, 2000], [Williams, 2001], [Williams, 2001], [Williams, 2001a], [Williams, 2002a], [Williams, 2003] , [Williams, 2003a] te bi za cilj imalo kreiranje univerzalnog rješenja za analize vremenskih serija u domeni temporalnog data mininga.

Osnovni nedostatak spomenutog Williamsovog modela proizlazi iz činjenice što je premala pažnja usmjerena na model transformacije, te je on stavljen u drugi plan prilikom definicije rješenja.

Nadalje, premala pažnja usmjerena na fleksibilnost prilikom analize (fleksibilno kreiranje razreda otklona od strane analitičara), te se ovaj problem predstavlja kao predmet budućih istraživanja. Predloženo rješenje također zanemaruje kvantitativne aspekte vremenskih serija (površinu ispod krivulje).

Zbog toga su najavljena buduća istraživanja usmjerena na povećanje fleksibilnosti, iznalaženje rješenja kako vremensku seriju povezati direktno sa tradicionalnim data mining metodama, te je izražen problem povezivanja vremenske serije sa fuzzy logikom unutar Williamsovog predloženog rješenja, i ovaj se problem spominje kao predmet budućih istraživanja. Paradoksalno je to što za primjerice povezivanje vremenske serije sa tradicionalnim data mining metodama, Williams namjerava prilagođavati tradicionalne data mining algoritme bez razmišljanja o izgradnji jedinstvenog modela transformacije vremenske serije. Jedinstven model transformacije definiramo kao polazišni model koji služi kao temelj za provođenje čitavog niza analiza iz domene temporalnog rudarenja podataka. Pojam jedinstvenosti ovdje je sinonim za univerzalnost transformacije vremenske serije.

Glavna kritika Williamsove koncepcije odnosi se na :

- Svi problemi proizlaze zbog nedovoljno dorađenog i nefleksibilnog modela transformacije vremenske serije koji se u hodu dorađuje s obzirom na analitičke potrebe [Williams, 2000], [Williams, 2001], [Williams, 2001], [Williams, 2001a], [Williams, 2002a], [Williams, 2003] , [Williams, 2003a]
- Zanemarivanje kontingencijskog pristupa analizi

- Nije predviđena metodologija povezivanje temporalnih atributa sa netemporalnim
- Kao aktivnosti budućih istraživanja napominje se prilagođavanje i dorada postojećih data mining algoritama, umjesto da se putem modela transformacije izgrade mostovi za spajanje, što naravno implicira i generalni problem sa primjenom fuzzy logike unutar modela

REFII model ima odgovore na probleme koji su navedeni kao rezultati budućih istraživanja u Williamsovom radu, te nudi puno fleksibilniji pristup analizi vremenskih serija od one koju je zamislio spomenuti autor. Williams upada u klasičnu zamku prilikom pokušaja objedinjavanja metoda, a to je fokusiranje na rješavanje parcijalnih analitičkih zadataka, pri čemu ne postoji univerzalno razrađeni model transformacije vremenske serije, već sakuplja dijelove niza parcijalnih rješenja koje se pokušava objединiti u konačni mozaik, posredstvom nedovoljno razrađenog modela transformacije koji bi trebao imati centralno mjesto u povezivanju različitih koncepcija.

Svrha disertacije je ukazati na skrivene potencijale za analizu koje leže u vremenskim serijama, te modalitete rješenja primjenom jedinstvenog modela transformacije vremenske serije i njegovu centralnu ulogu u povezivanju različitih metodoloških postupaka analize.

Isto tako rad namjerava ukazati na mogućnost modeliranja rješenja iz poslovne prakse posredstvom opisanog modela, te ukazati na čitav niz rješenja koji u sebi objedinjavaju dio modela ili cjelokupnog modela.

Karakteristika modela na kojem se temelji ovaj rad je kompatibilnost i otvorenost prema ostalim data mining metodama. To omogućava daljnje procesuiranje vremenskih serija posredstvom algoritama kao što su standardne statističke metode, *neuronske mreže, klasteriranje, stabla odlučivanja, memorijski temeljeno razlučivanje, metoda potrošačke košarice*. Na ovaj način vremenske serije dobivaju vrlo moćan instrumentarij za analizu. Osim što prezentirani model sam po sebi pruža niz mogućnosti po pitanju njihove analize, procesuiranje vrijednosti bilo konkretne serije bilo njenog odsjeka posredstvom raspoloživih data mining metoda te njihovim ulančavanjem pruža dodatne mogućnosti po pitanju ekstrakiranja znanja.

Cilj istraživanja je bio istražiti mogućnosti nadogradnje REFII modela, kao i njegovih mogućnosti sintetiziranja postojećih algoritama rudarenja podataka i analize vremenskih serija u jedinstven sustav analize tržišnih zakonitosti i segmentacije tržišta na temelju vremenskih serija.

S tim u skladu, cilj istraživanja jest ponuditi rješenja za različite modalitete segmentacije tržišta i otkrivanja tržišnih zakonitosti (modeli ponašanja tržišnih segmenata i subjekata) na temelju vremenskih serija, posredstvom REFII modela.

1.2. Hipoteze i obrazloženje hipoteza

Jedinstveni model transformacije vremenske serije (REFII) omogućuje povezivanje različitih konceptualnih modela analize vremenske serije, što je unapređenje u odnosu na tradicionalan način korištenja niza nepovezanih metoda prilikom analize.

REFII model također omogućuje analize vremenskih serija primjenom tradicionalnih metoda rudarenja podataka (stabla odlučivanja, klasteriranja, metode potrošačke košarice,...)

Povezivanjem različitih konceptualnih modela analize u okviru REFII modela možemo riješiti složene analitičke zadatke na temelju vremenskih serija, poput segmentacije tržišta te otkrivanja tržišnih zakonitosti (modeli ponašanja tržišnih segmenata i subjekata).

Primjena REFII modela nudi unapređenje u otkrivanju tržišnih zakonitosti iz vremenskih serija prilikom "ad hoc" analize tržišnih problema za razliku od tradicionalnoga pristupa, prilikom kojega koristimo niz nepovezanih i nekompatibilnih metoda, koje nam ponekad ne dopuštaju mogućnost dublje analize i modeliranje rješenja za nestandardne probleme povezivanjem niza različitih metodoloških postupaka analize.

Unapređenje pojmovno znači povezivanje različitih metodoloških koncepcija analize vremenskih serija, primjenu tradicionalnih metoda rudarenja podataka na vremensku seriju, ulančavanje metoda za analizu vremenskih serija te nadogradnja temeljnog REFII modela s izvornim algoritmima s ciljem izgradnje modela za segmentaciju tržišta na osnovi vremenskih serija i procjene ponašanja tržišnih segmenata i subjekata.

Dublja analiza pojmovno znači mogućnost provedbe različitih tipova analiza na reduciranom setu podataka iz vremenske serije, koji je proizašao kao rezultat obrade prije primijenjene metode ili analitičkoga postupka.

Povezivanje se ostvaruje posredstvom jedinstvenoga modela transformacije vremenske serije, koji je temelj analitičkim postupcima, metodama te razvoju novih metoda s obzirom na problemski prostor.

Dosadašnje metode, koje se bave analizom vremenskih serija, fokusirane su uglavnom na određeni problem te se ponašaju poput "crnih kutija". Na osnovi ulaznih parametara (vremenska serija) one daju informacije poput postojanja sezonskih oscilacija ili ponavljajućih uzoraka i slično. Za sve njih karakteristično je da imaju vlastite modele transformacije vremenske serije koji su uglavnom prilagođeni konkretnoj problematici kojom se bave te, osim što su međusobno nekompatibilni, ne može ih se ekstrahirati kao izlazne vrijednosti iz modela.

REFII model nudi jedan do sada sasvim novi pristup u analizi.

Novi pristup, koji donosi REFII model, očituje se u modelu transformacije kao osnovi analize, na kojem se temelje sve daljnje analitičke metode.

Analitičke metode su algoritamski postupci kojima se djeluje na transformirane vrijednosti, a ti algoritamski postupci mogu biti izvorni algoritamski postupci koji rješavaju određene tipove problema, poput tržišne segmentacije na temelju vremenskih serija ili pak tradicionalni algoritmi za rudarenje podataka, kao što je to, primjerice, stablo odlučivanja, odnosno novootkrivene metode za analizu vremenskih serija prilagođene REFII sintaksi.

Ovakva orijentiranost na model transformacije kao polazišne točke analize, omogućuje i temeljitiju analizu vremenskih zakonitosti ulančavanjem metoda koje djeluju na vremensku seriju te modeliranje rješenja za nestandardne probleme, proizašle iz otkrivanja tržišnih zakonitosti i segmentacije tržišta.

Pojam *temeljitosti* u ovom slučaju znači mogućnost provedbe različitih tipova analiza na reduciranom setu podataka iz vremenske serije koji je proizašao kao rezultat obrade prije primijenjene metode ili analitičkoga postupka, s ciljem rješavanja kompleksnih zadataka kao što je to npr. segmentacija tržišta na temelju vremenskih serija.

Kao primjer navodim hipotetsku mogućnost segmentacije klijenata prema kriteriju aktivnosti unutar vremenskoga perioda te procjenu vjerojatnosti kretanja trendova za svaki segment pojedinačno, sve u okviru istoga sustava.

Daljnja prednost ovoga pristupa očituje se i u mogućnosti definiranja stupnja željene pouzdanosti analize unutar jedinstvenoga modela, što implicira da analitičar sam određuje hoće li se u analizi baviti globalnim trendovima ili preciznom analizom, što, naravno, ovisi o karakteru problema koji se rješava.

REFII model transformacije konstruiran je s težnjom jednoznačnoga opisa krivulje vremenske serije, tako da se vremensku seriju transformira u niz vrijednosti koje predstavljaju površinu ispod vremenskoga odsječka, koeficijent kutnog nagiba pravca vremenskoga odsječka i oznaku kretanja trenda.

Nad takvim modelom transformacije moguće je vršiti analize primjenom otkrivenih algoritmiranih postupaka za analizu vremenske serija, ali takva koncepcija otvara mogućnost razvoja novih algoritama za analizu vremenskih serija kao što su, primjerice, algoritmi za tržišnu segmentaciju na osnovi vremenskih serija te primjenu tradicionalnih metoda rudarenja podataka u okviru istoga temeljnog modela transformacije.

1.3. Metode rada

U radu je osnovni teorijsko-znanstveni model podvrgnut empirijskim istraživanjima na realnim podacima te se na osnovi tih istraživanja izvode zaključci o primjenjivosti i efikasnosti modela.

Efikasnost modela je u konkretnom slučaju sposobnost povezivanja različitih analitičkih koncepcija, bilo izvornih bilo tradicionalnih, s ciljem provođenja složenih analiza vremenskih serija u području istraživanja tržišnih zakonitosti i segmentacije tržišta.

Iako je REFII model primjenjiv i na ostale vrste vremenskih serija za potrebe empirijskih istraživanja, u empirijskim istraživanjima uglavnom je korištenja baza štetnika - populacije korisnika police auto-osiguranja.

Osnovna metoda rada i istraživanja svodi se na primjenu algoritamskih metoda posredstvom REFII modela na uzorcima podataka te na stvaranje metodološkog i praktičnog temelja u povezivanju različitih koncepcija analize s ciljem dolaženja do

rješenja za različite tipove problema iz područja analize tržišta na temelju vremenskih serija.

Istraživanje se provelo na temelju podataka iz realnih transakcijskih baza podataka (štetnici – korisnici automobilske police osiguranja), a prilikom analize korištena je metodologija rudarenja podataka u *formiranju uzoraka* za učenje i testiranje modela.

Kroz rad se je pokazana učinkovitost REFII modela koja se odnosi na integraciju niza tradicionalnih koncepcija analize vremenskih serija, prilagođenih REFII notaciji s ciljem provođenja kompleksnih analiza vremenskih serija proizašlih iz podataka s tržišta (transakcijske baze podataka). U skladu s tim, istraženo je rješavanje problematike segmentacije tržišta i otkrivanja tržišnih zakonitosti (modeli ponašanja tržišnih segmenata i subjekata) na temelju vremenskih serija, posredstvom REFII modela.

Tijekom provođenja empirijskih istraživanja, korišteni su i analitički alati: *SPSS*, *GeNie*, *AnswerTree* s ciljem povezivanja REFII modela sa standardnim metodama rudarenja podataka, od kojih spominjem *Bayesove mreže*, *stabla odlučivanja*, *klasteriranje*.

S tim u skladu, primijenjeni su algoritmi rudarenja podataka (*stabla odlučivanja*, *klasteriranje*, *Bayesove mreže*,...), te tradicionalni i izvorni algoritmi analiza vremenskih serija na vremenske serije.

Tijekom provođenja empirijskih istraživanja, korišteni su programski jezici: *Visual Fox pro 8.0* i *Python* za programiranje objekata proizašlih iz REFII modela, kao i za razvoj vlastitih rješenja u okviru spomenutoga modela.

Teorijski modeli isprogramirani su u vidu objekata, i to objekte za jednoznačnu transformaciju vremenske serije te objekte za analizu vremenskih serija koji nisu do sada softverski riješeni, odnosno neprilagođeni su REFII notaciji. Programi su pisani u programskim jezicima : *Visual Fox Pro 8.0*, *Python*, *SPSS* skriptni jezik.

Tijekom provođenja empirijskih istraživanja, korišteni su programski jezici: *Visual Fox pro 8.0* i *Python* za programiranje objekata proizašlih iz REFII modela, kao i za razvoj vlastitih rješenja u okviru spomenutoga modela.

1.4. Očekivani znanstveni doprinos

Znanstveni doprinos očituje se u predstavljanju jedne nove koncepcije u analizi vremenskih serija u području istraživanja i segmentacije tržišta, koja, za razliku od tradicionalnoga pristupa ovoj problematici, u žarište stavlja model jednoznačne transformacije vremenske serije.

Prednost ovakvoga pristupa očituje se u sintezi niza različitih pristupa i metoda u analizi vremenskih serija, u koje spadaju i tradicionalne metode rudarenja podataka, kao i nadogradnja novim algoritamskim postupcima analize.

Ovakav pristup omogućuje provedbu različitih tipova analiza na reduciranom setu podataka iz vremenske serije, koji je proizašao kao rezultat obrade već primijenjene metode ili analitičkoga postupka.

Ovakav pristup daje odgovore na to kako provoditi kompleksne tržišne analize temeljene na vremenskim serijama koje je tradicionalnim pristupom vrlo teško ili gotovo nemoguće izvesti.

Kao ilustraciju svojoj tvrdnji, istražena su rješenja problema segmentacije tržišta i otkrivanja tržišnih zakonitosti (uzorak podataka korisnika police auto-osiguranja koji su imali prometnu nezgodu) na temelju vremenskih serija posredstvom REFII modela, što zahtijeva angažman niza ulančanih analitičkih postupaka, kako tradicionalnih, tako i izvornih.

Ove mogućnosti su do sada tradicionalnim pristupima analizi vremenskih serija zbog primjene niza nepovezanih metoda (nepostojanja generalne strategije povezivanja) bile nemoguće ili vrlo teško izvedive, a pojavljuju se kao nužnost u analizi tržišnih zakonitosti putem vremenskih serija.

1.5. Vremenske serije i analiza podataka u oblasti istraživanja tržišta

Vremenske serije možemo definirati kao niz kronološki uređenih vrijednosti neke pojave.

Postoje različiti pristupi u analizi vremenskih serija koji dobrim dijelom ovise o problemskom prostoru u okviru kojeg se vrše analize nad vremenskim serijama. Tako primjerice postoje specijalizirani modeli za makroekonomska istraživanja, predikcije vrijednosti u budućim razdobljima, telekomunikacijski modeli ili pak modeli koji se koriste u oblasti fizike.

Ovaj je rad fokusiran prvenstveno na otkrivanje znanja iz vremenskih serija iz oblasti analize tržišta.

Dosadašnja metodologija analize vremenskih serija u domeni analize tržišta ne poznaje jedinstveni model transformacije, kao ni usvojenu metodologiju pretprocesiranja vremenskih serija s ciljem analize.

Općenito gledajući postoji niz nužnih, često puta korištenih postupaka u analizama vremenskih serija koje predmnijevaju uspjeh analize bez obzira na korištenu metodu i bez obzira na njihovu međusobnu nekompatibilnost.

U jedan od nužnih postupaka koji prethode bilo kakvoj analizi vremenskih serija odnosi se na unifikaciju vremenskih pomaka na jednake intervale. Taj postupak naziva se normiranjem vremenskih distanci.

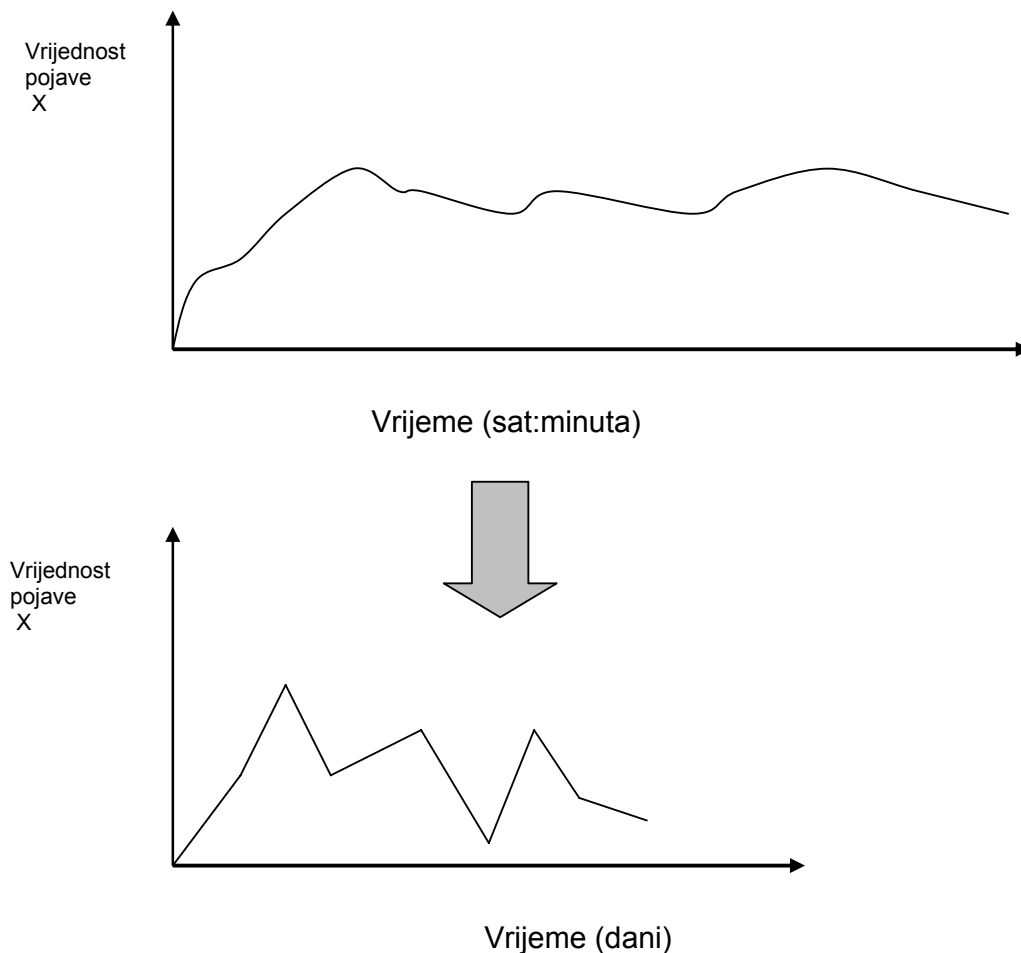
Pod pojmom normiranja podrazumijevam svođenje vremenske serije na jednake vremenske pomake između vrijednosti atributa. Prilikom normiranja moramo uzeti u obzir sve vremenske podintervale kako bismo dobili jedinstvenu vrijednost na kraju

vremenskog pomaka. U literaturi [Pyle, 2001] se prilikom provođenja ovog procesa prakticira metoda srednje vrijednosti, medijana moda, te metoda sumiranja.

Normiranje vremenske serije u uskoj je korelaciji sa stupnjem kondenziranosti (granulacije) vremenske serije.

Pojednostavljenim rječnikom rečeno mi vremensku seriju možemo “kondenzirati” na minute, sate, dane, mjesece. Procjena stupnja granulacije vremenske serije ovisi o konkretnom problemu kojega rješavamo.

Slika 1.1. prikazuje opisanu metodologiju i ilustrira razloge, zašto se prilikom analize vremenskih serija pristupa granulaciji koja se ujedno i može shvatiti kao pretprocesiranje podataka. Kao što je vidljivo iz slike 1.1. pojave/ poslovni događaji mogu se odvijati u nejednolikim vremenskim razmacima, što je neprikladno za provođenje analiza. Da bi podaci bili podesni za provedbu analiza, potrebno ih je svesti na jednolike vremenske razmake (sumiranje, prosjeci).



SLIKA 1.1. PRETPROCESIRANJE VREMENSKE SERIJE

Na primjer, prilikom pristizanja informacija koje čine vremensku seriju, te informacije mogu dolaziti u nejednolikim vremenskim razmacima (npr. sat i minuta obavljene transakcije na bankomatu) Iako je ta podjela detaljna i opisuje realnu situaciju, ona je analitičaru kod evaluacije vremenskih serija neprimjerena za analizu.

Prema nekim autorima [Westhpal, 2001] neobrađene vremenska serija može biti interesantna za promatranje nekim od alata za vizualizaciju, pri čemu možemo vizualno odrediti trendove, gomilišta i sezonske oscilacije usprkos nejednolikoj vremenskoj raspodjeli. No, ozbiljnijoj analizi vremenskih serija u domeni marketinških analiza možemo pristupiti nakon svođenja vremenske serije na jednake odsječke kao što je to prikazano na slici 1. Nakon ovog postupka vremenska serija se transformira u prepoznatljiv *zupčasti* oblik koji je podijeljen na jednake vremenske odsječke, što je polazišna osnova za veliki broj algoritama za analizu vremenskih serija, pa tako i REFII model .

Proces normizacije odnosno granulacije spada u etapu pretprocesiranja podataka, a ovom procesu prethodi proces čišćenja podataka. Pod pojmom čišćenja podataka u vremenskoj seriji podrazumijevamo procese dijagnosticiranja i otkrivanja nedostajućih vrijednosti atributa i njihovu supstituciju sa vrijednostima koje mogu biti izgenerirane posredstvom algoritmiranih postupaka linearne interpolacije, MLP algoritmima i slično.

Bitna predradnja prije samog pristupanja analizama vremenskih serija jest otkrivanje "ekstremnih vrijednosti" (eng. outliers).

Ekstremne vrijednosti mogu nastati smetnjama u sustavu, ali isto tako mogu biti činjenično stanje koje upućuje na određenu anomaliju koja može biti putokaz u otkrivanju nekih ekstremnih stanja promatrane pojave. Uobičajena metoda za otkrivanje eksternih vrijednosti koja je obuhvaćena u okviru softverskih paketa je metoda vizualizacije vremenske serije [Westhpal, 2001], nakon čega analitičar procjenjuje da li je određena ekstremna vrijednost nastala kao rezultat smetnji sustava, ili pak kao rezultat objektivnih vrijednosnih kretanja promatranog atributa na osnovu realnih tržišnih ili nekih drugih situacija. U slučaju smetnji ekstremne vrijednosti se izoliraju iz promatranog uzorka, u protivnom ostaju u uzorku kako bi se podvrgli procesu analize. Autori poput Pyle-a prednost prilikom otkrivanja ekstremnih vrijednosti podataka daju klasičnoj statistici [Pyle, 2001] .

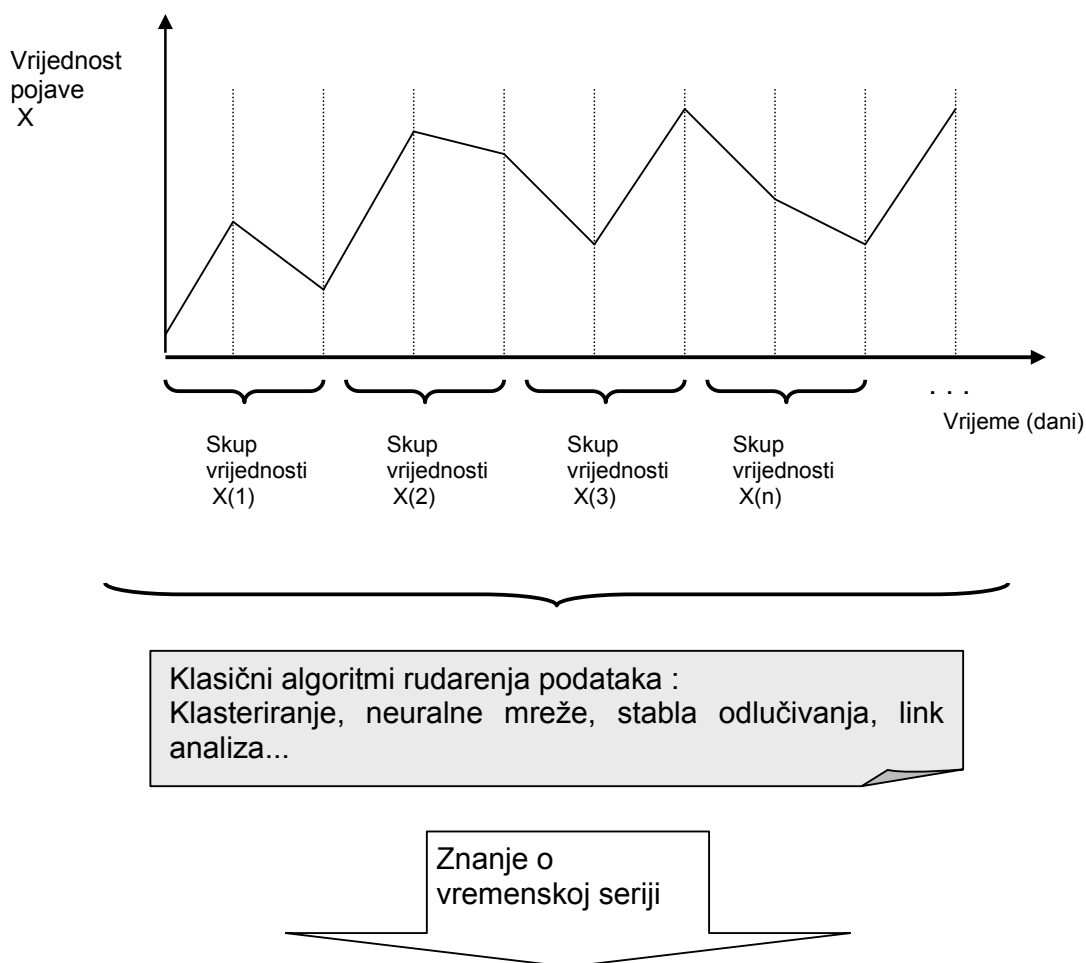
Najčešći tipovi analize koji se provode nad vremenskim serijama koncentrirani su na pravilnosti kretanja neke pojave unutar vremenskih odsječaka (npr. otkrivanje sezonskih oscilacija), predviđanje (eng. forecasting), otkrivanje senzibiliteta i jačine pomaka atributa u atomarnom vremenskom odsječku.

Algoritmi koji uspješno rješavaju ove tipove problema izvedeni su iz klasičnog matematičkog instrumentarija i svode se uglavnom na vektorski račun, a u pojedinim rješenjima i na fuzzy logiku.

Dosadašnji algoritmi pristupali su vremenskoj seriji prvenstveno kao području gdje se klasične data mining metode nisu direktno mogle primijeniti na vremensku seriju, nego se za taj tip problematike koristio instrumentarij specijaliziran samo za to područje. Model (REF II) koji će biti predstavljen specijaliziran je za analizu vremenskih serija, ali isto tako otvara vrata ka primjeni klasičnih data mining algoritama na vremensku seriju. Na taj način otvara se čitav niz novih mogućnosti kao što je primjerice klasteriranje odsječaka (uzoraka) vremenskih serija, klasifikacija (stabla odlučivanja) odsječaka (uzoraka) vremenskih serija, predviđanje (neuralne mreže) daljnjeg trenda vremenske serije na osnovu odsječaka (uzoraka) vremenske serije. Jednoznačna definicija vremenske serije posredstvom niza parametara otvara

moгуćnost korištenja klasićnih algoritama u otkrivanju znanja i prepoznavanja uzoraka.

Shema metodologije primjene odsjećaka vremenske serije u klasićnim algoritmima rudarenja podataka posredstvom REF II modela prikazana je na slici 1.2.



SLIKA 1.2 METODOLOGIJA PROCESUIRANJA VREMENSKE SERIJE DATA MINING ALGORITMIMA POSREDSTVOM REF II ALGORITMA

Jedna od vrlo bitnih karakteristika prikazanog modela svodi se na činjenicu da je njegovom primjenom moguće izmodelirati rješenja za određene tipove problema kako u bankarstvu tako i u područjima trgovine, medicine, prepoznavanja uzoraka i slično.

1.6. Različiti pristupi problematici analize vremenskih serija u domeni istraživanja tržišta

Gledajući sa perspektive fokusa interesa analizu vremenskih serija možemo podijeliti u nekoliko osnovnih kategorija:

- Analizu kretanje trenda serije
- Analize cikličkih pojava unutar serije
- Analize sezonskih oscilacija unutar serije
- Pronalaženje odsječka vremenske serije koji korespondira sa određenom tržišnom pojavom
- Pronalaženje sličnosti uzoraka unutar serije
- Pronalaženje i otklanjanje irelevantnih sekvenci iz serija
- Analiza korelatornih odnosa, kako između vremenskih serija tako i njenih odsječaka
- Autokorelatorna analiza vremenskih serija
- Predviđanje vremenskih serija
- Analiza prirasta i osjetljivosti vremenskih serija

Ovo je osnovna podjela koja uglavnom obuhvaća glavne tipove analize.

Neosporno tradicionalno najčešće korištena metoda koja često puta prethodni detaljnim analizama vremenske serije je metoda vizualizacije. Ovim načinom možemo na vrlo brz i efikasan način spoznati osnovne karakteristike promatrane pojave.

Metodi vizualizacije često puta prethodi čišćenje podataka, odnosno nadomještanje nedostajućih vrijednosti vremenske serije. Ako se je ta pojava dešavala unutar nejednakih vremenskih razmaka, potrebno je odrediti stupanj granulacije, kako je to pisano u prethodnom poglavlju s ciljem formiranja jednakih komparabilnih vremenskih odsječaka.

Ako zanemarimo nužnost vizualizacije prilikom pristupanja analizi vremenske serije, metodologija svođenja na jednake vremenske odsječke je proces koji je neminovan prije korištenja većeg broja algoritama za analizu vremenskih serija.

Klasična statistika (kao što je već spomenuto), nudi metode poput skupnih indeksa, srednje vrijednosti vremenskih serija, individualnih indeksa, regresijske metode, procjene trendova. Sa perspektive analitičara ove metode mogu dati korisne pokazatelje, ali naviknuti na snagu i efikasnost algoritama *data mining* algoritama iz tog područja ovo je vrlo skućeni niz pokazatelja za analitičara.

Metodologija otkrivanja znanja u vremenskim serijama u svjetlu *data mininga* vrlo se često oslanjala na izračun apsolutne vrijednosti prirasta funkcije (trenda kretanja) između vremenskih odsječaka. Osnovni element izračuna tako koncipiranog modela možemo iskazati narednom formulom:

$$t = D_i - D_{i-1}$$

Za t_i , gdje je t vrijednost trenda, D_i vrijednost neke pojave u vremenu i .

Ovaj pokazatelj daje vrijednost prirasta funkcije između dvije promatrane točke. Prikazana formula čest je temelj različitim tipovima algoritama koji u svojim temeljima imaju ugrađen pokazatelj prirasta funkcije. Ovo je formula iz koje je izveden čitav niza algoritama koji se bave analizama vremenskih serija.

Vrlo eksploatirana metoda koja se pojedinim slučajevima može iskoristiti kao metoda transformacije vremenske serije je metoda pomičnih prosjeka. Neki autori [Han, 2000]

je deklariraju kao metodu "kompresije" vremenske serije. Ovaj pristup se može koristiti prilikom vizualizacije podataka u situacijama kada želimo vizualno predložiti neku pojavu sa velikim brojem vrijednosti u seriji, ali isto i kao metoda za orijentacijsku procjenu vremenske serije.

Njen nedostatak leži u činjenici da modificira originalne vrijednosti serije, pa tako u pojedinim slučajevima nakon primjene ovog modela možemo uočiti određene pravilnosti unutar serije koje ne postoje na realnim podacima.

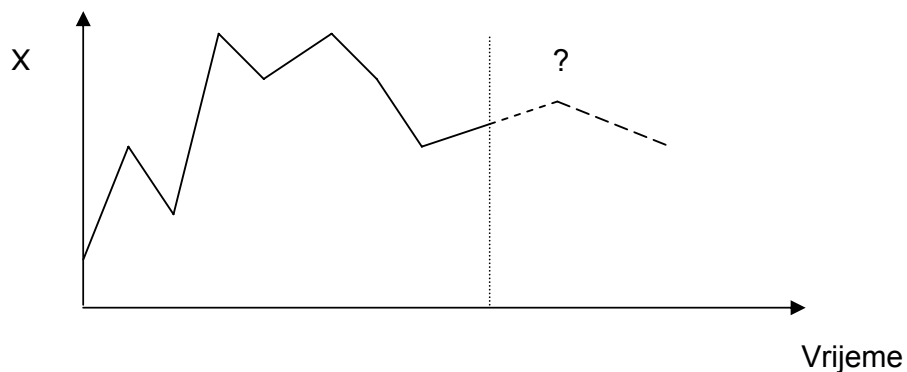
Osnovni matematički model pomičnih prosjeka glasi :

$$\frac{D_1 + \dots + D_n}{n}, \frac{D_2 + \dots + D_{n+1}}{n}, \frac{D_3 + \dots + D_{n+2}}{n}$$

Gdje je D_i vrijednost neke pojave u vremenu i , a n broj obuhvaćenih vremenskih segmenata u seriji za koju se računaju pomični prosjeci.

Traženje sezonskih oscilacija, odnosno traženje cikličkih pravilnosti unutar vremenskih serija jedno je od glavnih fokusa interesa analitičara. Do sada najeksploatiraniji pristup koji daje odgovore na tu vrstu pitanja su Fourierove transformacije [Pyle, 2001], [Han, 2000]. Nedostatak ovog pristupa očituje se u transformaciji originalnih podataka, tako da su oni u tom obliku nepodesni za daljnje analize. S obzirom na povijesni nastanak i namjenu ovog modela nova istraživanja nude nove pristupe u rješavanju ove problematike. Jedna od takvih metoda je i *leg* metoda, [Pratt, 2001]. koja u sebi ima i elemente korelatorne analize vremenskih odsječaka. Osobno smatram da u navedenom radu *leg* metoda nije dovoljno razrađena, jer ima potencijala (uz detaljniju razradu) uz spomenute mogućnosti otkrivanje sezonskih oscilacija i cikličkih pravilnosti.

Kada govorimo o predviđanju trendova kretanja vremenskih serija, tada cjelokupnu problematiku možemo promatrati iz dva aspekta. Prvi aspekt odnosi se na predviđanje trendova kretanja vremenskih serija primjenom metodologije neuronskih mreža. Ova problematika vrlo je efikasno riješena na način da model neuronske mreže predviđa nepoznate vrijednosti atributa u budućnosti kao što je to prikazano na slici 1.3.

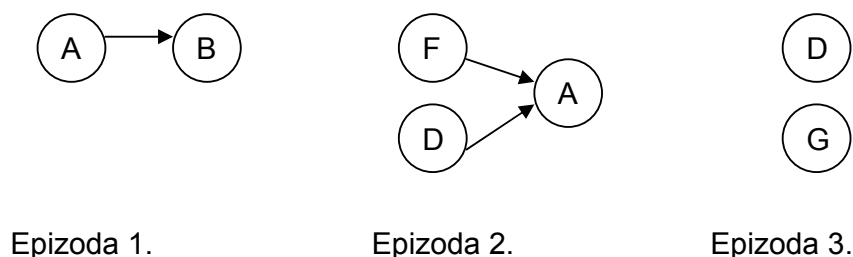


SLIKA 1.3. PREDVIĐANJE TRENDOVA PRIMJENOM NEURALNE MREŽE

Neuronska mreža (MLP) na osnovu povijesnih podataka (koje je prvo potrebno pročistiti i preprocesirati) nakon izvršenog modeliranja i treniranja predviđa izgled krivulje u budućnosti [Kliček, 2002], [Taylor, 1996], [Apostols, 1996], [Klepac, 2001].

Ovakvi modeli koriste se prvenstveno u previđanju vrijednosti neke pojave u budućnosti za koju ne postoji egzaktna formula, ili parametrizirani model.

Drugi pristup po pitanju predviđanja svodi se na dijagnostiku i otkrivanje epizoda i scenarija [Manilla, 1997]. Osnovna ideja je otkrivanje uzoraka unutar vremenskih serija koje determiniraju ponašanje u $n \times x$ vremenskih odsječaka. Znanstveni radovi koji se bave ovom problematikom pojam točke u vremenskoj seriji poistovjećuju sa događajem (događaji u alarmnom sustavu) [Manilla, 1997]. Ovaj pristup predviđanju unutar vremenskih serija približava se koncepciji formalne logike, što otvara neke nove mogućnosti po pitanju generiranja pravila. Slika 1.4. prikazuje niz hipotetskih epizoda koje se mogu pojaviti u vremenskoj seriji, a niz epizoda može činiti scenario.



SLIKA 1.4. EPIZODE U VREMENSKOJ SERIJI

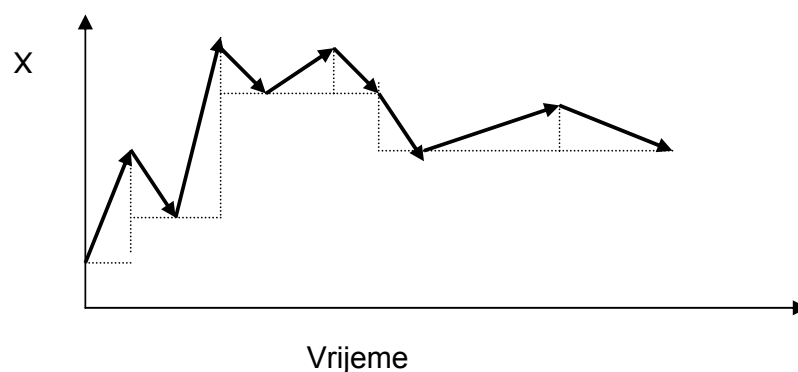
Ako promotrimo sliku 1.4., tada je lako uočiti da u ovom modelu postoje osnovni preduvjeti za primjenu link analize. Ovo je jedan od mogućih načina generiranja događaja na osnovu uzoraka. Način koji se spominje [Manilla, 1997] kao model generiranja događaja odnosi se na otkrivanje paralelnih ili sekvencijalnih epizoda u vremenskim serijama. Otkrivene uzorci mogu se kombinirati u događaje odnosno epizode, te se mogu promatrati njihovi korelatorni odnosi.

Ovaj model može biti vrlo interesantan kod pojava kao što su ponašanje potrošača, analiza elastičnosti i slično, gdje procjenjujemo uvjetnu vjerojatnost odvijanja kupnje proizvoda A ako je proizvod B kupljen prije vremena t .

Jedan od perspektive pogleda prilikom razmatranja problematike vremenskih serija jest oslanjanje na *fuzzy* logiku [Cheng, 1997], [Dougherty, 1988]. Ovaj pristup daje jednu sasvim novu dimenziju u kojoj možemo promatrati ovu problematiku. Postoji čitav niz algoritama koji se temelji na opisanoj koncepciji, gdje se mehanizmi fuzzy logike uključuju u modele vremenske serije, najčešće s namjerom izgradnje prediktivnih modela.

Jedan od pravaca u analizama vremenskih serija svodi se na transformaciju vremenske serije u vektorski model.

Odsječak vremenske serije između dvije točke možemo promatrati kao vektor u prostoru, što implicira da se na osnovnim elementima vektorskog računa razvijaju modeli. Reprezentacija vremenske serije posredstvom vektorskog prostora prikazana je slikom 1.5.



SLIKA 1.5. REPREZENTACIJA VREMENSKE SERIJE POMOĆU VEKTORA

Prikazani model koji je reprezentiran pomoću vektora može se iskoristiti kako za opis tako i za analizu odsječaka u vremenskoj seriji. Krivulja prikazana na slici 1.5, formira se na osnovu realnih podataka postupkom pretprocesiranja podataka, kako je to opisano u poglavlju 1.2. Ovakav pristup pretprocesiranju omogućava transformaciju vremenske serije u niz vektora. Udaljenosti između točaka u prostoru mogu se računati posredstvom *Euklidskih distanci*.

Ovako transformirana vremenska serija dobra je temelj za građenje analitičkog instrumentarija. Pomoću *Euklidskih udaljenosti* možemo izračunati dužine stranica "trokuta", a posredstvom trigonometrijskih funkcija možemo izračunati kuteve trokuta.

Ovaj model transformacije temelj je idejnog predložaka za jedan od tri glavna elementa REFII modela.

Vrijednosti pojave nekog obilježja u vremenskoj seriji mogu biti promatrane kroz frekvencije vrijednosti podciklusa

Vrijednosti frekvencija takvih pojava mogu se prezentirati u tablici 1.1. :

Tablica 1.1. : Frekvencija pojavnosti

Podciklus ciklusa	Vrijednost	Frekvencija pojave/ %
3	3	30 / 78%
2	7	25/ 68 %
8	9	35 / 82 %
11	34	41/ 92
...
n

Na osnovu frekvencija možemo dijagnosticirati ciklično ponavljanje vrijednosti neke pojave, odnosno sezonsku oscilaciju neke pojave. Iz tablice je vidljiva mogućnost formiranja histograma frekvencija pojavnosti vrijednosti podciklusa. Vrijednost frekvencije temelj je za izračun indeksa pojavnosti, odnosno faktora sigurnosti. U situacijama kada definiramo vrijednost faktora sigurnosti, algoritimiziranim postupkom možemo otkrivati sezonske oscilacije u krivulji. Nedostatak ove metode je "krutost",

koja proizlazi iz egzaktnosti vrijednosti. Fleksibilniji instrumentarij temeljen na ovoj ideji možemo izgraditi koristeći elemente *fuzzy* logike.

Prezentirani model svoje korijene vuče iz link analize (Anacapa model), a razvijen je kasnih šezdesetih godina prošlog stoljeća [Westhpal, 2001]. U svom izvornom obliku orijentiran je pronalaženju veza između dva objekta, te računanju njihovih frekvencija, s naglaskom na smjer veze. S obzirom da je u vremenskoj seriji smjer veze jednoznačno određen, prikazana tablica podrazumijeva smjer veze.

Vrlo značajno područje u analizi vremenskih serija u domeni tržišnih istraživanja je i otkrivanje uzoraka u vremenskim serijama, sezonskih oscilacija unutar vremenskih serija te otkrivanje sličnosti unutar vremenskih serija [Xsniaping, 1998.], [Han, 1998.], [Pratt, 2001.], [Bettini, 1998.], [Sheng, 2000.], [Manilla, 1997.].

Za navedene radove je karakterističan pristup koji vrlo malo pažnje poklanja modelima transformacije vremenske serije, te se često puta u okviru navedenih radova kao rezultati budućih istraživanja spominju problemi koji su već riješeni kod drugih autora, ali ih je nemoguće primijeniti na vlastiti model rješenja jer ne postoji kompatibilnost između modela.

Tako na primjer probleme koje je riješio Pratt [Pratt, 2001.] definicija vremenskih odsječaka na temelju empirijskih vrijednosti i procjena sličnosti, Manilla [Manilla, 1997.] u radu koji se bavi događajnošću vremenskih serija navodi kao predmet budućih istraživanja, dok Pratt rješavanje problema događajnosti navodi kao predmet budućih istraživanja.

Zbog nepostojanja jedinstvenog modela transformacije nemoguće je proširiti vlastite modele sa postojećim rezultatima istraživanja, već je potrebno kreirati problemski orijentirana, međusobno nekompatibilna rješenja.

Navedeni ilustrativni pristupi u analizi vremenskih serija spomenuti su s ciljem prezentacije različitih koncepcija koje na žalost nemaju zajedničkih dodirnih točaka koje bi osiguravale međusobnu kompatibilnost što je preduvjet za ulančavanje metoda.

Glavni zadatak REFII modela upravo je razrješavanje ovakve situacije.

Geneza njegovog nastanka započeta je REF modelom [Klepac, 2000], [Klepac, 2001], koji je nastao prvenstveno kao odgovor na krutost standardnih alata za procjenu sezonskih oscilacija. Tijekom godina primjene u praksi, pokazale su se prednosti ove koncepcije, kao manjkavosti, koji su riješene koncepcijom REFII modela.

Glavni nedostatak REF modela očitova se u njegovoj orijentiranosti na oblik krivulje, zanemarujući kvantitativne aspekte krivulje i jednoznačnost definicije krivulje. Temeljni REF model pokazao je vrlo dobra svojstva u praksi u domeni otkrivanja sezonskih oscilacija u vremenskoj seriji, epizoda u vremenskoj seriji, što je rezultiralo povećanjem koeficijenta obrtaja vezivanja sirovina i robe u skladištima.

Upravo iz tog razloga REF model je jedan od tri osnovne komponente REFII modela. Osnovne prednosti koje karakteriziraju REF model kao što su fleksibilnost, primjenjivost na različita područja, otvorenost i modularnost, karakteristika su i REFII modela. Njega također karakterizira i veća preciznost, kao i sloboda procjene stupnja preciznosti potrebnog za konkretnu analizu vezanu uz realan problem. Modularnost je ujedno i najsnažnije oružje koje karakterizira ovu koncepciju. REFII model osim što precizno opisuje vremensku seriju, i daje osnovne pokazatelje nakon transformacije, pruža široki spektar mogućnosti razvoja različitih pristupa obrade transformiranih vrijednosti.

U prvom dijelu bit će detaljno izložen matematički instrumentarij koji čini okosnicu modela, a ostatak rada fokusirat će se na konkretne probleme, i mogućnosti njihovog algoritmiziranog rješavanja korištenjem REFII koncepcije.

1.7. Karakteristike i ciljevi REF II modela

Osnovna karakteristika ovog modela je jednoznačnost opisa vremenske serije pomoću parametara modela. Matematička jednoznačnost implicira mogućnost provođenja temeljnih matematičkih operacija nad vremenskim odsječcima poput jednakosti, različitosti i sličnosti.

U uvjetima kada krivulju, ili neki njen segment možemo komparirati sa drugom krivuljom odnosno nekim njenim segmentom matematički precizno, tada dolazimo do snažnog instrumentarija na kome se može temeljiti čitav sustav analize vremenskih serija. Upravo na ovim osnovama počiva koncepcija REFII modela.

Ova je koncepcija nastojala zadovoljiti još jedan kriterij, a to je kriterij povezanosti sa algoritmima koji se primjenjuju u *data miningu*. Do sada poznate metode za analizu vremenskih serija davale su određene pokazatelje koji se kasnije nisu mogli procesuirati posredstvom nekog od poznatih algoritama u cilju ekstrakiranja dodatnog znanja. REFII model teži ka otvorenosti, odnosno uz svoj matematički instrumentarij koji služi kako za opis, tako i za generiranje znanja koje se krije u vremenskoj seriji, daje modalitete rješenja povezivanja sa ostalim *data mining* algoritmima.

Na taj način možemo iskoristiti snagu provjerenih algoritama na području vremenskih serija, u sklopu standardnih softverskih rješenja.

U skupinu klasičnih *data mining* algoritama ubrajamo neuralne mreže, klasteriranje, stabla odlučivanja, analizu potrošačke košarice, link analizu, i slično, kao i sve mutacije i izvedenice ovih algoritama.

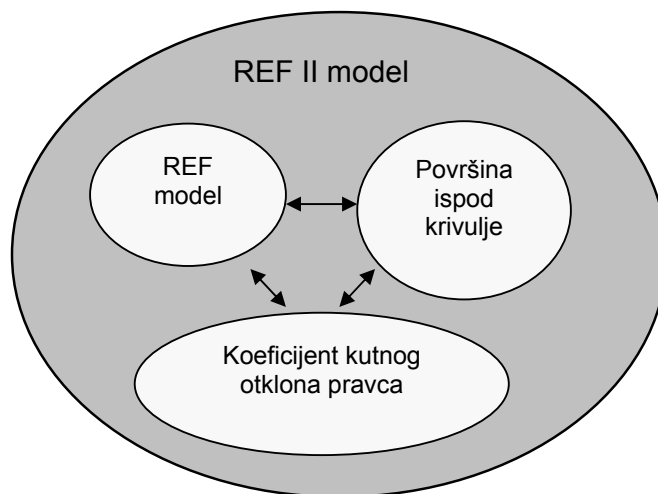
Lako je za pretpostaviti kolika snaga leži sustavu koji je sposoban klasterirati vremenske odsječke, ili sustavu koji posredstvom algoritma stabla odlučivanja klasificira vremenske odsječke ili pak kompletne vremenske serije. Isto tako možemo raditi link analize nad vremenskim serijama ili njihovim odsječcima, kao i proračunavati sličnosti vremenskih serija na osnovu funkcije udaljenosti.

Ovom koncepcijom nam se otvara čitavo jedno novo područje koje nudi detaljniji i precizniji analitički instrumentarij u domeni vremenskih serija.

REFII model se koncentrira se na tri osnovna segmenta kojima se jednoznačno može opisati krivulja, to su :

- oblik krivulje (opis izgleda vremenske serije)
- površina ispod krivulje (kvantifikacija vremenske serije)
- koeficijent kutnog nagiba pravca unutar vremenskog odsječka ("jačina" trenda)

Odnosi ovih triju elemenata koji čine REFII model prikazani su na slici 1.6.



SLIKA 1.6. KOMPONENTE REFII MODELA

U REFII modelu za opis oblika krivulje zadužen je REF model [Klepac, 2000], [Klepac, 2001]. Njegova karakteristika je dijagnosticiranje i modeliranje oblika krivulje. Na žalost, ovaj model ne može jednoznačno definirati krivulju, te se s toga služimo i sa ostala dva spomenuta elementa modela.

Površina ispod krivulje daje kvantitativnu dimenziju određenoj pojavi. Krivulja može imati jednaki oblik u odnosu na neku drugu krivulju, ali to ne znači da ima istu kvantitativnu vrijednost, iz čega prolazi pojam nejednakosti krivulje.

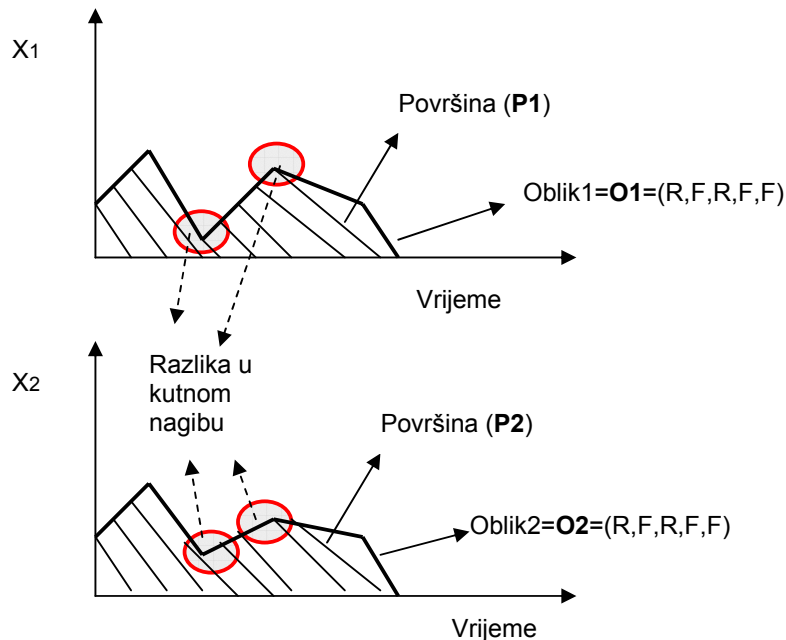
Površina ispod krivulje nam može pomoći u dobivanju tog pokazatelja.

Sa ova dva elementa možemo vrlo precizno, gotovo jednoznačno opisati vremensku seriju.

REFII model je jedinstveni model transformacije koji jednoznačno opisuje empirijsku krivulju (a sadrži diskretne vrijednosti), čija je zadaća povezati niz kako tradicionalnih, tako i *ad hoc* analitičkih postupaka.

Teoretski krivulje mogu imati i jednak oblik opisan REF modelom, i jednaku površinu a da ne budu u potpunosti jednake što se vidi iz slike 1.7. Vidljivo je da ako procjenjujemo sličnost na promatranom intervalu, na temelju *REF* oznaka i površine ispod krivulje za dvije vremenske serije prikazane na slici 1.7. možemo ustvrditi da su one slične. Ako pak u obzir uzmemo i koeficijent kutnog otklona za obje krivulje na promatranom intervalu, tada je vidljivo da je procjena puno preciznija, te ta procjena može rezultirati manjom pouzdanošću po pitanju sličnosti. To objašnjava korištenje sva tri elementa u modelu.

Iako bi nam za potrebe analize podataka ova dva elementa u velikom broju slučajeva bila dostatna za analize, naš je cilj matematički jednoznačno opisati model krivulje.



SLIKA 1.7. ULOGA KOEFICIJENTA KUTNOG NAGIBA PRAVCA U REFII MODELU

Kada kažemo “gotovo jednoznačno”, u pojedinim slučajevima ove dvije krivulje mogu zadovoljavati kriterij sličnosti. Ponekad je apsolutno nerealno očekivati potpunu jednakost krivulja za određene pojave. Ako spoznamo da je recimo petkom povećan broj transakcija na bankomatu, to ne znači da ćemo odbiti tvrdnju o povećanju transakcija petkom ako krivulje u potpunosti nisu identične.

Spoznaja o sezonskoj oscilaciji dostatan je indikator za poduzimanje akcija po pitanju ove problematike.

Da bismo u potpunosti uspjeli jednoznačno definirati vremensku seriju matematičkim modelom potrebno je uvesti i treći element, a to je kut unutar vremenskog odsječka, kojeg izražavamo koeficijentom kutnog nagiba.

Uvođenjem i ovog krajnjeg elementa dobili smo osnovne obrise REF II modela, koji je orijentiran na data mining terminologiju procjene, gdje je mjerilo pouzdanosti modela faktor sigurnosti.

Iz izloženog modela REF model se čini kao redundantni element, jer pomoću koeficijenta kutnog nagiba pravca možemo opisati oblik krivulje. Razlog uklapanja REF modela kao referentan segment sustava proizlazi iz činjenice da je vrijeme koje stroj utroši kod procjene sličnosti krivulja na temelju REF modela mnogo kraće nego kod modela koji bi se bazirao na koeficijentima kutnog nagiba pravca. Naravno, REF model je isto tako mnogo neprecizniji od modela baziranog na koeficijentima kutnog nagiba pravca, ali je isto tako mnogo “tromiji” u procesu obrade.

Prilikom procesa analize vremenske serije možemo uzeti sva tri elementa kao temelj za analizu. Ovaj model sam po sebi ne daje gotova rješenja već je polazišna osnova za kompleksnije modele. S tim u skladu i REF model i površina ispod krivulje i koeficijentima kutnog nagiba pravca mogu biti korišteni zajedno kod modeliranja rješenja složenijih problema, ali isto tako i parcijalno, što ovisi o karakteru problema koji rješavamo.

Kada govorimo o primjeni klasičnih data mining algoritama u vremenskoj seriji, tada govorimo o transferu pretprocesiranih vrijednosti opisanog modela u algoritme. S obzirom na prirodu problema kojeg rješavamo, u algoritme možemo transferirati vrijednosti proizašle iz svih segmenata REF II modela, ili samo određene vrijednosti.

U algoritme tako možemo prosljeđivati podatke o oblicima krivulja, površinama ispod krivulja, jačini trendova krivulja, te na osnovu toga možemo ekstrahirati nova znanja o vremenskim serijama. Ponekad će nam u analizi biti bitan samo obrazac trenda kretanja bez kvantitativnog aspekta i jačine nagiba, pa ćemo procesuirati podatke o obliku. U nekim slučajevima kvantitativni aspekt može igrati dominantnu ulogu te će on biti predmet razmatranja i tako redom.

Naredna tablica daje orijentacijske odnose između tipa analize i elemenata REFII modela.

Tablica 1.2. Orijentacijski odnosi između tipa analize i elemenata REFII modela

Vrsta analize	Segment REFII modela
Analiza oblika krivulje	REF i/ili koeficijent kutnog otklona
Kvantitativna analiza	Površina ispod krivulje
Traženje pravilnosti u vremenskoj seriji	REF i površina ispod krivulje
Jednoznačna definicija krivulje	REF, koeficijent kutnog otklona, površina ispod krivulje
Otkrivanje epizoda i scenarija u vremenskim serijama	REF i /ili koeficijent kutnog otklona i/ili površina ispod krivulje

Koncepcije poput otkrivanja epizoda i scenarija u vremenskim serijama isto je tako moguće realizirati primjenom REF II modela.

Osnovni zadatak i cilj REFII modela je uspješno modeliranje rješenja problema iz domene analize vremenskih serija. Kao što će biti prikazano kroz tekst REFII model je dobra polazišna osnova i temelj rješavanja problema iz poslovne prakse. Njegova primjenjivost leži u činjenici što ulazi u svaku poru vremenske serije, te ja na osnovu konkretnog problema moguće dijagnosticirati analogiju između elementa vremenske serije i konkretnog problemskog prostora.

Vremenska serija na kojoj smo primijenili model može se primjerice transformirati u niz objekata koji kao vrijednosti sadrže upravo vrijednosti proizašle iz REFII modela. To je vidljivo iz naredne tablice.

Tablica 1.3. Transformacija vremenske serije u objekte

Indeks prostorne lokacije u vremenskoj seriji (oznaka pozicije)
Oznake/oznaka REF sintakse
Vrijednost površine segmenta koju označava pozicija
Vrijednost kutnih koeficijenata promatranog segmenta

Niz vrijednosti opisane prethodnom tablicom čine osnovu REF II modela. Kao što je vidljivo iz tablice vrijednosti se mogu odnositi na odsječak vremenske serije (npr. t_0-t_4 , t_4-t_8 , t_8-t_{12} , ..., t_n-t_{n+4}), a isto tako i na odsječak između dva promatrana stanja vremenske serije (t_0-t_1 , t_1-t_2 , t_2-t_3 , ..., t_n-t_{n+1}).

Ovo je naravno ilustrativni primjer moguće strukturalne formalizacije REFII modela, koji nije jedini mogući ali zbog ilustrativnosti i jasnoće prikazan je ovdje kao hipotetski primjer.

Različiti modaliteti rješenja determiniraju različite strukturalne formalizacije modela. U određenim situacijama ta će se vremenska serija morati lomiti u manje segmente s ciljem traženja reprezentativnih uzoraka i znanja. Vrlo rijetko će ovako transformirana serija biti u originalnoj dužini. Za potrebe sofisticiranijih analiza bit će potrebno segmentirati seriju kako bi se olakšao proces analize.

1.8. Prednosti REF II modela

Standardni modeli za analizu vremenskih serija, kao što je već spomenuto koncentrirani su prvenstveno na generiranje određenih pokazatelja proizašlih iz analize te su fokusirani na usku problematiku. Ovakvi modeli ne pružaju mogućnost direktnog procesuiranja vremenske serije posredstvom klasičnih *data mining* algoritama, kao što je to slučaj sa REFII modelom.

Ovaj model nadalje pruža mogućnost jednoznačnog opisa vremenske serije, što rezultira primjenu egzaktnih matematičkih modela i procesa.

Model se oslanja na istovremeno korištenje sva tri podmodela u procesu analize, što rezultira vrlo visokim stupnjem jednoznačnosti i preciznosti u opisu krivulje. U pojedinim slučajevima analize gdje se ne zahtijeva toliko stupanj egzaktnosti i preciznosti, a potrebni su nam orijentacijski pokazatelji moguće je u proces analize ući sa jednim ili dva elementa modela.

Kao primjer navodim cikličke analize pojavnosti, gdje nas ne interesira kvantitativni aspekt pojave, nego samo pojava, tada iz navedenog sustava možemo izdvojiti REF podmodel kao temelj za analizu.

Primjenjivost REF II modela u različitim područjima od financija, trgovine, bankarstva, medicine, preko tekst *mininga*, *face recognition*, i slično karakteristika je koja ovom modelu daje dodatnu težinu kao univerzalnom sustavu za otkrivanju znanja u vremenskim serijama.

Generalno gledajući REFII modelom opisujemo krivulju nizom parametara, a skup tih parametara možemo kasnije procesuirati različitim tipovima algoritama u različitim problemskim prostorima s ciljem otkrivanja znanja.

Upravo procesi i problemski prostor obrade ovih parametara determiniraju primjenu.

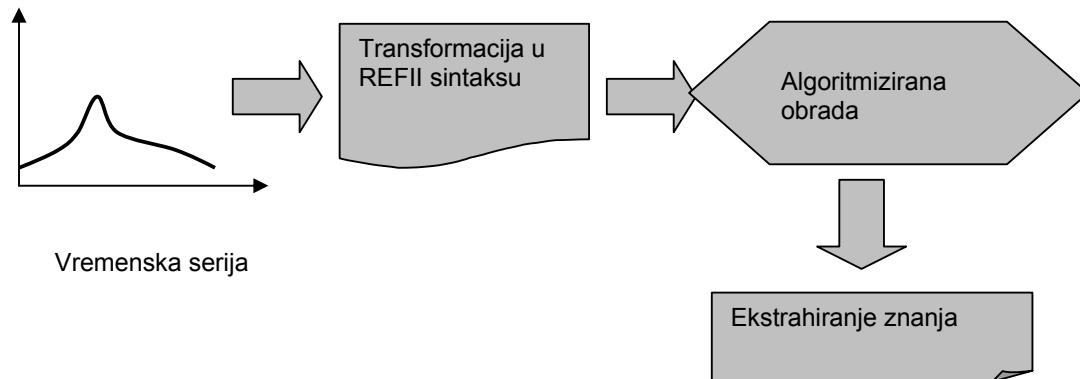
Ako promotrimo područja koje obuhvaća koncepcija REFII-modela , tada tu ubrajamo:

- Analizu kretanje trenda serije
- Analize cikličkih pojava unutar serije
- Analize sezonskih oscilacija unutar serije

- Pronalaženje odsječka vremenske serije koji korespondira sa određenom tržišnom pojavom
- Pronalaženje uzoraka unutar serije
- Pronalaženje i otklanjanje irelevantnih sekvenci iz serija
- Analiza korelatornih odnosa, kako između vremenskih serija tako i njenih odsječaka
- Autokorelatorna analiza vremenskih serija
- Predviđanje vremenskih serija
- Analiza prirasta i osjetljivosti vremenskih serija

Iz priloženog je vidljivo da je ovaj model primjenjiv na sve glavne tipove analiza u domeni vremenskih serija i šire.

Osnovna koncepcija primjene REFII modela svodi se na tri koraka. U prvom koraku vremenska se serija transformira u REFII sintaktički model. Drugi korak odnosi se na algoritmiranu obradu pokazatelja u formi REFII sintakse, dok se u trećem koraku ekstrahira znanje iz modela. Ove etape prikazane su na slici 1.8.



SLIKA 1.8. ETAPE U OTKRIVANJU ZNANJA POSREDSTVOM REFII MODELA

Upravo čitava lepeza algoritmiranih obrada koje se mogu primijeniti nad vremenskom serijom opisanom REFII sintaksom, jedna je od generalnih prednosti ove koncepcije, u odnosu na ostale koncepcije koje su procesno orijentirane.

Algoritmirana obrada može se izvršiti i posredstvom *black box* metodologije na način da algoritam bude dio nekog softverskog proizvoda. U tom slučaju naglasak će poslije transformacije u REFII sintaksu biti na pretprocesiranje podataka.

REFII model osim deskriptivne funkcije ima i procesne funkcije koje se očituju u algoritmiranim postupcima procjene faktora sigurnosti za svaki od tri elementa. Proračun faktora sigurnosti determiniran je tipom analize koju provodimo nad krivuljom, i stupnjem zahtijevane "preciznosti".

Ovi algoritmirani postupci za otkrivanje faktora sigurnosti dio su REFII modela. Karakteristika koja je ujedno i prednost ove koncepcije očituje se u modularnosti koja je determinirana problemskim prostorom i koja nudi modele rješenja unutar tog problemskog prostora primjenom raspoloživog instrumentarija unutar samog modela, ali isto tako procesuiranjem parametara posredstvom nadograđenih modula.

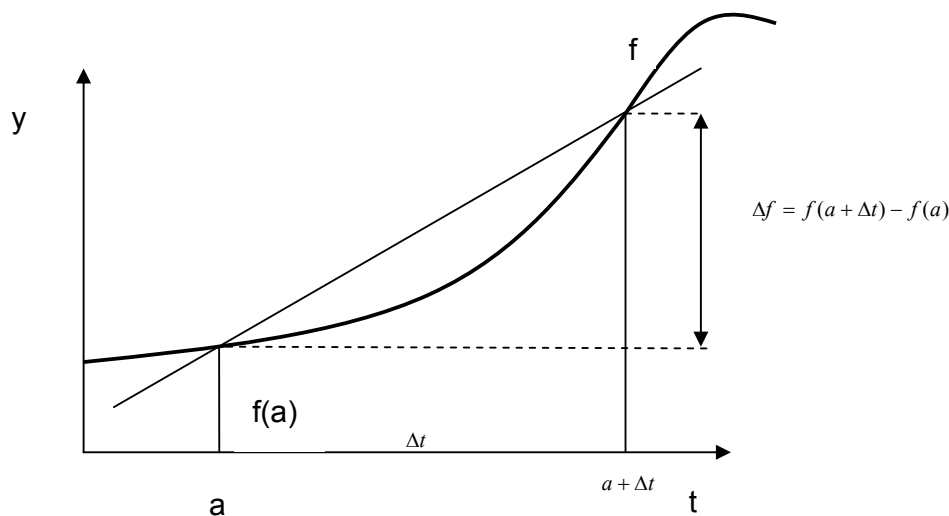
Upravo ova koncepcija pruža s jedne strane snažan instrumentarij orijentiran na sofisticiranu analizu vremenskih serija, a sa druge strane dobivamo vrlo otvorenu arhitekturu za nadogradnju modela rješenja iz domene različitih područja.

2. Temeljni koncept REF II modela

2.1. Matematička definicija modela

2.1.1. Promjena funkcijskih vrijednosti (prirast funkcije)

Kod kontinuirane funkcije f na intervalu I gdje je a pripada intervalu I . Ako je t takav broj da je $a + \Delta t$ pripada intervalu I onda broj $\Delta f = f(a + \Delta t) - f(a)$, a , $a + \Delta t$ nazivamo promjenom funkcijske vrijednosti kao što je vidljivo iz slike 2.1. [Javor, 1988], [Fanchi, 2000], [Mardešić, 1977]



SLIKA 2.1. PRIRAST KONTINUIRANE FUNKCIJE

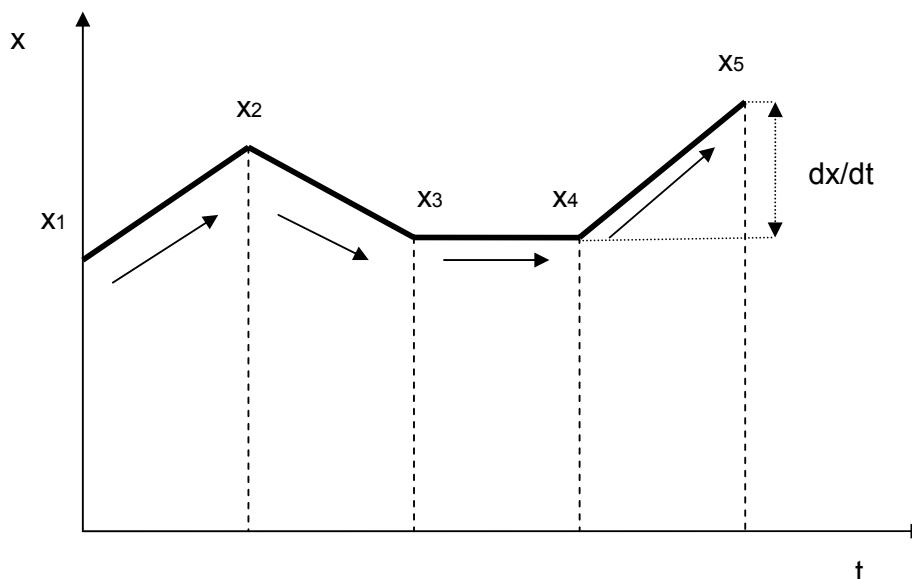
Derivaciju kontinuirane funkcije u točki t definiramo kao [Javor, 1988], [Fanchi, 2000], [Mardešić, 1977] :

$$f'(t) = \lim_{\Delta t \rightarrow 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} = \frac{\Delta f}{\Delta t}$$

Pomoću matematičke analize na kontinuiranim funkcijama možemo izračunati različite pokazatelje, poput ekstrema funkcije, i nul točaka.

Kako se u rudarenju podataka, i području analize vremenskih serija iz domene otkrivanja tržišnih zakonitosti susrećemo sa vremenskim nizovima predstavljenih nizom tabličnih vrijednosti, funkcije koje ulaze u proces data mining analize vremenskih serija su diskontinuirane funkcije kao. Naime, za razliku od kontinuiranih funkcija poput kvadratne funkcije, ili funkcije kosog hitca, tržišne pojave se na razini podataka promatraju u vremenskim razmacima i to najčešće jednolikim vremenskim

razmacima. Isto tako, poslovni događaji i tržišne aktivnosti odvijaju se u diskontinuiranim vremenskim intervalima. S obzirom na to, REFII model i modeli rudarenje podataka analiza vremenskih serija baziraju se na diskontinuiranim funkcijama, kao što je to i prikazano na slici 2.2.



SLIKA 2.2. PRIRAST DISKONTINUIRANE FUNKCIJE

REF komponentu unutar REFII modela definiramo kao :

Tablica 2.1. Definicija REF oznaka na temelju prirasta diskontinuirane funkcije

dx / dt	REF oznaka
= 0	E
>0	R
<0	F

Prilikom korištenja REFII modela u svjetlu klasične logike REF oznake označuju trendove prirasta.

Unutar REFII modela diskontinuirane vrijednosti se normiraju u intervalu <0,1>.

Pri čemu se za normiranje koristi min-max formula normiranja :

$$x_{\text{normirano}} = \frac{x_{\text{nenormirano}} - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$

Gdje je :

$x_{\text{normirano}}$ - normirana vrijednost (u intervalu između 0 i 1)

$x_{\text{nenormirano}}$ - vrijednost koja se normira

x_{max} – maksimalna vrijednost promatrane vremenske serije

x_{min} – minimalna vrijednost promatrane vremenske serije

Koeficijenti kutnog odklona računaju se prema formuli :

$$\frac{dx}{dt} = |x_normirano_t - x_normirano_{t-1}|$$

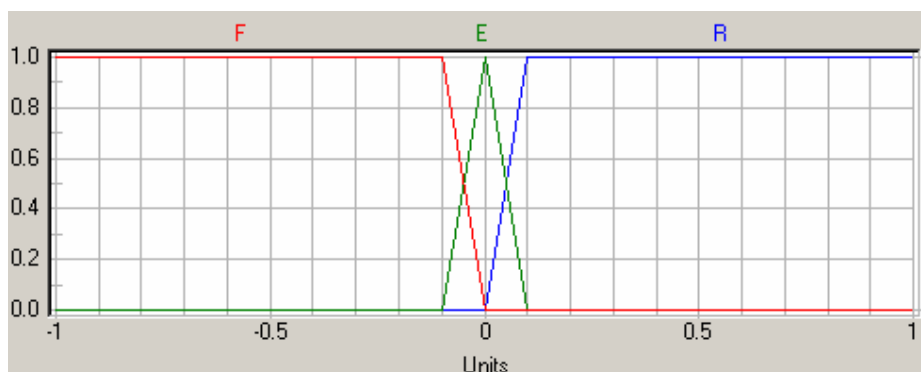
Na temelju koeficijenta kutnog odklona nadopunjujemo REF model kao što je to prikazano u tablici 2.2.

Tablica 2.2. Primjer nadopune REF modela sa na temelju koeficijenata kutnog odklona

REF dx / dt	REF oznaka	Koeficijent kutnog odklona dx / dt	Razred temeljem koeficijenta kutnog odklona (crisp)
= 0	E	0	Jednako
>0	R	0-0.2	Slab rast
		0.2-0.6	Srednji rast
		0.6-1	Oštar rast
<0	F	0-0.2	Slab pad
		0.2-0.6	Srednji pad
		0.6-1	Oštar pad

Prirast funkcije unutar REFII modela može se izraziti i kroz fuzzy logiku, pri čemu izrazitu logika možemo promatrati kao podskup fuzzy logike.

Funkciju pripadnosti kreiramo na temelju vrijednosti koeficijenta kutnog odklona prema slici 2.3. Vrijednosti na x osi predstavljaju normirane vrijednosti, i to za trend pada u intervalu od -1 do 0, a za trend rasta na intervalu od 0 do 1. Y os prikazuje stupanj pripadnosti REF (lingvističkih) varijabli.



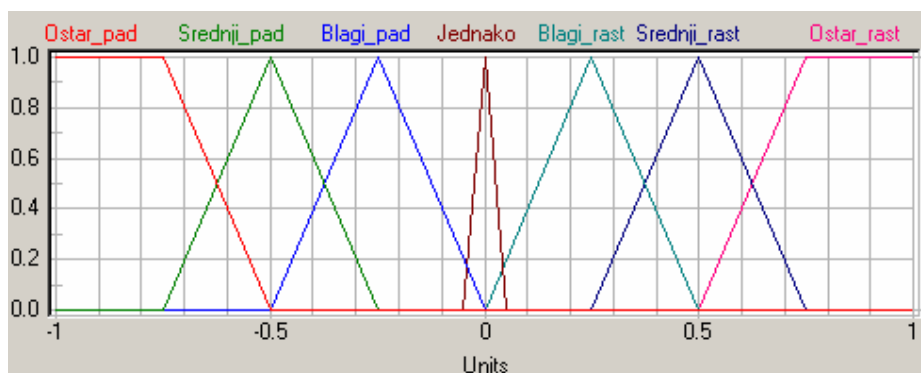
SLIKA 2.3. FUNKCIJA PRIPADNOSTI REF VARIJABLI

Tablica 2.3. Definicija točaka funkcije pripadnosti REF lingvističke varijable

Oznaka	Vrsta krivulje	Definicija točaka (x, y)		
F	linear	(-1, 1) (1, 0)	(-0.1, 1) (1, 0)	(0, 0)
E	linear	(-1, 0) (0.1, 0)	(-0.1, 0) (1, 0)	(0, 1)
R	linear	(-1, 0) (1, 1)	(0, 0) (0.1, 1)	(0.1, 1)

Kao što je vidljivo na temelju definicije lingvističke varijable, postoji stanovita tolerancija u definiranju nepromijenjenosti prirasta funkcije, a R,E i F oznake, deklarirane su kroz koeficijent kutnog otklona. Funkcije pripadnosti su definirane proizvoljno, te su se prijelomne točke mogle definirati i sa manjim/većim stupnjem tolerancije.

Ako želimo kreirati preciznije razrede u okviru fuzzy logike, kao što je to već prikazano za klasičnu "crisp" logiku, tada lingvističku varijablu kutni otklon možemo pomoću fuzzy logike definirati kao što je to prikazano na slici 2.4. na temelju istog obrasca značenja x i y osi.



SLIKA 2.4. FUNKCIJA PRIPADNOSTI VARIJABLI KUTNIH OTKLONA

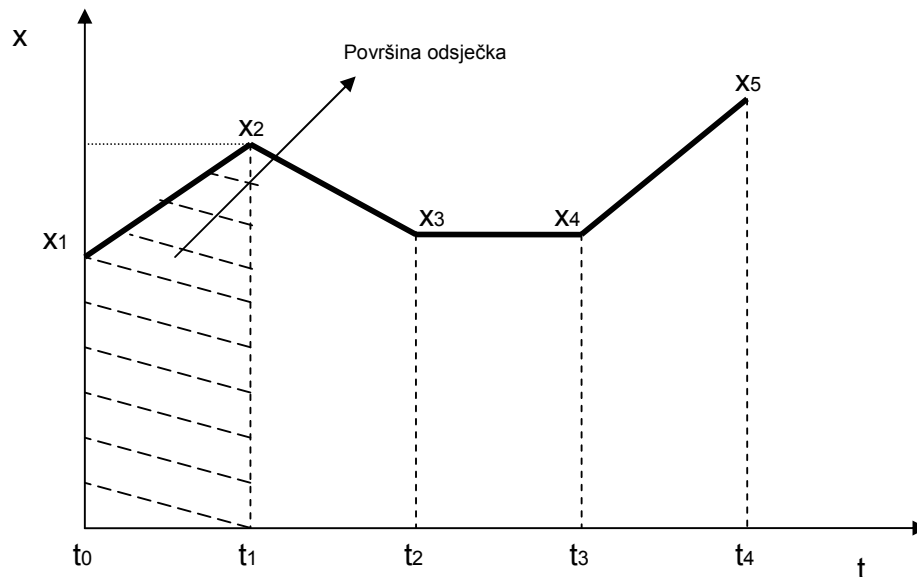
Tablica 2.4. Definicija točaka funkcije pripadnosti lingvističke varijable "kutni otklon"

Razred	Vrsta krivulje	Definicija točaka (x, y)		
Ostar_pad	linear	(-1, 1) (1, 0)	(-0.75005, 1) (1, 0)	(-0.5, 0)
Srednji_pad	linear	(-1, 0) (-0.25, 0)	(-0.75005, 0) (1, 0)	(-0.5, 1)
Blagi_pad	linear	(-1, 0) (0, 0)	(-0.5, 0) (1, 0)	(-0.25, 1)
Jednako	linear	(-1, 0) (0.05, 0)	(-0.05, 0) (1, 0)	(0, 1)
Blagi_rast	linear	(-1, 0) (0.5, 0)	(0, 0) (1, 0)	(0.25, 1)
Srednji_rast	linear	(-1, 0) (0.75, 0)	(0.25, 0) (1, 0)	(0.5, 1)
Ostar_rast	linear	(-1, 0) (1, 1)	(0.5, 0) (0.75, 1)	(0.75, 1)

Na temelju definiranih vrijednosti kategoriziraju se kutni otkloni, koji čine dva od tri temeljna elementa REFII modela.

2.1.2. Površina ispod krivulje

Kod kontinuiranih funkcija prilikom proračuna površine ispod krivulje koristimo integraciju funkcija. Kako je REFII model baziran na diskontinuiranim vrijednostima, za proračun površine ispod krivulje koristimo metode numeričke integracije.



SLIKA 2.5. NUMERIČKA INTEGRACIJA

Površina odsječka računa se prema formuli numeričke integracije metodom pravokutnika [Fanchi, 2000], [Mardešić, 1977] :

$$p = \frac{(x_n * \Delta t) + (x_{n+1} * \Delta t)}{2}$$

Ovako izraženu površinu u fizikalnom smislu možemo interpretirati kao prosječnu kvantitativnu vrijednost neke pojave između dva promatrana intervala.

Tako primjerice površina unutar odsječka uzorka između dva intervala (npr. dva mjeseca) gdje se promatra primjerice stanje tekućeg računa možemo interpretirati kao prosječan iznos sredstava na tekućem računu između ta dva perioda promatranja (npr. svibnja i lipnja)

Površina ispod krivulje na intervalu $\langle t_a.. t_n \rangle$ računa se kao zbroj površina odsječaka prema formuli:

$$P = \sum_{i=a}^n p_i$$

Pri čemu je vremenska distanca izražena kao Δt jednaka za sve vremenske odsječke u modelu i predstavlja mjeru vremena koja može biti izražena u mjesecima, danima, satima, minutama i slično.

Površinu na razini vremenskog odsječka također možemo normirati.

Kada u REFII model uvedemo površinu ispod odsječka kao treći element, dobijemo tablicu 2.5.

Tablica 2.5. Primjer nadopune REF modela sa površinom ispod krivulje

REF dx / dt	REF oznaka	Koeficijent kutnog otklona dx / dt	Razred temeljem koeficijenta kutnog otklona (crisp)	Površina ispod odsječka (normirana)	Razred površine
= 0	E	0	Jednako	0-0.3	Mala
				0.3-0.6	Srednja
				0.6-1	Velika
>0	R	0-0.2	Slab rast	0-0.3	Mala
				0.3-0.6	Srednja
				0.6-1	Velika
		0.2-0.6	Srednji rast	0-0.3	Mala
				0.3-0.6	Srednja
				0.6-1	Velika
		0.6-1	Oštar rast	0-0.3	Mala
				0.3-0.6	Srednja
				0.6-1	Velika
<0	F	0-0.2	Slab pad	0-0.3	Mala
				0.3-0.6	Srednja
				0.6-1	Velika
		0.2-0.6	Srednji pad	0-0.3	Mala
				0.3-0.6	Srednja
				0.6-1	Velika
		0.6-1	Oštar pad	0-0.3	Mala
				0.3-0.6	Srednja
				0.6-1	Velika

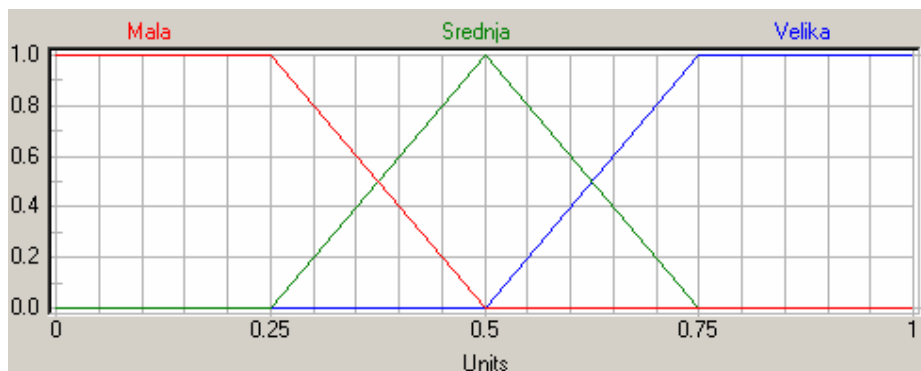
Istaknute vrijednosti u tablici 2.5. pokazuju primjer definicije elementarnog uzoraka na temelju empirijskih vrijednosti.

Tablica 2.5. je ujedno i jedinstven shematski prikaz elementarnih uzoraka koji čine REFII model, i koji povezuje otklon diskontinuirane funkcije sa površinom sa svrhom izgradnje elementarnih uzoraka.

Elementarni uzorci se mogu za potrebe analiza promatrati zasebno i ulaziti u obradu kao zasebne vrijednosti, ili se mogu promatrati kao niz spojenih vrijednosti.

Tablica 2.5. ujedno daje i obrazac za klasifikaciju odsječaka.

Ako odsječke želimo klasificirati uz pomoć fuzzy logike, tada je potrebno definirati lingvističku varijablu površina prema obrascu prikazanom na slici 2.6.



SLIKA 2.6. FUNKCIJA PRIPADNOSTI NORMIRANE VARIJABLE POVRŠINA

Tablica 2.6. Definicija točaka funkcije pripadnosti lingvističke varijable "površina"

Razred	Vrsta krivulje	Definicija točaka (x, y)		
Mala	linear	(0, 1)	(0.24998, 1)	(0.5, 0)
		(1, 0)		
Srednja	linear	(0, 0)	(0.24998, 0)	(0.5, 1)
		(0.75, 0)	(1, 0)	
Velika	linear	(0, 0)	(0.5, 0)	(0.75, 1)
		(1, 1)		

Da bi se kreirao ekvivalentan sustav klasifikacije u fuzzy logici kakav je kreiran u tablici 2.5. za crisp logiku potrebno je kreirati fuzzy ekspertni model prikazan na slici 2.7. baziran na definiranim lingvističkim varijablama



SLIKA 2.7. FUZZY SUSTAV KLASIFIKACIJE

Gdje su pravila u blokovima pravila sustava prikazana u tablici 2.7.

Tablica 2.7. Pravila klasifikacije uzoraka

IF		THEN	
Kutni_otklon	Povrsina	DoS	Klasifikacija
Ostar_pad	Mala	1.00	UzorakTipa1
Ostar_pad	Srednja	1.00	UzorakTipa1
Ostar_pad	Velika	1.00	UzorakTipa1
Srednji_pad	Mala	1.00	UzorakTipa1
Srednji_pad	Srednja	1.00	UzorakTipa1
Srednji_pad	Velika	1.00	UzorakTipa1
Blagi_pad	Mala	1.00	UzorakTipa1
Blagi_pad	Srednja	1.00	UzorakTipa1

IF		THEN	
Blagi_pad	Velika	1.00	UzorakTipa2
Jednako	Mala	1.00	UzorakTipa1
Jednako	Srednja	1.00	UzorakTipa1
Jednako	Velika	1.00	UzorakTipa2
Blagi_rast	Mala	1.00	UzorakTipa1
Blagi_rast	Srednja	1.00	UzorakTipa1
Blagi_rast	Velika	1.00	UzorakTipa2
Srednji_rast	Mala	1.00	UzorakTipa1
Srednji_rast	Srednja	1.00	UzorakTipa1
Srednji_rast	Velika	1.00	UzorakTipa2
Ostar_rast	Mala	1.00	UzorakTipa1
Ostar_rast	Srednja	1.00	UzorakTipa1
Ostar_rast	Velika	1.00	UzorakTipa2

2.1.3. Definicija elementarnih uzoraka

REFII model objedinjava trendove diskontinuiranih funkcija i površine na razini vremenskog odsječka, čineći pri tome elementarni uzorak. Elementarni uzorak je predstavljen sa tri osnovne vrijednosti :

- Oznakom trenda rasta (REF)
- Koeficijentom kutnog odklona koji možemo klasificirati u razrede bilo uz pomoć klasične crisp logike, bilo pomoću fuzzy logike
- Površinom ispod odsječka koji možemo klasificirati u razrede bilo uz pomoć klasične crisp logike, bilo pomoću fuzzy logike

Važno je napomenuti da klasifikacija u razrede unutar REFII modela ovisi o prirodi problema koji se rješava, te je u procesima analize moguće koristiti i izvorne vrijednosti koeficijenata kutnih odklona, kao i površina, što je vidljivo u empirijskom dijelu istraživanja disertacije.

Elementarni uzorci također mogu tvoriti kompleksnije strukture nizova uzoraka, te kao takvi mogu ulaziti u procese analize. Ovaj postupak je vidljiv u dijelu empirijskih istraživanja koji se bave otkrivanjem uzoraka, otkrivanjem sličnosti i motiva.

Elementarni uzorak definiran kroz REFII model, njegov je temeljni element.

2.2. Definicija i opis REF metode

Osnovna namjena REF modela jest upravo dijagnostika oblika krivulje, što može biti temelj za postupke poput prepoznavanja sezonskih oscilacija unutar vremenskih serija.

Ovaj model je nastao kao reakcija na krutost standardnih statističkih metoda, koje se bave analizom sezonskih oscilacija, a zametak je REFII modela . U REFII modelu REF model je segment koji je zadužen za detekciju oblika krivulje, ne ulazeći pri tome u procjenu jačine trendova od kojih je krivulja konstituirana, niti u detalje kvantitativnih aspekata vremenskih odsječaka, što je zadatak ostalih komponenti REF II modela.

Zadatak REF modela prvenstveno je dijagnostika oblika krivulje u određenim segmentima vremenske serije. Njegov povijesni nastanak vezan je uz otkrivanje sezonskih oscilacija u vremenskim serijama, pa će u svjetlu te problematike biti ilustrativno prezentiran.

Razlog njegovog uključivanja kao dijela REFII modela, usprkos tome što se trendovi mogu procijeniti na temelju koeficijenata kutnog otklona proizlazi iz činjenice, da je on podesan za brzu i efikasnu procjenu oblika krivulje na velikom broju elemenata vremenske serije.

To znači da na brz način u velikim serijama možemo dobiti osnovne pokazatelje o postojanju generalnih trendova, bilo sličnosti, bilo uzoraka a da pri tome ne vršimo proračune kutnih otklona i formiramo razrede otklona. Ovaj podmodel može znatno uštedjeti vrijeme i pružiti elementarne informacije o obliku krivulje na jednoj generalnoj razini, uz znatno kraće vrijeme obrade dugih vremenskih serija.

Matematički instrumentarij, koji je korišten prilikom razvoja ove metode svodi se na teoriju o monotonosti funkcija, koja je ujedno i okosnica cijele metode.

Kao popratni elementi modela korišteni su izračuni stupnja pouzdanosti, standardne devijacije i ukupne greške.

Osnovna ideja modela proizlazi iz matematičke teorije o monotonosti funkcije [Javor, 1988] .

Funkcija

$$f : D \rightarrow R$$

$$D \subseteq R$$

monotono raste na D ako za svaki

$$x_1, x_2 \in D$$

vrijedi

$$(x_1 < x_2) \Rightarrow f(x_1) \leq f(x_2)$$

ili monotono pada na D ako za svaki

$$x_1, x_2 \in D$$

vrijedi

$$(x_1 < x_2) \Rightarrow f(x_1) \geq f(x_2)$$

Gdje je D domena a R je skup realnih brojeva.

U situacijama kada je funkcija u formi vremenske serije, a dobivena je iz baze podataka, možemo izračunati parcijalni trend iz izraza:

$$f = e_i - e_{i-1} \quad (2.1)$$

gdje

f - predstavlja trend neke pojave

e- predstavlja vrijednost neke pojave u periodu vremena (sat, dan, mjesec...)

i – predstavlja vremenski pomak

Ako je $f > 0$, to znači da funkcija raste.

Ako je $f < 0$, to znači da funkcija pada.

Ako je $f = 0$, funkcija ne raste niti ne pada, ima istu vrijednost za vremenski pomak

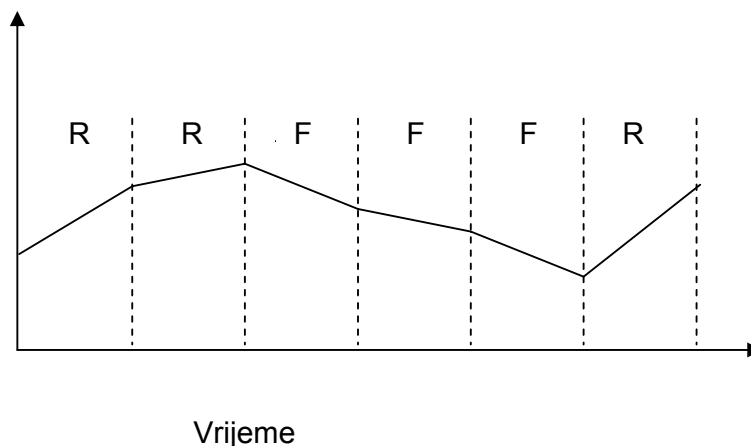
Rastući trend ($f > 0$) dijela funkcije proizašlog iz ove formule na određenom odsječku možemo prezentirati oznakom R.

Padajući trend ($f < 0$) dijela funkcije proizašlog iz ove formule na određenom odsječku možemo prezentirati oznakom F.

Neutralan trend ($f = 0$) dijela funkcije proizašlog iz ove formule na određenom odsječku možemo prezentirati oznakom E.

Upotrebom ove notacije, kompletnu vremensku seriju možemo prezentirati sa ove tri oznake.

Za funkciju (R,R,F,F,F,R) možemo reći da raste za dvije vremenske, da nakon toga pada za tri vremenske jedinice i raste za jednu vremensku jedinicu.



SLIKA 2.8. PRIKAZ VREMENSKE SERIJE KROZ REF MODEL

Slika 2.8. prikazuje kako REF model transformira vremensku seriju u svoju notaciju.

Ovakva transformacija vremenske serije i matematički instrumentarij koji će biti prikazan može biti korisna za različite tipove analiza vremenskih serija, koje ovise o zadanim ciljevima. Ilustrativni primjer primjene REF modela bit će prikazana metodologija otkrivanja sezonskih oscilacija na temelju oblika krivulje.

Za tu namjenu potrebno je vremensku seriju razlomiti na vremenske odsječke (npr. po tjednima), transformirati je na osnovu prikazane notacije te provoditi analize metodom usporedbe.

2.2.1. REF model i otkrivanje sezonskih oscilacija

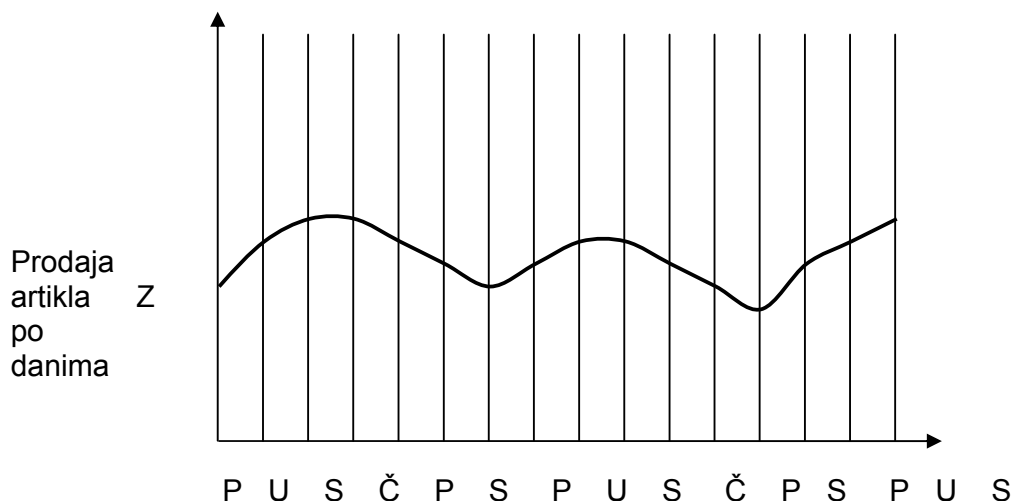
Ako primjerice promatramo prodaju artikla Z kroz period vremena, te želimo saznati da li pojedini dani u tjednu bilježe sezonske oscilacije u prodaji tog artikla, u prvom koraku formiramo vremensku seriju, čije su vrijednosti na koordinati Y vrijednost prodane robe Z, a X os prikazuje dane.

Ovako formiranu vremensku seriju transformiramo u REF oznake i podijelimo na tjedne.

Tek sada možemo prijeći na korak koji se svodi na usporedbe uzoraka.

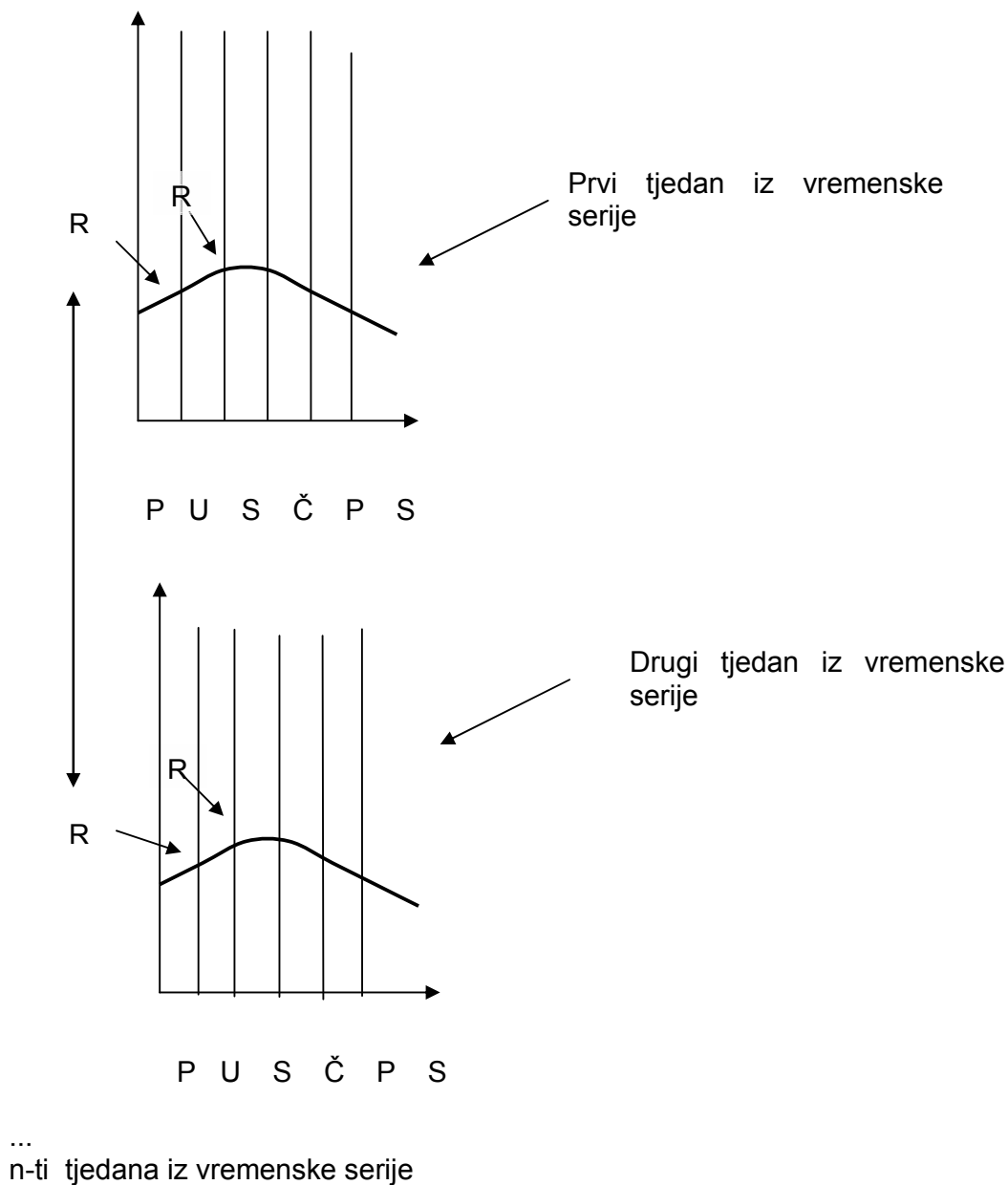
Ako postoji očigledna zakonitost ti će se uzorci podudarati sa vjerojatnošću 1, no u praksi možemo prihvatiti neku tvrdnju sa pouzdanošću između 0,7 i 1.

Slika 2.9. prikazuje vremensku seriju prodanosti artikla Z po danima.



SLIKA 2.9. VREMENSKA SERIJA PRODAJE ARTIKLA Z PO DANIMA

Ovako formiranu vremensku seriju reprezentiramo REF oznakama i radi usporedbe razlomimo je po tjednima, kako bi bila spremna za analizu i to na način prikazan na slici 2.10.



SLIKA 2.10. LOMLJENJE VREMENSKIH SERIJA U MANJE VREMENSKE JEDINICE

Iz slike 2.10. vidljivo je da se uspoređuju elementi vremenske serije (u ovom slučaju dani unutar svih tjedana) u cilju otkrivanja zakonitosti.

U slučaju da se određeni uzorci podudaraju, možemo zaključiti o postojanju određenog modela ponašanja (sezonske oscilacije), ne uzimajući pri tome u obzir kvantitativni aspekt pojave.

Kao što je spomenuto, nerealno je za očekivati da će se uzorci podudarati sa vjerojatnošću 1.

Zato se koristi stupanj pouzdanosti, o kome ovisi prihvaćanje ili odbijanje tvrdnje.

Stupanj pouzdanosti računamo izrazom:

$$c = \frac{\sum_{i=1}^{n-1} \frac{p_i}{mp}}{n-1} \quad (2.2)$$

Gdje je:

c – stupanj pouzdanosti (0 -1)

p – broj jednakih elemenata sa bazičnim uzorkom

mp – broj elemenata u bazičnom uzorku

n – broj promatranih uzoraka izraženih u REF formatu (n-1 zbog bazičnog uzorka)

Stupanj pouzdanosti je mjera kojom određujemo pouzdanost modela, odnosno mjera o kojoj ovisi prihvaćanje/neprihvaćanje tvrdnje.

Bazični uzorak je izabrani uzorak iz osnovnog skupa uzoraka (npr. jedan od tjedana prezentiran REF oznakama), koji je usporediv sa ostalim uzorcima, također prezentiranim u REF formatu.

Ako je stupanj pouzdanosti veći ili jednak graničnom stupnju određenom sa strane analitičara, prihvaća se tvrdnja o postojanju sezonske oscilacije u vremenskoj seriji.

Bazični uzorak možemo selektirati na temelju slučajnog odabira, što je lošija strategija, na temelju "maksimalne sličnosti", ili pak na temelju primjene formule stupnja pouzdanosti sa svakim od postojećih uzoraka. Strategija biranja ili kreiranja proizvodnog bazičnog uzorka preporuča se u situacijama kada želimo saznati da li se neka pojava ponaša prema unaprijed zacrtanom obrascu koji predstavlja bazični uzorak. Isto tako, bazični uzorak možemo izabrati iz osnovnog skupa uzoraka.

Situacija procjene obrasca ponašanja u kome je bazičan uzorak izabran iz osnovnog skupa uzoraka bit će obrađena u daljnjem tekstu.

Ako smo zainteresirani jedino za određivanje postojanja određenog grubog modela ponašanja u vremenu (između ostalog i sezonskih oscilacija), ovakav model nam je dostatan, međutim ako želimo otići korak dalje i kvantificirati odstupanja između uzoraka, koristimo narednu formulu, koja svoje temelje nalazi u statističkoj formuli za određivanje standardne devijacije.

$$Er = \sum_{j=1}^{n-1} \sqrt{\frac{\sum_{j=1}^j (X_j - X_{pj})^2}{j}} \quad (2.3)$$

Gdje

Er – izražava grešku cjelokupnog modela

n – broj promatranih vremenskih serija u REF uzorku (n-1 zbog bazičnog uzorka)

j – broj elemenata u promatranim serijama

X_j – vrijednost j-tog REF člana dobivenog na temelju formule 2.1

X_{pj} – vrijednost j-tog REF člana iz bazičnog uzorka dobivenog na temelju formule 2.1

U narednoj tablici vidljivi su odnosi ovih varijabli:

Tablica 2.8. Indeksirane varijable

	Elementi uzorka	Elementi bazičnog uzorka
Uzorak 1	$X_{11(j)}; X_{12(j)}..X_{1m(j)}$	$X_{11(pj)}; X_{12(pj)}..X_{1m(pj)}$
Uzorak 2	$X_{21(j)}; X_{22(j)}..X_{(j)2m(j)}$	$X_{11(pj)}; X_{12(pj)}..X_{1m(pj)}$
Uzorak 3	$X_{31(j)}; X_{32(j)}..X_{3m(j)}$	$X_{11(pj)}; X_{12(pj)}..X_{1m(pj)}$
Uzorak ,,,	$X_{...1(j)}; X_{...2(j)}..X_{...m(j)}$	$X_{11(pj)}; X_{12(pj)}..X_{1m(pj)}$
Uzorak n	$X_{n1(j)}; X_{n2(j)}..X_{nm(j)}$	$X_{11(pj)}; X_{12(pj)}..X_{1m(pj)}$

Indeksirani elementi predstavljaju vrijednosti dobivene iz formule 2.1

Na ovaj način možemo kvantificirati dobivene vrijednosti i dobiti konkretnije informacije o standardnoj grešci.

Navedena mjera je korisna za proračun greške modela, ali na žalost to je indikativna mjera koja kao rezultat izračuna nudi indikator. U slučajevima kada dobijemo informaciju o odstupanju, idealno bi bilo da možemo saznati uslijed čega se je pojavilo odstupanje, odnosno gdje je razlog odstupanju. Gledano iz perspektive vremenskih serija, bilo bi poželjno saznati koji odsječak vremenske serije odstupa kvantitativno, usprkos sličnom obliku krivulje. Odgovor na ovo pitanje ne možemo dobiti posredstvom REF modela, na to pitanje odgovor možemo dobiti primjenom preostala dva modela koji također pripadaju REFII koncepciji.

U svrhu ilustrativnog prikaza REF modela izvršit će se proračun postojana sezonskih oscilacija za naredne vremenske serije, koje su razlomljene u komparativne vremenske odsječke (serija 1, serija 2), gdje t može predstavljati dane u tjednu, mjesece i slično.

Tablica 2.9. Vremenske serije za REF model

Vrijeme (t)	1	2	3	4	5
Serija 1	4	5	7	9	11
Serija 2	4	9	7	9	11
Serija 3	4	5	8	9	11
Serija 4	3	7	8	9	12
Serija 5	4	5	7	6	17

Kada vremenske serije transformiramo u REF notaciju (npr. prvi element serije1 u REF notaciji dobiven je na način da $5-4 > 0$ što implicira R), tada dobivamo tablicu 2.10.

Tablica 2.10. Vremenske serije u REF notaciji

	Odsječak 1	Odsječak 2	Odsječak 3	Odsječak 4
Seriya 1	R	R	R	R
Seriya 2	R	F	R	R
Seriya 3	R	R	R	R
Seriya 4	R	R	R	R
Seriya 5	R	R	F	R

Prilikom biranja bazičnog uzorka možemo krenuti od pretpostavke da primjerice želimo procijeniti da se sve vremenske serije ponašaju kao serija 1.

Na temelju formule 2.2 dobiven je stupanj pouzdanosti 0.875 $((0.75+1+1+0.75)/4)$.

Grešku modela možemo izračunati na temelju formule 2.3 kao što je to prikazano u tablici 2.11.

Tablica 2.11. Proračun greške modela

						SQRT(SUM($X_j - X_{pj}$)))
$(X_1 - X_p)^2$	0	16	0	0	0	1,79
$(X_2 - X_p)^2$	0	0	1	0	0	0,45
$(X_3 - X_p)^2$	1	4	1	0	1	1,18
$(X_4 - X_p)^2$	0	0	0	9	36	3,00
				Er		1,60

Ovakav pristup preporuča se u situacijama, kada na temelju bazičnog uzorka želimo procijeniti da li se i ostale vremenske serije ponašaju po tom obrascu. Ova metodologija izračuna je bliska problemima u kojima želimo dobiti odgovor na pitanje da li se neka pojava ponaša po obrascu X, i sa kolikim stupnjem pouzdanosti. No na temelju REF modela za procjenu čistih sezonskih oscilacija, preporuča se postupak računanja frekvencija pojavnosti REF oznaka s obzirom na vremenski odsječak.

Ova metodologija detaljnije je razrađena u poglavlju 6.3. na temelju REFII modela. U slučaju REF modela procjena postojanja sezonskih oscilacija za dani primjer prikazana je u tablici 2.12. izračunata na temelju frekvencije pojavnosti u odsječcima iz tablice 2.10.

Tablica 2.12. Sezonske oscilacije

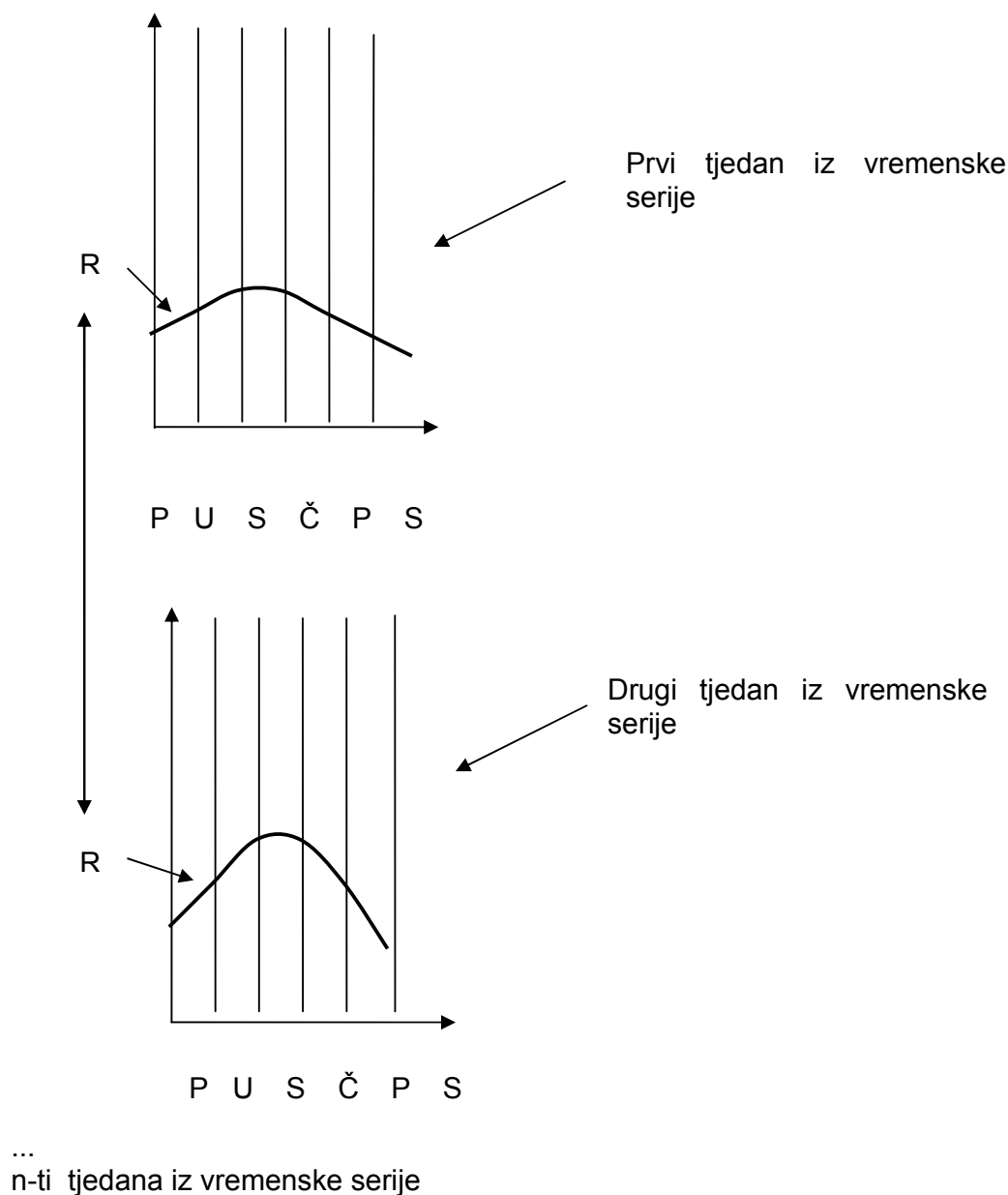
Prvi odsječak	Drugi odsječak	Treći odsječak	Četvrti odsječak
R=5/5=1	R=4/5=0.8 ; F=1/5=0.2	R=4/5=0.8 ; F=1/5=0.2	R=5/5=1

Ovisno o zahtijevanom stupnju pouzdanosti prihvaća se tvrdnja o postojanju sezonskih oscilacija. Tako sa pouzdanošću 1 možemo tvrditi da postoji oscilacija rasta za prvi i četvrti odsječak. Nedostatak ovakvog pristupa proračuna sezonskih oscilacija između ostalog odnosi se na nemogućnost sagledavanja sezonskih oscilacija kao povezane cjeline odsječaka u nizu, već se sezonske oscilacije promatraju na razini svakog pojedinačnog odsječka. Poboljšani model koji rješava taj problem razrađen je u poglavlju 6.3., te on obuhvaća cjelokupan REFII model.

Naredni primjer ilustrira važnost kvantitativnog aspekta koji je vezan uz pojam vremenske serije. Prilikom analize podataka može doći do prihvaćanja tvrdnje o sezonskoj oscilaciji koja je formalizirana tvrdnjom AKO dan= ponedjeljak ONDA trend povećanja prodaje proizvoda X sa stupnjem pouzdanosti 0.92.

Usprkos tome, kao što je već spomenuto može doći do rezultata velike relativne greške modela.

Razlog tome leži u činjenici da, usprkos otkrivenom pravilu, postoje odstupanja u kvantitativnim pokazateljima, što je ilustrirano narednom skicom.



SLIKA 2.11. UZROCI VELIKE STANDARDNE GREŠKE USPRKOS PRIHVAĆANJU MODELA

Slika 2.11. prikazuje uzroke velike standardne greške modela usprkos visokom stupnju pouzdanosti modela.

Proračun standardne greške daje nam mogućnost kvantifikacije te pojave, na način koji nas navodi na daljnje analize.

Upravo ovdje leži slabost korištenja isključivo REF modela u analizama vremenskih serija, jer on ne daje jednoznačan opis vremenske serije. S toga su nastala daljnja dva elementa o kojima će kasnije biti govora, a koji zajedno sa REF modelom čine REF II model.

U slučaju dijagnoze velike standardne greške modela, možemo krenuti u daljnje ispitivanje gdje je to odstupanje najočitije, odnosno krenuti u detaljniji stupanj analize sezonske oscilacije.

Koristeći se navedenim primjerom o dinamici prodaje artikla kroz period vremena, može se detaljnijim analizama doći do zaključka da ta pojava prati trend sezonskih oscilacija, ali da su prva dva tjedna u mjesecu siječnju, ožujku i studenom, s perspektive kvantitativnog aspekta pojave, mnogo manjeg kvantitativnog intenziteta, nego u istim komparativnim razdobljima u ostalim mjesecima.

Ovo je naravno samo jedan od mogućih razloga postojanja velike standardne greške usprkos prihvatanja tvrdnje o postojanju pravila u vremenskoj seriji koje je izgenerirano REF modelom.

REF model prvenstveno daje odgovore na pitanja o postojanju pravilnosti u vremenskim serijama.

Proračuni standardnih grešaka omogućuju nam dijagnosticiranje većih kvantitativnih odstupanja, što može biti podloga za ulazak u naredni model REF II modela.

Iako je cjelokupna primjena modela prikazana kroz otkrivanje sezonskih oscilacija, to nije jedino područje u kome je REF model efikasan. Problematika sezonskih oscilacija dana je kao ilustrativni primjer primjene modela, i u skladu s tim prikazana je formula za računanje greške modela i standardne greške modela.

Ovaj podmodel REF II modela sposoban je otkrivati sezonske oscilacije u vremenskoj seriji, otkrivati epizode, procjenjivati inverzne oscilacije, pri čemu moramo prilagoditi formule za proračun greške modela i standardne greške.

Ovu metodu prvenstveno moramo promatrati kao dio REFII modela, i segment zadužen za manipulacije oblikom krivulja, te izvođenje zaključaka na osnovu oblika krivulje.

Algoritamski postupak prikazan u ovom poglavlju, treba promatrati prvenstveno sa perspektive ilustrativnosti, jer je on povijesna preteča REFII modelu, te je upravo zbog čitavog niza otvorenih pitanja prilikom njegove primjene u praksi razvijen REFII model. Otvorena pitanja odnosila su se na problematiku optimalnog izbora bazičnog uzorka, problematike kvantifikacije i voluminoznosti. Prikazani primjer sa otkrivanjem sezonskih oscilacija prikazan je kako bi ilustrirao širinu primjene podmodela i mogućnosti kontingencijskog pristupa prilikom rješavanja određenih tipova problema.

2.2.2 Algoritamska interpretacija primjene REF modela

Iako generalno gledajući, postoji niz nedostataka koji se odnose na preciznost REF modela, on može biti od velike koristi prilikom dijagnostike obriša promatranih empirijskih krivulja. Ova značajka posebice dolazi do izražaja kod vremenskih serija od nekoliko milijuna elemenata, gdje želimo dobiti osnovnu informaciju o obliku i karakteristikama vremenske serije, te o razradi daljnje strategije analize.

Analitičar mora imati fleksibilan alat, koji se temelji na ovoj metodi, jer upravo vremenski razredi, koji sadrže numeričke vrijednosti, primjerice prodaje, mogu biti od ključne važnosti za otkrivanje pravilnosti.

Fleksibilan alat, koji se temelji na ovoj metodi, morao bi nam omogućiti i opciju mijenjanja stupnja granulacije podataka posredstvom funkcija sumiranja, prosjeka, broja elemenata (SQL- SUM(), AVG(), COUNT()) i slično.

Kod nedostajućih vrijednosti unutar osnovnog seta podataka, ako se već mora pribjeći nekoj od metoda za proračun izgubljenih vrijednosti, smatram da je za to najpodesnija neuronske mreža.

Za pripremu podataka prije procesa obrade najpodesniji su SQL upiti na bazu, koji mogu posredstvom GROUP BY klauzule, formirati željeni stupanj zrnatosti.

Algoritam pisan u pseudo kodu, za potrebe transformacije vremenske serije u ŽREF model glasi:

m= broj elemenata vremenske serije

brojač = 1

ZA $m > \text{brojač}$

$$f = e_{i-1} - e_{i-2}$$

AKO JE $f > 0$

$$\text{niz} = \text{niz} + R$$

AKO JE $f < 0$

$$\text{niz} = \text{niz} + F$$

AKO JE $f = 0$

$$\text{niz} = \text{niz} + E$$

brojač=brojač+1

KRAJ

2.2.3. REF model – područje efikasnosti / neefikasnosti

Kao što je već navedeno, u svojoj ranoj fazi ovaj modela bio je razvijen s ciljem otkrivanja sezonskih oscilacija unutar vremenskih serija.

Praktičnom primjenom, a kasnije i integracijom ovog modela u REFII model on dobiva na širini primjene. REF model treba prvenstveno promatrati kao dio REFII modela, što implicira samo jednu od tri karike u lancu cjelokupnog sustava. Opisana koncepcija prvenstveno je zadužena za definiciju oblika krivulje, i podatkovno gledajući krivulja transformirana posredstvom opisanog modela predstavlja niz R,E,F oznaka. Kako

ćemo toj podatkovnoj strukturi pristupiti, ovisi o konkretnom problemu, pa se u skladu s tim formiraju algoritmizirani postupci koji u sebi sadrže modificirane formule za procjenu stupnja pouzdanosti i greške modela, ako što je to bio slučaj na konkretnom ilustrativnom primjeru.

Ovakva podatkovna struktura polazišna je osnovica, za obradu svih onih problema koji se mogu uspješno razriješiti analizom oblika krivulje. Primjer koji se je odnosio na otkrivanje sezonskih oscilacija, ilustrativan je primjer kako na osnovu oblika krivulje, gdje imamo vremensku seriju transformiranu REF modelom otkriti sezonske oscilacije.

Prikazana metodologija otkrivanja sezonskih oscilacija rudimentarni je oblik otkrivanja znanja iz vremenskih serija, te je u kontekstu REFII modela prikazana isključivo u ilustrativne svrhe, iako je ova metodologija primjenjivana u praksi.

U okviru REFII modela kao cjeline prikazani su fleksiblniji i moćniji algoritmi za rješavanje problematike otkrivanja sezonskih oscilacija.

Primjenjujući samo REF model moguće je provoditi i analizu događajnosti, kao i pretragu uzoraka.

Osnovni nedostatak oslanjanja isključivo na REF model je njegova nepreciznost i "grubost" u definiciji vremenskih odsječaka.

Jedna od glavnih karakteristika REF modela svodi se na mogućnost generiranja tvrdnji primjenom adekvatnih algoritama nad modelom u formi AKO – ONDA, što je veoma interesantno ako problematiku analize vremenskih serija promatramo iz perspektive generiranja znanja.

Pozitivne strane modela:

- Sposoban je izgenerirati dobivene rezultate u obliku AKO- ONDA
- Orijentiran je prvenstveno pronalaženju zakonitosti
- Bliži je koncepciji metoda data mininga nego klasične statističke metode
- Fleksibilniji je jer uvažava stanovita odstupanja od bazičnih uzoraka
- Generira tvrdnje uz faktor sigurnosti
- Može "procijeniti" ukupnu grešku modela

Negativne strane REF modela ako se ne koristi u sklopu REFII modela:

- Nepreciznost
- Nejednoznačno opisuje krivulju
- Nema adekvatnu kvantificirajuću komponentu, pomoću koje bi se moglo ući u detaljnije analize krivulje

2.3. Proračun površine ispod krivulje

Opisani REF model može nam pomoći koda analize vremenskih serija kada nas ne interesira kvantitativni aspekt vremenske serije. Pokazatelj koji je na tragu kvantifikacije a pripada REF modelu je proračun greške modela. Njegov nedostatak očituje se u tome što je to indikativni pokazatelj koji se odnosi na cjelokupni model. Na osnovu tog pokazatelja možemo procijeniti pouzdanost modela sa kvantitativnog

aspekta, ali je to vrlo gruba mjera kada želimo spoznati dublje kvantitativne aspekte. Ako na primjer želimo vremenske odsječke analizirati s obzirom na kvantitativnu komponentu, u tom slučaju nam pomaže proračun površine ispod krivulje. Metodologija granulacije i pretprocesiranja podataka opisanu u prvom poglavlju, također je nužan predkorak primjene ove metode.

Osnovna ideja izračuna površine ispod krivulje oslanja se na integralni račun, i to na korištenje određenog integrala.

Za naše potrebe izračuna površine ispod krivulje, s obzirom na transformaciju iste u niz pravaca, metoda trapeza će davati vrlo precizne rezultate.

Površina ispod krivulje može se računati metodom numeričke integracije parcijalno za pojedine odsječke, kao i globalno za dani interval, što naravno ovisi o cilju analize.

Slika 2.5. prikazuje integraciju diskontinuiranih funkcija koja je izražena formulom :

$$p = \frac{(x_n * \Delta t) + (x_{n+1} * \Delta t)}{2}$$

Površina ispod krivulje na intervalu $\langle t_a, t_n \rangle$ računa se kao zbroj površina odsječaka prema formuli:

$$P = \sum_{i=a}^n p_i$$

Pri čemu je vremenska distanca izražena sa Δt jednaka za sve vremenske odsječke u modelu i predstavlja mjeru vremena koja može biti izražena u mjesecima, danima, satima, minutama i slično.

Ta površina može korespondirati sa odsječkom krivulje, ili kompletnom krivuljom. Ovu kategoriju određuje interval (a,b) .

Kada govorimo o kompatibilnosti unutar modela gledano sa perspektive REF modela i proračuna površine ispod krivulje, tada je bitno odrediti jedne vremenske intervale na koje ćemo razlomiti krivulju kako bi model bio transparentan.

S obzirom na to da smo vremensku seriju prikazali nizom razlomljenih pravaca, metoda numeričke integracije bazirana na trapezu najpodesnija je za izračunavanje površine ispod krivulje.

Rezultat izračuna za krivulju ili njene segmente je numerička vrijednost koja reprezentira vrijednost površine ispod krivulje. Ako na primjer vremenska serija predstavlja plog sredstava na tekući račun površina ispod krivulje u određenom intervalu, odnosno podintervalu predstavlja sumu prosječnih stanja između dva perioda, položenih sredstava na tekući račun. Što je vremenski interval duži, to je veća mogućnost pogrešne prognostike o jednakosti krivulje na osnovu površine ispod krivulje. Upravo iz tog razloga ova metoda može biti pomoćni element kod analize vremenskih serija, a nikako glavni element procjene. Primjerice, površine između dva odsječaka vremenskih serija mogu biti jednake, ali to nužno ne znači da je i oblik krivulje jednak, što upućuje na nejednakost vremenskih serija. Cilj analize determinira stupanj preciznosti i intervale pouzdanosti prihvatanja, odnosno odbijanja hipoteze.

Površina ispod krivulje je u svakom slučaju dragocjeni pokazatelj , koji daje precizniju sliku kada je riječ o vremenskoj seriji.

Ako promotrimo sliku 2.10., gdje se isključivo na osnovu REF modela može donijeti sud o jednakosti krivulja, odnosno postojanju sezonske oscilacije, izračunom greške modela dolazimo do pokazatelja koji upućuje na signifikantno odstupanje. Gledajući problematiku prvenstveno kroz prizmu REF modela, ovdje razmatranje prestaje. Konstatirali smo kvantitativna odstupanja, ali neosporno je da postoji pravilnost u obliku. Ako stvar promatramo šire i primijenimo REFII koncepciju, tada tražimo kvantitativne aspekte pojava na način da računamo površine ispod krivulja. Površinu možemo računati kroz cjelokupni interval, pri čemu nemamo mogućnost dubljeg ulaženja u uzroke odstupanja. U tom slučaju interval površine ispod krivulje računamo kao:

$$P(i)=P(a)...P(b)$$

Gdje i označava indeks razlomljene vremenske serije, dok su $a...b$ predstavlja interval razlomljene vremenske serije.

Na osnovu ovog proračuna možemo komparirati ukupnu vrijednost neke pojave uspoređujući površine ispod krivulja.

$$P(1)=P(2)=P(3)=...=P(n)$$

Precizniji način je izračunati parcijalne površine, i tražiti parcijalne uzroke kvantitativnih odstupanja uzoraka, jer tek na taj način dobivamo analitički instrumentarij za preciznu dijagnostiku odstupanja. To činimo narednom formulom:

$$P(i(x))=<P(ia(x))...P(ib(x))>$$

Gdje $i(x)$ označava indeks podintervala razlomljene vremenske serije, dok su $ia(x)...ib(x)$ podintervali razlomljene vremenske serije.

Nakon ovakvog izračuna parcijalne dijelove vremenske serije možemo uspoređivati sa ostalim parcijalnim dijelovima vremenske serije.

Generalna strategija usporedbe vremenskih serija mogla bi se svesti na primjenu REF modela, i nakon toga procjenu pouzdanosti modela. Ako postoji odstupanje, pomoću metode numeričke integracije možemo izračunati parcijalne površine ispod krivulje, i tražiti zakonitosti u odstupanju.

2.3.1. Algoritamska interpretacija numeričke integracije

Numerička integracija u REFII modelu ima ulogu proračuna površine ispod krivulje. Kao što će biti vidljivo kroz tekst, ovaj podsegment modela uvijek se koristi u kombinaciji sa ostalim elementima modela.

Korištenje ove metode svodi se na računanje površina svih podsegmenata krivulje koja je podijeljena u n jednakih odsječaka. Jednostavna algoritamska interpretacija ove metode prezentirana pseudo kodom glasila bi :

m = broj odsječaka vremenske serije

brojač = 1

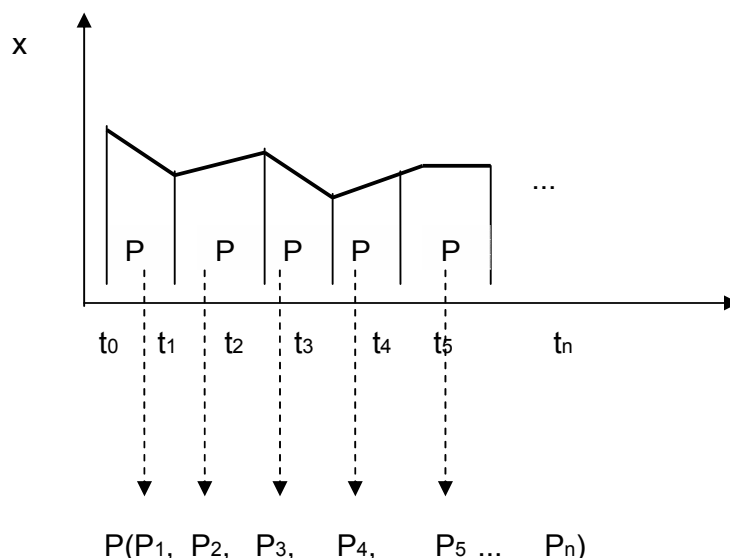
ZA $m > \text{brojač}$

*$niz(\text{brojač}) = ((x_1 * t_1) + (x_2 * t_2)) / 2$*

brojač = brojač +1

KRAJ

U podatkovni niz punimo vrijednosti površina odsječaka vremenske serije. kao što je to prikazano na slici 2.13.



SLIKA 2.12. GRAFIČKA INTERPRETACIJA NUMERIČKE INTEGRACIJE U REFII MODELU

Nakon transformacije vremenske serije kao što je to prikazano na slici 2.13., u ovisnosti o željenom cilju primjenjujemo metode rudarenja podataka na dobivene vrijednosti, ali u koordinaciji sa svim vrijednostima proizašlima iz REFII modela.

Kao i transformacija posredstvom REF modela, vrijednosti površina podsegmenata krivulje, također su nositelji informacija koji tek trebaju ući u proces obrade. Cilj analize determinira metodu koja će se primijeniti nad ovim vrijednostima.

Značajka ovog pristupa leži u činjenici što ovim pristupom kvantificiramo vremensku seriju, dajemo joj kvantitativnu dimenziju kako u njenim pojedinačnim segmentima, tako i u cjelini.

2.3.2. Područje efikasnosti/neefikasnosti proračuna površine ispod krivulje

Unutar REFII modela proračun površine ispod dijela krivulje ima osnovni zadatak dati kvantitativnu dimenziju cjelokupnome modelu. Kao što je već napomenuto niti jednu od ova tri podmodela ne treba promatrati zasebno već kao element mozaika koji se zove REFII i koji je sposoban rješavati kompleksne zadatke. Površina ispod krivulje daje nam kvantitativni pokazatelj, ali nam ništa ne govori o obliku krivulje za koju smo izračunali površinu. Što se odsječak predprocesirane krivulje za koju računamo površinu sastoji od većeg broja odsječaka to je prognostika o svojstvima iste na osnovi površine nepreciznija.

Procesuiranje vrijednosti koje predstavljaju niz površina ispod krivulja u slučaju obrade pretprocesiramo i klasificiramo na osnovu njihova oblika, te ima pridodjeljujemo indekse. Upravo vrijednost površine ispod krivulje nakon indeksiranja vrlo je podesna za obradu posredstvom algoritama poput stabla odlučivanja, *memory based reasoning*, klasteriranje i slično.

Osnovni zadatak proračuna površine ispod krivulje unutar REFII modela je pridodjeljivanje karakteristika voluminoznosti cjelokupnom modelu elementarnim

uzorcima. Teoretski gledano dvije krivulje mogu imati isti oblik, ali mogu biti različite s obzirom na površinu ispod krivulje. Ako je voluminoznost vremenskog odsječka jedan od elemenata važnih za analizu, tada u proces analize uključujemo i komponentu površine.

2.4. Proračun koeficijenta kutnog otklona

Posljednji element koji upotpunjuje REFII model je proračun koeficijenta kutnog nagiba. Nužnost uvođenja ovog elementa ilustrirana je slikom 1.7. U situaciji kada imamo vrijednost površine ispod krivulje, i definiran oblik krivulje, REF model nam može dati konceptualni oblik krivulje sa stupnjem vjerojatnosti kada procjenjujemo sličnost krivulja. Precizno određivanje odstupanja sa egzaktnim numeričkim pokazateljima moguće je posredstvom ovog podmodela. On je u osnovi vrlo precizan, i daje egzaktnu matematičke pokazatelje, što upućuje na određeni stupanj "krutosti". S obzirom da mi nastojimo postići određeni balans između fleksibilnosti i preciznosti, REF model i ovaj model se nadopunjuju na način da ovaj model služi kao pomoćno sredstvo za precizno otkrivanje odstupanja, odnosno rušenje hipoteze o visokom stupnju sličnosti krivulja kao što je to prikazano na slici 1.7.

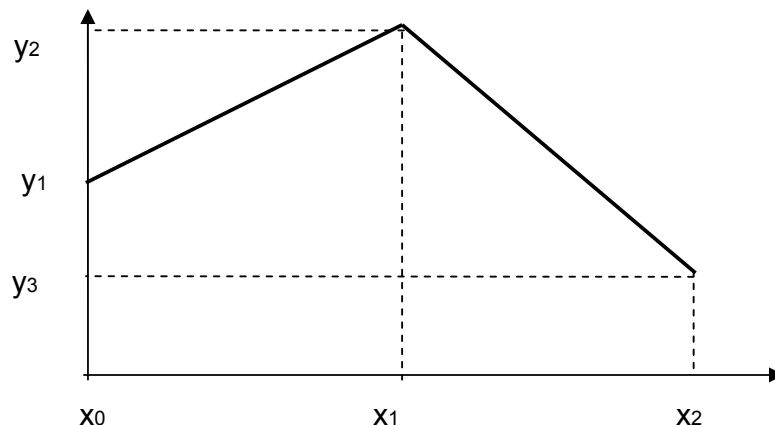
Korištenje isključivo ovog modela prilikom analize vremenskih serija može rezultirati potrebom za vrlo visokim stupnjem sličnosti oblika krivulje, što podrazumijeva vrlo visoku sličnost kutova svih pravaca, odnosno dužine pravaca koji reprezentiraju krivulju.

Kada usporedimo REF model sa ovim modelom, stupanj očekivane preciznosti kod ovog modela je krajnje precizan. U REF modelu rast nekog trenda u vremenskom odsječku definiramo sa R, dok se on može izraziti koeficijentom kutnog otklona 0.234. U REF modelu oznaku R dobiva kutni otklon vrijednosti koeficijenta kutnog otklona 0.234, i koeficijent kutnog otklona 0.876, dok između njih postoji velika razlika. Naravno, greška modela upućivala bi na stanovita odstupanja, ali se to odstupanje tim modelom ne može toliko precizno odrediti. Ponekad, ako imamo vrlo veliku vremensku seriju, koju uspoređujemo sa drugim vremenskim serijama, posredstvom REF modela ako postoji odstupanje u specifičnom vremenskom odsječku, i to u vrlo malom broju slučajeva tada se ta greška vrlo lako "izgubi". Upravo iz tog razloga razvijen je ovaj podmodel, kojemu glavna snaga leži u preciznosti. Ta preciznost može ponekad biti i kočnica, pa tako možemo ući u drugu krajnost, da algoritam "ne vidi očito". Upravo iz tog razloga ovaj model i REF model se savršeno nadopunjuju. REF model služi za brzu detekciju oblika krivulje, jer ne zahtjeva velik broj proračuna, a tada po potrebi stupa na scenu ovaj model s ciljem preciznog otkrivanja potencijalnih skrivenih zakonitosti, odnosno s ciljem prihvatanja ili odbijanja hipoteze.

Ilustracije praktične primjene svih modela, pa tako i ovog, kao i njihovu interakciju te efikasnost u pojedinim situacijama bit će prikazana kroz rad. Važno je napomenuti da ne postoji receptura na koji način i kako kombinirati REF model sa ovim modelom. Ona ovisi o konkretnoj situaciji i preferencijama analitičara.

Prednost proračuna koeficijenata kutnog otklona manifestira se kroz mogućnost definiranja razreda otklona pomoću koeficijenata kutnog otklona. Tako generalne trendove izražene sa oznakama "R" i "F" možemo svesti na niži stupanj granulacije – "Slab rast", "Srednji rast", "Oštar rast" ...

Model proračuna koeficijenta kutnog otklona prikazan je na slici 2.14.



SLIKA 2.13. MODEL PRORAČUNA NAGIBA PRAVCA I DULJINE KRAKA U VREMENSKOJ SERIJI

U prvom koraku pretprocesiranu vremensku seriju $S=(y_1, \dots, y_n)$ transformiramo u vremensku seriju $T=(t_1, \dots, t_n)$ na temelju izraza

$$t_i = \frac{y_i - \min}{\max - \min}$$

gdje je t_i element niza $T=(t_1, \dots, t_n)$

gdje min predstavlja minimalnu vrijednost na cijelom intervalu vremenske serije t , a max predstavlja maksimalnu vrijednost na cijelom intervalu vremenske serije t

Daljnji korak se svodi na računanje koeficijente kutnog odklona prema izrazima

$$k_j = t_{i+1} - t_i \quad \text{za} \quad t_{i+1} - t_i > 0 ;$$

$$k_j = t_i - t_{i+1} \quad \text{za} \quad t_{i+1} - t_i < 0 ;$$

$$k_j = 0 \quad \text{za} \quad t_{i+1} - t_i = 0$$

Vrijednost koeficijenta kutnog odklona kreće se u rasponu od 0-1. Ako bismo hipotetski željeli imati krajnje precizan model za procjenu oblika krivulje, tada bismo u analizama koristili vrijednosti koeficijenta kutnog odklona direktno bez dodatnih transformacija poput lingvističkih varijabli. Naravno, ovakav pristup povlači čitav niz tehničkih pitanja, ali je hipotetski u modelima moguće baratati i izvornim vrijednostima koeficijenta kutnog odklona. U tom slučaju moramo biti svjesni činjenice da će algoritam za primjerice procjenu jednakosti, vrijednost koeficijenta koji pripadaju odsječcima T_1 i T_2 , a iznose 0.231 i 0.232 interpretirati kao različite.

Alternativni pristup za potrebe analiza svodi se na korištenje intervalnih vrijednosti gdje se primjerice odklon u intervalu $<0.0001-0.3>$ deklarira kao interval niskog rasta.

Na taj način će algoritam za procjenu jednakosti vrijednost koeficijenta koji pripadaju odsječcima T_1 i T_2 , a iznose 0.231 i 0.232 interpretirati kao jednake.

U poglavlju 2.2.1. dana je definicija kutnih otklona pomoću fuzzy logike, gdje se umjesto intervalnih vrijednosti koriste lingvističke varijable.

Naredna tablica prikazuje primjer definicije intervalnih vrijednosti otklona kutova pomoću "crisp" logike:

Tablica 2.13. Intervalne vrijednosti otklona kutova

Raspon koeficijenata kutnog otklona	Trend odsječka (REF)	Lingvistička varijabla
0.7-1.0	R	Visok rast
0.4-0.7	R	Srednji rast
0.01-0.4	R	Nizak rast
0	E	Bez promjene
0.01-0.4	F	Nizak pad
0.4-0.7	F	Srednji pad
0.7-1.0	F	Oštar pad

Ovisno obrađivanoj problematici, rasponi intervalnih vrijednosti varijable mogu varirati.

Definicija intervalnih vrijednosti kako za koeficijente kutnog otklona, tako i za površinu ispod krivulje, temelj je za definiciju pojmovnih struktura intervalnih jednakosti, intervalnih nejednakosti i sličnosti.

2.4.1. Algoritamska interpretacija proračuna koeficijenata kutnog otklona

Algoritamska interpretacija ovog podsustava svodi se na primjenu prikazanih formula. Primjenom ovih formula dobivamo niz vrijednosti koje korespondiraju sa koeficijentima kutnog otklona između pravaca.

Daljnja prednost primjene koeficijenta kutnog otklona u ovom modelu očituje se u lakoj vizualnoj predodžbi nagiba na osnovu vrijednosti kuta.

Algoritamski postupak možemo opisati na naredni način:

m = broj odsječaka vremenske serije

brojač = 1

ZA *m* > brojača

$Tr > 0$ (R) koeficijent = (normiraj(*y*_{2,min},max)- normiraj(*y*_{1,min}, max))

$Tr < 0$ (F) koeficijent = (normiraj(*y*_{1,min},max)- normiraj(*y*_{2,min}, max))

$Tr = 0$ (E) koeficijent = 0

niz(brojač)=koeficijent

brojač = brojač +1

KRAJ

Funkcija normiraj (*v*,minimum, maximum)

$n = (v - \text{minimum}) / (\text{maximum} - \text{minimum})$

return n

Ovaj algoritam puni vrijednosti koeficijenata kutnog otklona u niz.

Isto tako, ovaj element modela prati pravilnosti ostalih dijelova modela s ciljem postizanja kompatibilnosti i istovremene primjenjivosti svih njegovih segmenata, te se uklapa u matricu transformacije REFII modela. Svakom odsječku vremenske serije pripada niz pokazatelja koji korespondiraju sa elementima modela, u koje između ostalih spada i proračun koeficijenta kutnog otklona.

2.4.2. Prednosti i nedostaci proračuna koeficijenta kutnog otklona

Prednost ovog modela očituje se u njegovoj egzaktnosti i preciznosti, što doprinosi jednoznačnom opisu vremenske serije. Ova preciznost isto tako može biti i nedostatak, jer zahtijevani stupanj preciznosti može rezultirati odbacivanjem određenih hipoteza, koje se za potrebe analize mogu protumačiti kao prihvatljive.

Spomenuti problem možemo riješiti definicijom raspona vrijednosti otklona u razrede, što ovisi o cilju i željenom karakteru same analize (intervalne jednakosti kod korištenja "crisp" logike, lingvističke varijable prilikom korištenja fuzzy logike.). Osnovni fokus ovog modela je jednoznačan opis oblika krivulje, i to u situacijama kada očekujemo vrlo visoki stupanj sličnosti. Može se reći da je ovaj model sve ono što REF model nije i obratno. REF model je vrlo fleksibilan model koji je prvenstveno fokusiran na oblik krivulje, ali njegov je nedostatak u prevelikom stupnju tolerancije koji barata samo pojmovima rasta, pada i jednakosti vremenskih odsječaka. Model baziran na koeficijentima kutnog otklona između vremenskih odsječaka ulazi u drugu krajnost, on traži egzaktnost i preciznost.

U analizama nam ponekad treba konceptualni oblik krivulje, pa tada koristimo REF model, odnosno vrlo precizan pristup određivanju oblika krivulje, što rezultira primjenu modela baziranog na koeficijentima kutnog otklona. Kroz rad će biti prikazana važnost obje koncepcije, kao i načini paralelne primjene obje koncepcije. Metodu baziranu na otklonima kutova možemo "ojačati" matematičkim instrumentarijem koji proračunava stupnjeve tolerancije. Na taj način možemo dobiti "hibridni model" koji dopušta određena odstupanja, gdje analitičar određuje stupnjeve slobode, ili pak možemo uvesti kategorije razreda. Ovaj pristup je determiniran određenim problemskim prostorima.

Cjelokupni REFII model fokusiran je na dva osnovna pokazatelja koja determiniraju vremensku seriju, a to su oblik i površinu krivulje.

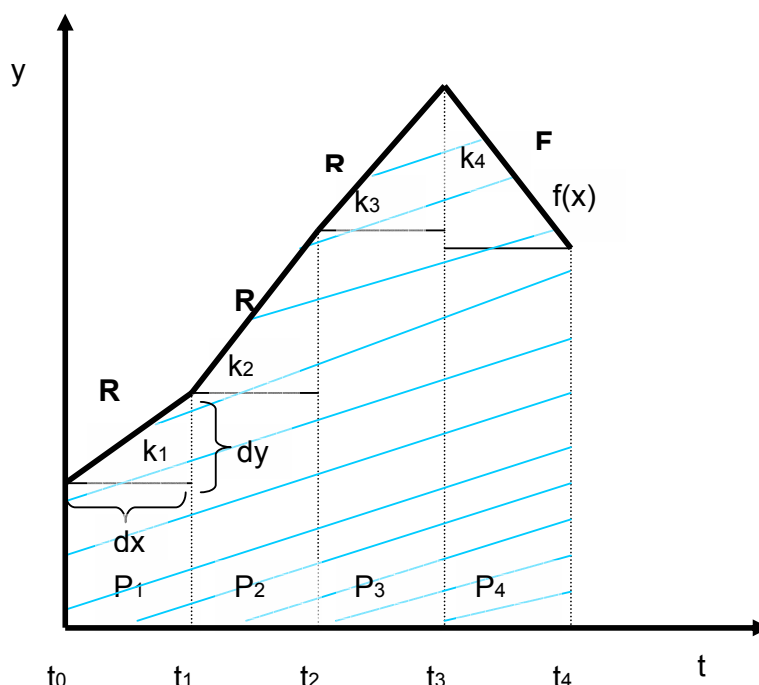
Gledano sa matematičke perspektive, pojedini elementi ovog modela mogu supstituirati potrebu za proračunom površine ispod krivulje, jer za tu namjenu možemo izračunati Euklidsku distancu između krajnjih točaka pravca koji omeđuju vremenski odsječak i X osi. Ove pokazatelje bismo mogli primijeniti u cilju dobivanja elementa kvantifikacije vremenske serije. Nedostatak takvog pristupa očitovao bi se u kompliciranom matematičkom instrumentariju za određivanje vrlo važnog elementa koji rezultira jednoznačnim opisom krivulje.

Postoje i drugi načini jednoznačnog opisa vremenske serije, ali su ovi modeli kao i pokazatelji izabrani s ciljem što efikasnijeg rješavanja problema baziranih na vremenskim serijama. Cilj je bio naći balans između efikasnog matematičkog instrumentarija, postojeće tehnologije, primjenjivosti sa ostalim algoritmiziranim postupcima koji se primjenjuju u analizi podataka i otkrivanju znanja, što kraćem vremenu procesuiranja algoritmiziranih postupaka, i ono što je najvažnije uspješnoj primjeni modela na vremenskim serijama s ciljem pronalaženja zakonitosti prema

različitim konceptualnim postavkama. Upravo ovaj zadnji element determinira postojanje cijelog modela i njemu je podređen kompletan matematički instrumentarij.

2.5. Shematski prikaz REF II modela

REFII model objedinjava prikazane metode na način na koji to prikazuje slika 2.15.



REF - Trend odsječka
 P_n - Površina odsječka
 K_n - Koeficijent kutnog otklona vremenskog odsječka

SLIKA 2.14. SHEMA REF II MODELA

Prikazana shema REFII modela objedinjava sve njegove glavne elemente. Kao što će biti prikazano navedeni pokazatelji su nositelji cjelokupnog modela, ali isto tako uz glavne pokazatelje možemo koristiti i izvedene pokazatelje, što ovisi o karakteru analize. Prikazani model pokazuje konceptualno objedinjenje i vezu između elemenata modela pomoću kojih analiziramo vremenske serije. Vremensku seriju za potrebe analize možemo promatrati bilo na razini pojedinačnih odsječaka, bilo kao niz grupe odsječaka.

2.6. Algoritamska interpretacija REFII modela

Krajnji algoritam mora objediniti sve tri spomenute cjeline i stvoriti temelj za primjenu analitičkih postupaka.

Algoritam za transformaciju vremenske serije u REFII model odvija se u nekoliko koraka

Vremensku seriju možemo deklarirati kao niz vrijednosti $S(s_1, \dots, s_n)$, gdje S predstavlja vremensku seriju, a (s_1, \dots, s_n) elemente serije S .

1. Korak - Vremenska interpolacija

Formiranje samostalnog vremenskog niza $V_i(v_{i1}, \dots, v_{in})$ na intervalu $\langle 1..n \rangle$ (Dani, tjedni, mjeseci, kvartali, godine) sa vrijednostima 0. Na temelju tako formiranog niza potrebno je provesti interpolaciju nedostajućih vrijednosti u $S(s_1, \dots, s_n)$ sa 0 na temelju formiranog niza V_i . Rezultat ovakve obrade je niz $S(s_1, \dots, s_n)$ sa interpoliranim nedostajućim vrijednostima iz niza $V_i(v_{i1}, \dots, v_{in})$.

2. Korak - Vremenska granulacija

U ovom koraku definiramo stupanj sažimanja vremenske serije $S(s_1, \dots, s_n)$ koja se nalazi u elementarnoj vremenskoj jedinici (dan, tjedan, mjesec...). U drugom koraku elemente postojeće vremenske serije sažimamo korištenjem statističkih funkcija poput sume, prosječne vrijednosti, moda na razini granuliranog odsječka. Na taj način vremensku seriju možemo svesti na veći stupanj granulacije (dani u tjedne, tjedni u mjesec ...), te dobivamo vremensku seriju $S(s_1, \dots, s_n)$ sa većim stupnjem granulacije. Na ovaj korak možemo se vraćati tijekom procesa analize s obzirom na ciljeve analize, što podrazumijeva obavezno ponovno provođenje procesa opisanima u narednim koracima

3. Korak - Normiranje

Postupak normiranja podrazumijeva transformaciju vremenske serije $S(s_1, \dots, s_n)$ u vremensku seriju $T(t_1, \dots, t_n)$ pri čemu je svaki element niza podvrgnut postupku min-max normizacije na intervalu $\langle 0, 1 \rangle$ i to

a) Vremenska serija T se sastoji od elemenata (t_1, \dots, t_n) , pri čemu se t_i računa kao $t_i = ((s_i - \min(S)) / (\max(S) - \min(S)))$, gdje su $\min(S)$ i $\max(S)$ minimalna i maksimalna vrijednost vremenske serije S .

b) Vremenski pomak između elementarnih uzorka (mjerilo vremenske kompleksnosti) odsječka na X osi određen je sa $d(t_i, t_{i+1}) = a$

4. Korak - transformacija u REF notaciju

Prema formuli $T_r = t_{i+1} - t_i$ $T_r > 0 \Rightarrow R$; $T_r < 0 \Rightarrow F$; $T_r = 0 \Rightarrow E$, gdje su Y_i elementi niza N_s

5. Korak – Proračun nagiba pravca na osnovu kuta

Koeficijent kutnog odklona \Rightarrow

$T_r > 0$ (R)	Koeficijent $= t_{i+1} - t_i$
$T_r < 0$ (F)	Koeficijent $= t_i - t_{i+1}$
$T_r = 0$ (E)	Koeficijent $= 0$

6. Korak – Proračun površine ispod krivulje

Numerička integracija metodom pravokutnika

$$p = ((t_i * a) + (t_{i+1} * a)) / 2$$

7. Korak - Kreiranje vremenskih indeksa

Građenje hijerarhijskog stabla indeksa ovisno o karakteru analize, gdje element strukturiranog indeksa može biti i atribut poput šifre klijenta

8. Korak - Kreiranje razreda

Kreiranje izvedenih vrijednosti atributa na temelju površine ispod krivulje i otklona kutova.

Moguće je kreirati razrede primjenom klasične *crisp* logike, ili primjenom *fuzzy* logike.

9. Korak - Povezivanje tablice transformacije REFII modela sa relacijskim tablicama koje sadrže attribute koji nemaju vremensku dimenziju

Ovih devet osnovnih koraka temelj su algoritmiziranog postupka na kojem se temelji REFII model čiji je krajnji rezultat formiranje matrice transformacije. Matrica transformacije je temelj za provođenje daljnjih analitičkih postupaka s ciljem analize vremenske serije.

Naredni jednostavni primjer pokazuje kako transformirati izvornu vremensku seriju u REFII na temelju opisanog postupka.

Zadana je vremenska serija u tablici 2.14.

Tablica 2.14. Izvorna vremenska serija

Datum	Iznos
01.1.2005	0
02.1.2005	11
03.1.2005	10
04.1.2005	26
05.1.2005	10

Obrazac za proračun intervalnih vrijednosti računa se na temelju tablice 2.13.

U prvom koraku izvrši se vremenska interpolacija za dani period. Kako unutar tablice 2.14. nema nedostajućih vrijednosti za dani period nije potrebno interpolirati nul vrijednosti. Primjerice ako za 03.01.2005 ne bi postojao podatak, u ovom koraku bi vrijednost za taj datum bila 0.

Kako želimo zadržati stupanj granulacije na razini dana, tada nije potrebno sažimati podatke na nekoj većoj razini granulacije, primjerice tjedana, što spada u drugi korak.

Minimalna vrijednost u vremenskoj seriji iz tablice 2.14. je 0, a maksimalna vrijednost je 26. Na temelju toga vršimo normiranje vremenske serije prema formuli za normiranje i 4. koraka te dobivamo tablicu 2.15.

Tablica 2.15. Normirane vrijednosti vremenske serije

Datum	Iznos	Normirane vrijednosti
01.1.2005	0	0
02.1.2005	11	0,423076923
03.1.2005	10	0,384615385
04.1.2005	26	1
05.1.2005	10	0,384615385

Primjerice, prva normirana vrijednost iz tablice (01.01.2005.) dobivena je iz formule za normiranje iz trećeg koraka, $0-0/26=0$.

Na temelju četvrtog i petog koraka, za prvi odsječak između 01.1.2005. i 02.1.2005. na temelju formule iz četvrtog koraka $(0,423076923-0)>0$, za taj odsječak je dodijeljena REF oznaka R, a koeficijent kutnog odklona je prema formuli iz petog koraka $(0,423076923-0)=0,423076923$. Površinu ispod krivulje računamo na način kako je to prikazano u šestom koraku, te ona za odsječak između 01.1.2005. i 02.1.2005. na temelju formule iz šestog koraka iznosi $(0+11)/2=5,5$ ($a=1$). Unutar REFII modela moguće je računati površinu ili na temelju normiranih ili na temelju originalnih vrijednosti. U ovom primjeru prikazan je način izračuna na temelju originalnih vrijednosti.

U daljnjem postupku prema sedmom koraku kreiraju se vremenski indeksi, koji mogu predstavljati vrijednosti tipa 1- trend sa subote na nedjelju, 2- trend s nedjelje na ponedjeljak i slično.

U osmom koraku vrši se pridodjeljivanje intervalnih vrijednosti na temelju tablice 2.13., te se prema potrebi na isti način mogu klasificirati i površine. Primjerice, prvom se odsječku između 01.1.2005. i 02.1.2005. na temelju vrijednosti iz tablice i vrijednosti koeficijenta kutnog odklona pridodjeljuje kategorija *Srednji rast*.

Nakon provedbe ovih operacija nad svim vrijednostima iz tablice 2.15. dobije se tablica 2.16.

Tablica 2.16. Vremenska serija u REFII notaciji

Indeks	REF	Koeficijent kutnog odklona	Površina	Intervalna vrijednost
1	R	0,423076923	5,5	Srednji rast
2	F	0,038461538	10,5	Nizak pad
3	R	0,615384615	18	Srednji rast
4	F	0,615384615	18	Srednji pad

2.7. Kako analizirati vremenske serije posredstvom REFII modela

REFII model u osnovi je koncepcija sastavljena iz tri opisane podcjeline, kojoj je prvenstveni cilj transformacija vremenske serije u niz pokazatelja koji jednoznačno definiraju vremensku seriju. REF, površina ispod krivulje, te koeficijent kutnog odklona su pokazatelji koji jednoznačno opisuju odsječak vremenske serije, a niz takvih odsječaka zajedno čine transformiranu vremensku seriju. Ovakav niz pokazatelja poredani su redoslijedom pojavnosti u jedinici vremena i objedinjeni u zajedničku koncepcijsku strukturu transformirane vremenske serije koju nazivamo *matricom transformacije*.

Takva struktura gledano sa perspektive dinamičke memorije može biti matrica reda $4 \times t_n - 1$, odnosno gledano sa perspektive trajnog zapisa na disku datoteka sa 4 atributa i dužine $t_n - 1$.

Grafički prikaz ove strukture dan je u narednoj tablici:

Tablica 2.17. Matrica transformacije

Indeks vremenskog odsječka	I_1	I_2	...	I_n
REF oznaka	REF ₁	REF ₂	...	REF _n
Koeficijent kutnog odklona	Koeficijent kutnog odklona ₁	Koeficijent kutnog odklona ₂	...	Koeficijent kutnog odklona _n
Površina vremenskog odsječka	P_1	P_2	...	P_n

Pokazatelji vremenskog odsječka izračunati su na temelju koordinata dvije susjedne vrijednosti u vremenskoj seriji. Tako je primjerice odsječak sa indeksom I_1 formiran na osnovu vrijednosti koordinata vremenske serije t_0 i t_1 . Indeks vremenskog odsječka služi za jednoznačnu identifikaciju vremenskog odsječka s ciljem njegove analize, odnosno *kreiranje bazičnog uzorka* o čemu je bilo riječi prilikom predstavljanja matematičkih osnova modela.

Indeksi mogu biti složeno strukturirani te mogu u sebi sadržavati hijerarhijske elemente, ako i elemente pripadnosti, te vezne elemente prema ostalim izvorima podataka.

Elementi obuhvaćeni u prethodnoj tablici temeljni su elementi REFII modela, sa kojima je moguće jednoznačno opisati krivulju i izvršiti sve analize zbog kojih je i ovaj model i razvijen. Osim opisanih pokazatelja moguće je obuhvatiti i izvedene pokazatelje prikazane kroz razvoj modela, ali to je opcionalni pristup koji ovisi o karakteru analize. Pokazatelji navedeni u prethodnoj tablici temelj su za sve analize zbog kojih je ovaj model razvijen.

Nakon transformacije vremenske serije posredstvom REFII modela dobijemo transformiranu vremensku seriju u obliku prikazanu u tablici.

Ovako transformirani podaci procesuiraju se algoritmiziranim metodama s ciljem rješavanja konkretnih problema iz domene vremenskih serija. Algoritmizirane metode i postupci koje služe za rješavanje konkretnih problema spadaju u širi pojam REFII modela

O samom karakteru analize ovisi kako će se tako transformiranoj vremenskoj seriji pristupiti po pitanju "lomljenja" ovako formirane strukture u manje logičke cjeline. Ako na primjer promatramo tjedne, a vremenska serija sadrži podatke za svaki dan u tjednu kroz cijelu godinu, tada logički razlomimo seriju na tjedne i analiziramo vremenske odsječke posredstvom konkretnog algoritma. Ovaj postupak "lomljenja" vremenske serije na manje analitički usporediva logičke odsječke vežemo uz pojam *vremenske kompleksnosti*. Ova mjera određuje točku logičkog loma vremenske serije s obzirom na cilj analize, a dio je algoritmiziranog postupka analize. Tako primjerice ovaj koeficijent za tjedan može imati vrijednost 7, ili 5 (radni dani u tjednu), pri čemu

treba voditi računa o nepostojanju vrijednosti u određenoj vremenskoj točki, sa čim se ova mjera također mora nositi.

Postoji čitav niz mogućih primjena, što će biti prikazano u tekstu. U poglavlju koje opisuje REF model dane su osnovne smjernice kako se provodi analiza sezonskih oscilacija primjenom ove metode. Kao što je vidljivo REF model koristi i neke vrijednosti koje nisu obuhvaćene u prethodno prikazanoj tablici, što je slučaj i sa ostalim podmodelima. Te vrijednosti proizlaze iz specifičnosti svakog podmodela i bit će prikazani u razmatranjima koje se odnose na konkretne analize.

Ako uzmemo primjer sezonskih oscilacija, tada nakon što otkrijemo sezonsku oscilaciju sa visokim stupnjem greške koristimo ostala dva podmodela za precizno otkrivanje uzroka grešaka. Model prikazan u tablici inicijalni je model iz kojeg počinju sve navedene analize. Izvedene tablice usko vezane uz pojedini podmodel razmatraju se posebno s obzirom na cilj analize.

Važno je napomenuti da se svaki od tri navedena elementa REFII modela međusobno nadopunjuju, ali da su nekim situacijama neki od elemenata irelevantni.

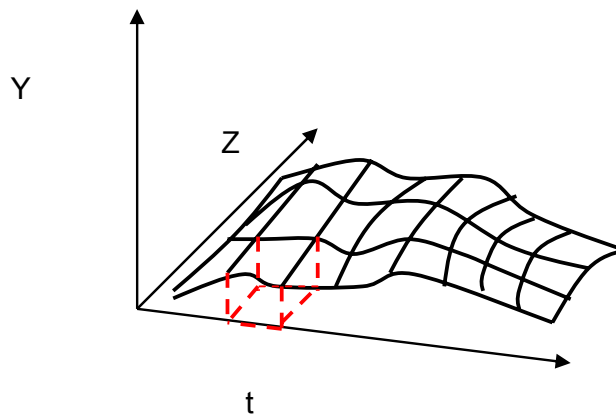
Sve to naravno ovisi o cilju analize i analitičkom pristupu utjelovljenom u algoritmiziranom postupku. Na prikazanu tablicu moguće je djelovati različitim algoritmiziranim postupcima koji u konačnici daju željene rezultate. Isto tako je moguće nad istom tablicom primjenjivati i uspoređivati efikasnost primjene različitih vrsta algoritama s ciljem rješavanja istog problema. Redoslijed i prioritet analize s obzirom na problematiku ovise od slučaja do slučaja. Osim kvantitativnog elementa i elementa oblika krivulje pristup analizi i algoritmiziranom postupku determinira i željeni stupanj preciznosti. Što je stupanj preciznosti i analitičkih zahtijeva veći to će više biti iskorišteni svi mogući pokazatelji koji proizlaze iz modela.

Ovaj transformacijski model, polazišna je točka za niz različitih vrsta analiza koje se provode nad vremenskim serijama, a koje ova koncepcija uspješno rješava. Kao što će biti vidljivo kroz daljnji tekst elementi prikazani u tablici proceduralno procesuirani kroz različite vrste algoritama mogu otkriti različite vrste znanja iz vremenskih serija, i upravo u tome leži snaga REFII koncepcije.

Ovaj koncept može biti uspješno primijenjen za razvoj upitnog jezika za vremenske serije. Da bi slika bila što potpunija nikako ne trebamo zanemariti izvedene pokazatelje koji spadaju u specifične pokazatelje podmodela. Njihova prisutnost može omogućiti kvalitetniji pristup algoritmiziranog rješavanja određenih tipova problema iz promatranog problemskog prostora. S obzirom na vrlo širok raspon tih pokazatelja, njih nećemo opisivati pojedinačno nego će biti razmatrani u sklopu konceptualne primjene modela, i algoritmiziranih rješenja za ekstrahirana znanja iz vremenskih serija.

2.8. Trodimenzionalna interpretacija REFII modela

Vremenske serije promatramo prvenstveno kroz 2D prostor. U takvom prostoru promatramo kretanje zavisne varijable u odnosu na nezavisnu varijablu (vrijeme). REFII model dopušta mogućnost promatranja i analize više od jedne varijable u jedinici vremena. Slika 2.15. daje grafičku interpretaciju takvog REFII modela



SLIKA 2.15. REFII I 3D PROSTOR

Što dobivamo takvom interpretacijom?

U svjetlu vremenskih serija možemo mnogo preciznije promatrati korelatorne odnose dviju varijable ovisne o vremenskom pomaku. Naravno u takvim situacijama kada REFII model promatramo u n-dimenzionalnom prostoru dolazimo do potrebe modifikacije matematičkog instrumentarija. Tako za potrebe računanja površine ispod krivulje sada moramo koristiti Monte Carlo metodu, REF model dobiva nove dimenzije, pa sada ne postoji jednosmjerno (2d) praćenje trendova, već u obzir moramo uzeti korelatorne odnose. Za razliku od 2d prostora gdje se je vremenski odsječak reprezentirao krnjim pravokutnikom u 3d prostoru dolazimo do pojma krnjeg kvadra. Korelatorne odnose, načine interpretacije određuje sam tijek analize.

Ova interpretacija dokazuje da REFII model nije samo ograničen na analizu vremenskih serija, već on u ovakvom obliku može biti primijenjen za primjerice prepoznavanje oblika, kao što je *face recognition*. Prvenstvenu kompleksniju i temeljitiju primjenu ovog modela moguće je ostvariti u korelatornoj analizi dviju varijabli u jedinici vremena.

Ovdje se suočavamo sa problematikom kombinatorne eksplozije, tako da je potrebno vrlo precizno odrediti ciljeve analize posredstvom ovog modela. Daljnji problem svodi se na model transformacije koji mora biti n-dimenzionalna matrica. U takvom modelu transformacije potrebno je razraditi postupke analize koji će biti brzi i efikasni prilikom obrade pokazatelja. Uzmimo kao primjer procjenu sličnosti 3D krivulje. Svaka *krnja kocka* koja je prikazana na slici 18 mora biti transformirana u model. Svaki taj model mora imati korelatorne veze sa susjednim krnjim kockama. Razrađeniji modeli mogli bi imati korelatorne odnose sa $t \rightarrow t+1(x)$, $t \rightarrow t(y)$, $t \rightarrow (t, y+1)$

Kod računanja površine sada se suočavamo sa problemima lokacije područja za koje vršimo analizu.

Ovaj model može biti vrlo koristan u procjeni korelatornih odnosa, no u radu će biti prikazane alternativne metode procjene korelatornih odnosa koje se svode na 2D REFII modela, koje mogu dati sasvim dostatne rezultate gledano kroz prizmu problematike kojom se bavimo. Interpretacija 3D prostora kroz REFII model dana je zbog ilustracije širokog raspona njegove primjene.

3. Pojemovne strukture i njihova definicija na osnovu REFII modela

3.1. Pojam jednakosti vremenskih odsječaka

Jedna od temeljnih operacija koju možemo provesti posredstvom REFII modela je operacija uspoređivanja i to uspoređivanja jednakosti vremenskih odsječaka, kao i jednakosti vremenskih serija.

Jednakost možemo definirati kao

Eksplicitnu (Jednakost bazirana na izvornim vrijednostima REFII modela)

Intervalnu (Jednakost bazirana na vrijednostima klasificiranu u razrede)

Za dva vremenska odsječaka (odsječak 1 i odsječak 2) vremenske kompleksnosti od jedne vremenske jedinice kažemo da su eksplicitno jednaki ako vrijedi:

$REF(1) = REF(2)$ i {oznake REF modela}
 $k(1) = k(2)$ i {koeficijenti kutnog otklona}
 $P(1) = P(2)$ {površine ispod krivulja}

S obzirom na karakter analize u proces obrade uzimamo one elemente REFII modela koji su relevantni za proces analize.

Ovim relacijama je jednoznačno definirana jednakost vremenskih odsječaka. Ako želimo pojam jednakosti definirati za više od jednog vremenskog odsječaka, tada primjenjujemo ovu temeljnu jednakost na rasponu od 1..n vremenskih odsječaka. To znači da za pojam jednakosti niza vremenskih odsječaka, odnosno vremensku seriju vrijedi pravilo jednakosti temeljeno na prethodnoj definiciji :

Vremenska serija 1 (V_{s1}) i vremenska serija 2 (V_{s2}) su eksplicitno jednake, ako i samo ako su eksplicitno jednaki svi njihovi korespondentni vremenski odsječci ($V_{o1}..V_{on}$) sastavljeni od elemenata REFII modela.

$V_{s1}(V_{o1}..V_{on}) = V_{s2}(V_{o1}..V_{on})$ ako i samo ako $V_{s1}(V_{o1}) = V_{s2}(V_{o1})$, $V_{s1}(V_{o2}) = V_{s2}(V_{o2})$, $V_{s1}(V_{o3}) = V_{s2}(V_{o3})$... $V_{s1}(V_{on}) = V_{s2}(V_{on})$.

Pojednostavljeno rečeno dva su vremenska odsječaka jednaka ako i samo ako su eksplicitno jednaki svi elementi REFII modela koji opisuju vremenski odsječak.

Tako primjerice ako uspoređujemo dva vremenska odsječaka od kojih jedan ima REF oznaku R, otklon vrijednost koeficijenta kutnog otklona 0.304576 i površinu ispod krivulje 30.34, možemo reći da je taj odsječak jednak drugom odsječku jedino ako taj odsječak ima potpuno iste vrijednosti REFII varijabli.

Na ovaj način definiramo krutu matematičku definiciju jednakosti odsječaka vremenskih serija posredstvom REFII modela, odnosno eksplicitnu jednakost odsječaka.

Imajući u vidu problematiku analiza rudarenja podataka te problematiku pretraživanja uzoraka gdje bi analiza provedena na temelju eksplicitne jednakosti odsječaka bila neprikladna, u REFII model uvodimo pojam intervalne jednakosti.

Intervalnu jednakost možemo definirati kao jednakost razreda, odnosno jednakost intervalnih vrijednosti.

Intervalne vrijednosti definiramo kao grupirane eksplicitne vrijednosti u intervalne logičke cjeline izvedene na temelju vrijednosti varijabli REFII modela.

Tako se primjerice intervalna vrijednost otklona kuta može definirati kao oštar rast za $REF=R$, i kutni otklon u u intervalu od 0.8-1.0, ili pak površina u rasponu od 10-40 .

Za dva vremenska odsječka (odsječak 1 i odsječak 2) kažemo da su intervalno jednaki ako vrijedi :

$$\begin{aligned} REF(1)_{intervalno} &= REF(2)_{intervalno} \text{ i} \\ k(1)_{intervalno} &= k(2)_{intervalno} \text{ i} \\ P(1)_{intervalno} &= P(2)_{intervalno} \end{aligned}$$

Pojam jednakosti intervalnih odsječaka možemo definirati i kao pripadnost istoj grupi. Na primjer : "Oštar rast"="Oštar rast", "Oštar rast" \neq "Oštar pad"

Na temelju intervalnih vrijednosti možemo definirati i jednakost vremenskih serija.

Vremenska serija 1 (V_{s1}) i vremenska serija 2 (V_{s2}) su jednake, ako i samo ako su međusobno jednaki svi njihovi intervalni odsječci ($V_{o1}..V_{on}$) sastavljeni od elemenata REFII modela.

$$V_{s1}(V_{o1}..V_{on})=V_{s2}(V_{o1}..V_{on}) \text{ ako i samo ako } V_{s1}(V_{o1}) = V_{s2}(V_{o1}), V_{s1}(V_{o2}) = V_{s2}(V_{o2}), \\ V_{s1}(V_{o3}) = V_{s2}(V_{o3})... V_{s1}(V_{on}) = V_{s2}(V_{on}).$$

Ako pojam jednakosti definiramo na temelju eksplicitne jednakosti, tada bi i minorno odstupanje pojedinih pokazatelja primjerice koeficijenta kutnog otklona za 0.00001 rezultiralo odbacivanjem hipoteze o jednakosti vremenskih odsječaka, odnosno to bi u konačnici imalo utjecaja na procjenu jednakosti vremenskih serija.

S obzirom na to da je REFII model konstruiran s ciljem prilagodljivosti i univerzalnosti s obzirom na ciljeve analize, o cilju analize ovise definicija intervalnih vrijednosti za sva tri elementa REFII modela.

Isto tako istovremeno unutar REFII modela možemo istovremeno koristiti intervalnu i eksplicitnu jednakost.

Prednost REFII modela očituje se u tome što analitičar ima mogućnost utjecaja na definiciju jednakosti vremenskih odsječaka s obzirom na karakter analize.

Definicija jednakosti na razini vremenskog odsječaka temeljna je operacija na kojoj se temelje algoritmi za otkrivanje uzoraka.

Naredni, vrlo važan termin koji se pojavljuje u analizi posredstvom REF II modela a ima veze sa definicijom pojma jednakosti je mjerilo vremenske kompleksnosti.

Ovaj termin koriste autori koji se bave problematikom epizoda u vremenskim serijama [Manilla, 1997]. Primjerice pojam srijede i četvrtka u terminologiji vremenske kompleksnosti mogli bismo označiti sa 2, ako je duljina vremenskog odsječaka jednaka mjerilu od jednog dana.

Manilla [Manilla, 1997] pojam vremenske kompleksnosti koristi prilikom definicije vremenskih prozora (eng. time window). Sa perspektive jednakosti vremenskih odsječaka pojam uključivanja mjerila vremenske kompleksnosti kao elementa jednakosti bitan je kod prepoznavanja sezonskih oscilacija.

U tom slučaju nije samo dovoljno procijeniti jednakost vremenskih odsječaka, nego i vremenski period pojavnosti te jednakosti unutar vremenske serije.

3.2. Pojam nejednakosti vremenskih odsječaka

Vrlo je važno definirati pojam nejednakosti u vremenskim serijama s obzirom na REFII model. Pojam nejednakosti ovisi o cilju analize. Sa jedne strane cilj analize može biti usmjeren na oblik krivulje, s druge strane na kvantitativna obilježja (površina ispod krivulje), ili pak na oba aspekta.

Analitičar na osnovu ciljeva analize određuje koeficijente sigurnosti koji su direktno odgovorni za donošenje suda o jednakosti/nejednakosti vremenskih odsječaka.

Pojam nejednakosti vremenskih odsječaka izvodimo iz pojma jednakosti vremenskih odsječaka, i ona može biti eksplicitnu i intervalna.

Za razliku od pojma jednakosti vremenskih odsječaka gdje pojam eksplicitne jednakosti definiramo uz uvjet :

$$\begin{aligned} \text{REF}(1) &= \text{REF}(2) \text{ i} \\ k(1) &= k(2) \text{ i} \\ P(1) &= P(2) \end{aligned}$$

pojam eksplicitne nejednakosti može biti vezan uz nejednakost svih elemenata ili barem jednog elementa REFII modela.

Tako primjerice za dva odsječka (odsječak 1 i odsječak 2) možemo definirati eksplicitnu nejednakost uz uvjet :

$$\begin{aligned} \text{REF}(1) &\neq \text{REF}(2) \text{ ili} \\ k(1) &\neq k(2) \text{ ili} \\ P(1) &\neq P(2) \end{aligned}$$

Odnosno intervalnu nejednakost ako vrijedi :

$$\begin{aligned} \text{REF}(1)_{\text{intervalno}} &\neq \text{REF}(2)_{\text{intervalno}} \text{ ili} \\ k(1)_{\text{intervalno}} &\neq k(2)_{\text{intervalno}} \text{ ili} \\ P(1)_{\text{intervalno}} &\neq P(2)_{\text{intervalno}} \end{aligned}$$

Ovisno o cilju analize postoji različita kombinacija definicija eksplicitne i intervalne nejednakosti, gdje se primjerice neinteresantni elementi za analizu u okviru REFII modela zanemaruju.

Isto tako potrebno je definirati i pojam nejednakosti vremenskog odsječaka koji možemo definirati kao :

Vremenska serija 1 (V_{s1}) i vremenska serija 2 (V_{s2}) su nejednake, ako je međusobno jednako i od n korespondentnih odsječaka, pri čemu je $i < n$. Elementi tih vremenskih serija sastavljeni od komponenti REFII modela.

U slučaju da je $i = n$ to znači da su dvije vremenske serije jednake.

Pojam jednakosti i nejednakosti u REFII modelu treba razlikovati od pojma "sličnosti". Sličnost je puno širi pojam od pojma jednakosti. Slični vremenski odsječci ne moraju istovremeno biti i jednaki.

3.3. Pojam inverznosti vremenskih odsječaka

U svjetlu REFII modela možemo definirati pojam inverznosti vremenskog odsječaka, odnosno inverznosti vremenske serije. Za dva vremenska odsječaka kažemo da su inverzni ako vrijede pravila :

$REF(1) \neq REF(2)$ i $REF(1), REF(2) \neq "E"$ i
 $k(1)=k(2)$ odnosno $k(1)_{intervalno} = k(2)_{intervalno}$

U svjetlu REF modela, svaki element promatranog vremenskog odsječaka oznake "F" u prvoj vremenskoj seriji korespondira sa oznakom "R" drugog vremenskog odsječaka u drugoj vremenskoj seriji. Isto tako svaki element promatranog vremenskog odsječaka oznake "R" u prvoj vremenskoj seriji korespondira sa oznakom "F" drugog vremenskog odsječaka u drugoj vremenskoj seriji.

Kod pojma inverzije bitno je odrediti mjerilo vremenske kompleksnosti na kojoj procjenjujemo inverziju. Proračun površine ovdje možemo izostaviti, jer se inverzija više oslanja na oblik krivulje.

Temeljem REF modela procjenu inverzije funkcija možemo izračunati izrazom:

$$I = \frac{In}{n} * 100$$

Gdje je:

I- stupanj inverznosti u vremenskom odsječku

In- broj inverznih elemenata

n – dužina promatranog vremenskog odsječaka ili serije (mjerilo vremenske kompleksnosti).

Mjera stupnja inverznosti koja u sebi uključuje REF elemente može biti interpretirana i kao faktor sigurnosti inverzije, odnosno pomoćna mjera za prihvatanje/ odbijanje hipoteze o inverznosti vremenskih odsječaka.

Kada računamo površinu ispod krivulje kao element pojma inverzije, tada se koncentriramo na površinu trokuta koji zatvaraju dva kraka. Element površine trokuta u pojmu inverzije uključujemo s perspektive jednakosti. Tako za precizno određivanje inverzije krivulja možemo reći da je odsječak inverzan ako su njegovi kutovi inverzni, REF oznake inverzne i površine jednake. Ovako definirana inverzija je precizna definicija potpune inverzije, koja mora dopuštati i uvođenje pojma vjerojatnosti kao što je to slučaj sa stupnjem inverznosti u vremenskom odsječku.

Za računanje stupnja inverznosti baziranog na pokazateljima površine, služimo se definicijom jednakosti (Prosječni intervalni koeficijent sigurnosti).

Pojam inverzije vremenskih serija, odnosno odsječaka vremenskih serija koristan je element otkrivanja uzoraka koji pokazuje na stanovitu pravilnost unutar serije.

Stupanj preciznosti i orijentiranost na segmente kao što je površina ili oblik krivulje određuje analitičar. Kao sve pojmovne kategorije u REF modelu, gdje postoji visok stupanj tolerancije i sloboda izbora po pitanju selekcije ključnih elementa, tako i pojam inverzije vremenskog odsječaka može biti promatran sa perspektive analiziranih elemenata i pripadajućih faktora sigurnosti. Inverzija vremenskih odsječaka realizira se posredstvom algoritmiziranih postupaka, te može biti procesuirana poput traženja

jednakosti unutar vremenskih odsječaka i usmjerena prema potvrđivanju hipoteze, ili pak automatskom pronalaženju inverznih uzoraka. Važno je napomenuti da između inverznih uzoraka ne mora nužno biti t_0+1 kao mjerilo vremenske kompleksnosti, nego mjerilo vremenske kompleksnosti koje određuje distancu u vremenskom prostoru može biti i t_0+n .

Inverzija u vremenskim serijama kao pojam može biti promatrana i kroz prizmu korelacije vremenskih odsječaka i prizmu epizoda u vremenskim odsječcima kojima mjerilo vremenske kompleksnosti može biti t_0+1 odnosno t_0+n .

Premda je pojam korelacije i vremenskih epizoda mnogo kompleksniji od onog što nam nudi pojam inverzije, ovo može biti alternativni način procjene ovih kategorija u određenim situacijama, gdje na osnovu pojma inverznosti vremenskih odsječaka donosimo sud o specifičnim korelatornim odnosima odsječaka i epizoda u vremenskim serijama.

Ovaj pojam možemo proširiti i pojmom djelomične inverzije vremenskog odsječaka, gdje sa jedne strane ovaj pokazatelj korespondira sa koeficijentom sigurnosti cjelokupnog odsječaka, a sa druge strane možemo analizirati područje inverznosti, odnosno dio odsječaka koji možemo nazvati inverznim, i dio koji to nije.

3.4. Pojam korelacije vremenskih odsječaka

Korelacija kao mjera procjene međuzavisnosti između dviju pojava vrlo se efikasno može iskoristiti gledanu kroz problematiku vremenskih nizova i REFII modela.

Teško je govoriti o korelatornoj analizi kompletne vremenske serije. Naš fokus bit će usmjeren na korelatornu analizu odsječaka vremenske serije.

Na ovaj način možemo promatrati korelatorske odnose pojedinih vremenskih odsječaka kroz prizmu REFII modela. Ako na primjer imamo podatke o određenim bankovnim transakcijama u periodu od godine dana kroz vremensku dimenziju, tada možemo promatrati korelatorske odnose transakcija u odnosu na dane u tjednu. Primjerice da li postoji korelatorni odnos određenih dana u tjednu po pitanju promatrane bankovne transakcije.

Na ovaj način možemo kako testirati hipoteze o korelatornim odnosima (AKO postoji smanjenje obima transakcija u petak, tada će se povećati obim transakcija u ponedjeljak). Isto tako možemo istražiti potencijalno postojanje ovakvih korelatornih odnosa na promatranom uzorku.

Ostvarenje ove ideje svodi se na korištenje osnovnih elementa REFII modela i statističke mjere korelacije. Kada govorimo o REFII elementima koji se analiziraju, tada se prvenstveno misli na površinu ispod krivulje, REF modela, i koeficijenta kutnog odklona.

Grube korelativne odnose možemo procijeniti upotrebom čistog REF modela i statističke formule za korelaciju. U tu svrhu moramo napraviti supstituciju elemenata REF modela kao je to prikazano u narednoj tablici:

Tablica 3.1. Supstitucija elemenata

REF oznaka	Supstituirajući element
R	1
E	0
F	-1

Površine odsječaka i kutovi koji zatvaraju krakove mogu se u formulu uvrstiti u izvornom obliku.

U nizu vremenskih serija 1..n, gdje svaka vremenska serija ima 1..e vremenskih odsječaka korelatorne odnose računamo između niza korespondentnih vremenskih odsječaka u n vremenskih serija kao što je to prikazano slikom 19.

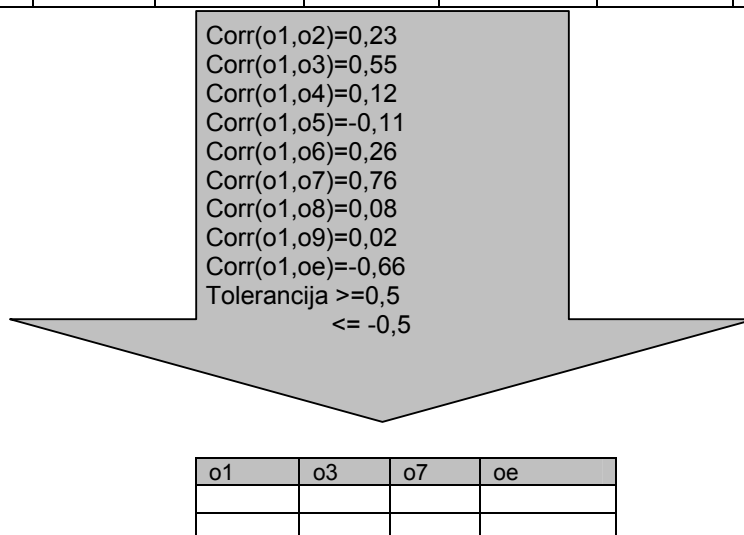
Na slici je vidljivo da se uspoređuju korespondentni odsječci. Provodi se korelatorska analiza korespondentnih odsječaka kako bi se otkrili korelatorski odnosi unutar određenih segmenata vremenskih serija. Odsječak odgovara pojmu stupnja granulacije u vremenskoj seriji. Tako primjerice u vremenskoj seriji koja sadrži podatke o mjesečnim transakcijama, gdje želimo istražiti korelatorne odnose na razini dana stupanj granulacije je dan a mjerilo vremenske kompleksnosti je tjedan.

Slika 19. prikazuje korelatorski odnos 1:1. Što znači da prikazuje korelaciju između dviju pojava. Ako želimo istražiti korelatorski odnos 1:n. Tada provodimo prvi korak koji je prikazan na slici 19. Pretpostavimo hipotetski primjer da takve korelatorne analize provodimo svaki mjesec unutar 12 mjeseci. Nakon 12 mjeseci istražujemo korelatorne odnose 1:n na način da potvrdimo hipotezu primjerice ako ponedjeljkom raste broj transakcija, tada broj transakcija utorakom, srijedom, i petkom pada.

Dublja kvantitativna analiza korelatorskih odnosa značila bi usredotočenost na kutove između krakova krivulje. Tvrdnja o postojanju korelacije može biti usredotočena na tezu o postojanju korelacije ili nepostojanju korelacije, kao i o kvantifikacije te pojave, te prihvatanju hipoteze na osnovu granice prihvatanja hipoteze zadane od strane analitičara.

Tablica 3.2. Tablica korelatorskih odnosa

Serija 1..n	Odsječak o1	Odsječak o3	Odsječak o4	Odsječak o5	Odsječak o6	Odsječak o7	Odsječak ...	Odsječak oe
1								
2								
...								
n								



SLIKA 3.1. RAČUNANJE KORELATORSKIH ODNOSA U VREMENSKIM SERIJAMA

Da bismo dobili čitav niz navedenih pokazatelja, potrebno je izračunati korelaciju između nizova vremenskih odsječaka. Statistička mjera korelacije ima za zadatak procjenu međuzavisnosti između nizova vrijednosti promatranih varijabli. Korelacijom procjenjujemo međuzavisnost da li je varijabla A u odnosu na varijablu B zavisna ili nezavisna.

Da bismo dobili odgovor na to pitanje, da li postoji korelacije između varijabli A i B u vrijednosti nizova varijable A i varijable B uvrštavamo u formulu :

$$r_{A,b} = \frac{\sum (A - \bar{A})(B - \bar{B})}{(n-1)\sigma_A \sigma_B}$$

gdje n predstavlja broj elemenata nizova a-potez i b-potez su prosječne vrijednosti A i B, dok sigme označavaju standardne devijacije za A i za B.

Vrijednost korelacije kreće se u intervalu od -1 do +1.

Naredna tablica prikazuje tumačenje vrijednosti r.

Tablica 3.3. Tumačenje koeficijenta korelacije [Han, 2001]

r(A,B)	Tumačenje
<0	Negativna korelacija
0	Nezavisne varijable
>0	Pozitivna korelacija

Što je vrijednost koeficijenta korelacije bliža +1 to je jača pozitivna međuzavisnost između varijabli (ako se poveća A povećat će se i B). Ako je vrijednost koeficijenta korelacije jednaka 0, to znači da ne postoji međuzavisnost među varijablama. Što je vrijednost koeficijenta bliže -1 to je jača negativna međuzavisnost (ako se poveća A povećat će se i B).

Do koje mjere ćemo tolerirati i deklarirati odnos između varijabli kao korelatoran, ovisi o dizajneru sustava za procjenu korelacije.

Tako na primjer možemo formirati pravila :

AKO $r(A,B) \geq 0.6$ TADA postoji pozitivna međuzavisnost ili

AKO $r(A,B) \leq -0.6$ TADA postoji negativna međuzavisnost

U primjeru iz slike 19 varijable koje imaju međusobni stupanj korelacije < 0.6 ili < -0.6 sustav ne bi deklarirao kao varijable između kojih postoji korelacija.

Korelacija u osnovi pruža podatak o međuzavisnosti između dviju pojava predstavljenih u vidu niza vrijednosti varijabli. U slučajevima kada tražimo međuzavisnost utjecaja određene varijable na neku drugu varijablu, tada nam analiza korelacije daje osnovnu informaciju o stupnju međuzavisnosti.

3.5. Pojam sličnosti vremenskih odsječaka

Jedna od glavnih karakteristika REFIL modela jest rješenje koje nudi u segmentu procjeni sličnosti vremenskih odsječaka. Pojam sličnosti je relativno teško određiva kategorija, zbog poteškoća u jednoznačnoj definiciji.

Kada govorimo o problematici sličnosti vremenskih serija tada se prvenstveno fokusiramo na oblik krivulje, i površinu ispod vremenskog odsječka. O cilju analize ovisi koji je zahtijevani stupanj preciznosti, odnosno da li je fokus analize prvenstveno usmjeren na oblik krivulje ili i na kvantitativne aspekte.

Sa perspektive ljudske percepcije čovjek je u vrlo kratkom roku svojim perceptivnim mehanizmima odrediti postojanje sličnosti između dva ili više odsječka vremenske serije. Čovjekove kognitivne sposobnosti dolaze do izražaja na malom uzorku. Zamislimo situaciju u kojoj imamo vremensku seriju koja je segmentirana na n vremenskih odsječaka ($n > 400$). Čak i u slučaju da čovjek svojim perceptivnim mehanizmima promatra ovakvu vremensku seriju koja je grafički prezentirana, nemoguće je pronaći sve slične uzorke. Ako ih i kojim slučajem pronađe kako kvantificirati određene uzorke u vremenskoj seriji, objasniti razloge "sličnosti", i zašto je određeni vremenski odsječak sličniji nekom drugom vremenskom odsječku od nekog trećeg vremenskog odsječka u seriji.

Stvar se sve više komplicira kada radimo analizu sličnosti za n vremenskih serija

Sličnost možemo definirati preko pojma intervalne jednakosti vremenskih odsječaka.

S obzirom na cilj analize možemo postaviti uvjet sličnosti koji govori da su dva odsječka slična ako postoji intervalna jednakost između otklona kutova odsječaka.

Isto tako možemo postaviti uvjet sličnosti koji govori da su dva odsječka slična ako postoji intervalna jednakost između otklona kutova odsječaka i postoji intervalna jednakost površina ispod krivulje odsječka.

Pretpostavimo da vremenski odsječak T_1 , vremenske kompleksnosti od jedne vremenske jedinice ima vrijednost koeficijenta kutnog otklona 0.8, a vremenski odsječak T_2 , vremenske kompleksnosti od jedne vremenske jedinice ima vrijednost koeficijenta kutnog otklona 0.85. Pretpostavimo da je intervalno zadana kategorija "oštrog rasta" u intervalu od 0.7-1.0.

Prema kriteriju sličnosti postojanja intervalne jednakosti između otklona kutova odsječaka gdje je T_1 i T_2 spadaju u kategoriju "oštrog rasta", zaključujemo da su odsječak T_1 i T_2 slični.

U kriterij ocjenjivanja sličnosti možemo uvesti i površinu ispod krivulje, te istom metodologijom određivati kategorije i procjenjivati intervalnu jednakost površine.

U tom je slučaju potrebno definirati uvjete prihvatanja i odbijanja tvrdnji o sličnostima vremenskih odsječaka T_1 i T_2 .

AKO $k(1)_{\text{intervalno}} = k(2)_{\text{intervalno}}$ i $P(1)_{\text{intervalno}} \neq P(2)_{\text{intervalno}}$
 $\Rightarrow T_1$ slično T_2

AKO $k(1)_{\text{intervalno}} = k(2)_{\text{intervalno}}$ i $P(1)_{\text{intervalno}} = P(2)_{\text{intervalno}}$
 $\Rightarrow T_1$ slično T_2

AKO $k(1)_{\text{intervalno}} \neq k(2)_{\text{intervalno}}$ i $P(1)_{\text{intervalno}} \neq P(2)_{\text{intervalno}}$
 $\Rightarrow T1$ nije slično $T2$

AKO $k(1)_{\text{intervalno}} \neq k(2)_{\text{intervalno}}$ i $P(1)_{\text{intervalno}} = P(2)_{\text{intervalno}}$
 $\Rightarrow T1$ nije slično $T2$

Ovako postavljena pravila rješavaju problematiku sličnosti na razini jednog vremenskog odsječka vremenske kompleksnosti od jedne vremenske jedinice. Kod procjene sličnosti dijela vremenskih serija, ili vremenskih serija u cjelini, potrebno je algoritamskim putem procjenjivati sličnosti na osnovu definirane metodologije.

3.6. Pojmovne strukture i fuzzy logika

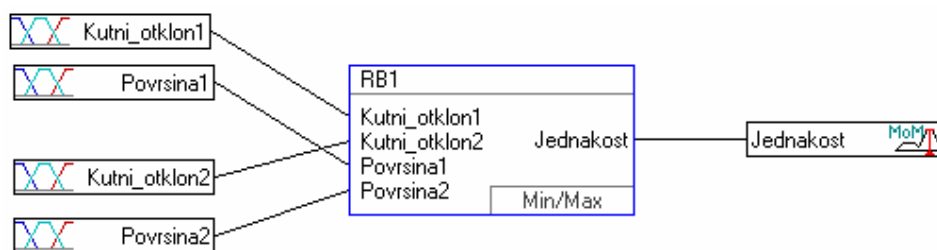
Pojmovne strukture u svjetlu fuzzy logike mogu se definirati kroz jednostavan fuzzy ekspertni sustav.

Kroz sustav pravila (eng. Rule block) definiraju se pojmovi poput jednakosti, sličnosti i inverznosti.

Kao ilustrativni primjer prikazan je fuzzy ekspertni sustav za klasifikaciju odsječaka vremenske serije po kriteriju jednakosti.

Ovakav sustav može zamijeniti proračunski mehanizam procjene jednakosti baziran na *crisp* logici, a koji je korišten kod empirijskih istraživanja kod pretraživanja uzoraka, traženja motiva i slično.

Ulazni parametri u takav sustav su korespondentni vremenski odsjecci predstavljeni u vidu fuzzy varijabli kao što je to prikazano na slici 3.2 :

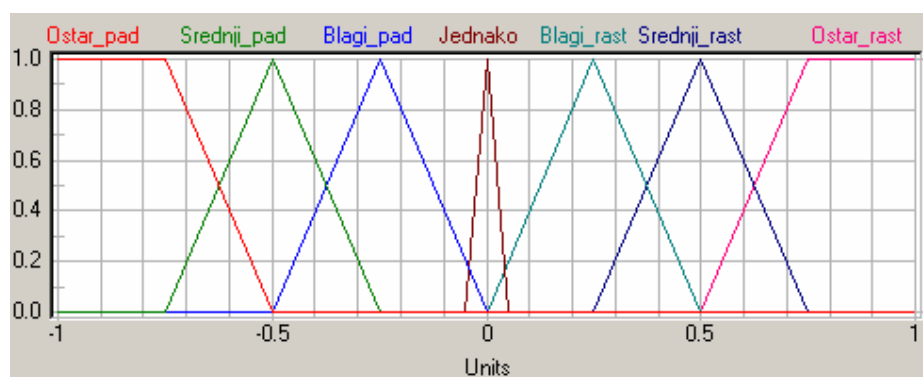


SLIKA 3.2. FUZZY SUSTAV ZA PROCJENU JEDNAKOSTI ODSJEČAKA

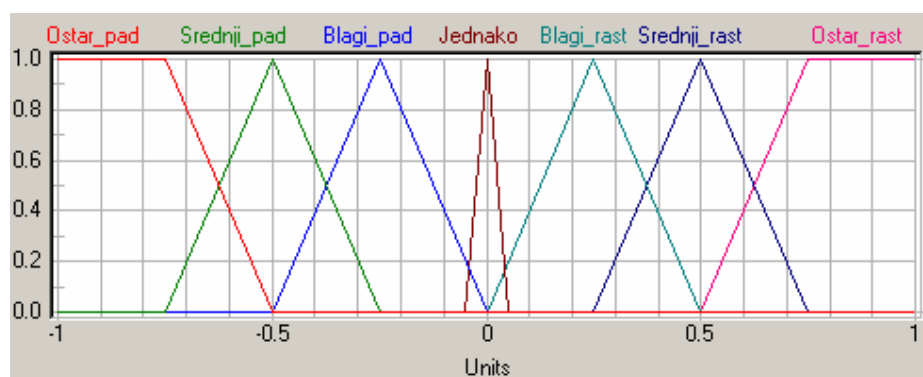
Varijable *kutni_otklon1* i *Povrsina1* odnose se na *n*-ti vremenski odsječak iz prve vremenske serije za koju procjenjujemo jednakost, a *kutni_otklon2* i *Povrsina2* na *n*-ti vremenski odsječak iz druge vremenske serije za koju procjenjujemo jednakost. Izlazna varijabla "Jednakost" klasificira odsječke kao jednake ili nejednake. Korištena defuzifikacijska metoda za ovu varijablu je *Mean of Maximum* (MoM).

Lingvističke varijable možemo definirati kao na slici 3.3. :

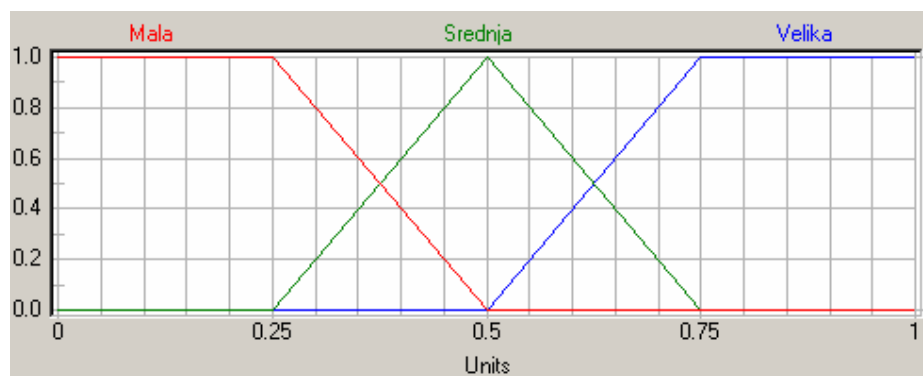
"Kutni_otklon1"



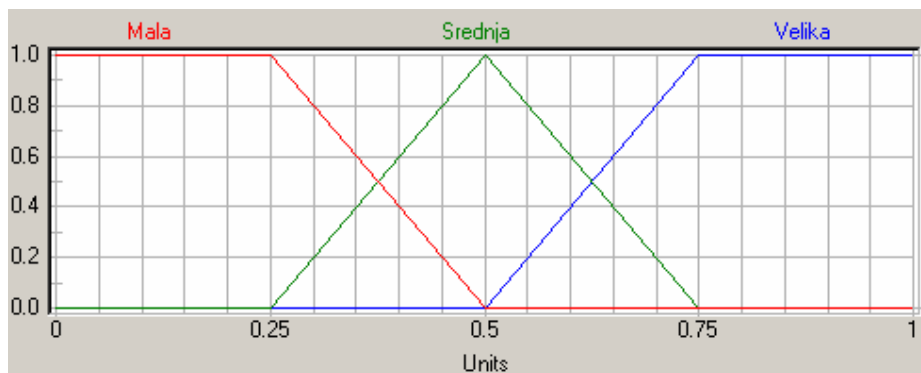
"Kutni_otklon2"



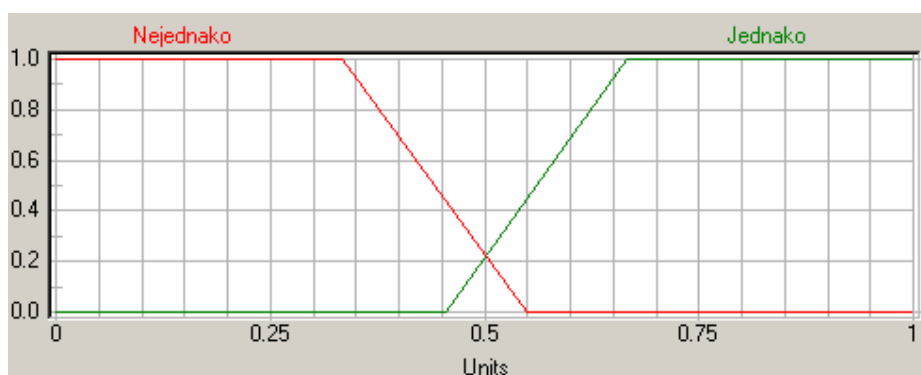
"Povrsina1"



"Povrsina2"



"Jednakost"



SLIKA 3.3. DEFINICIJA FUZZY VARIJABLI U SUSTAVU ZA PROCJENU JEDNAKOSTI ODSJEČAKA

Jednakost vremenskih odsječaka definiramo sustavom pravila kakav je prikazan u prilogima pod naslovom "Fuzzy definicija jednakosti odsječaka".

Identičan fuzzy sustav pravila sa izmijenjenim vrijednostima pravila u THEN klauzuli može se iskoristiti za definiciju sličnosti, pa je moguće definirati sličnost unutar istog sustava.

Sličnost unutar istog sustava možemo definirati u vidu pravila kao primjerice :

Odsječak 1= Oštar rast AND Odsječak 2= Oštar rast AND Odsječak 1= Velika površina AND Odsječak 2= Srednja površina THEN Slično

Kao rezultat obrade dobivamo klasifikaciju parova vrijednosti promatranih vremenskih odsječaka.

Fuzzy sustav definiran na ovaj način može se integrirati unutar rešetaka samoorganizirajućih mapa koje u procesima primjerice pretraživanja uzoraka imaju zadatak klasificirati uzorke koristeći fuzzy operatore, a ne crisp operatore, kao što je to slučaj u provedenim empirijskim istraživanjima za potrebe rada.

Isto tako je moguće i ostale algoritme dizajnirane na temelju crisp logike za potrebe empirijskih istraživanja prikazane kroz rad bazirati na fuzzy logici (npr. Sličnost, traženje motiva ...)

4. Konceptualna primjena modela

4.1. REFII model i konstrukcija problemskih rješenja

Izraženi problemi u tradicionalnoj *data mining* analizi vremenskih serija su problemi automatskog pretprocesiranja vremenske serije [Pyle, 2001], te problemi ulančavanja niza analitičkih metoda u okviru jedne analize, kao i problemi primjene tradicionalnih *data mining* metoda nad vremenskom serijom i uključivanja nevremenskih atributa kao elementa koji ulaze u proces analize vremenskih serija.

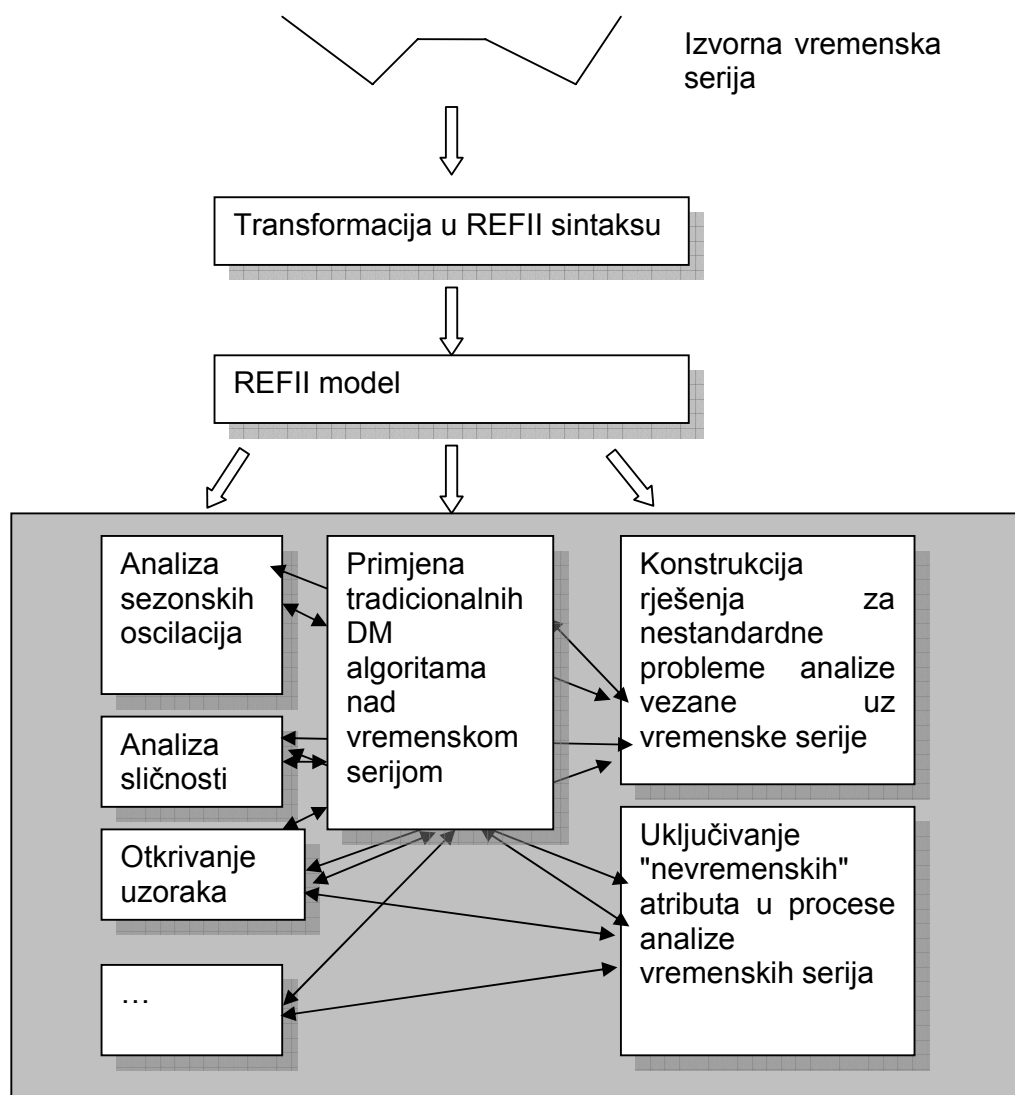
Isto tako, tradicionalna *data mining* analiza vremenskih serija sklona je primjeni različitih nekompatibilnih modela transformacije vremenske serije ovisno o primijenjenom algoritmu.

Osnovna karakteristika REFII modela je svojstvo jednoznačne transformacije vremenske serije, koja je temelj za konstrukciju algoritamskih rješenja s obzirom na problemsku domenu.

Nakon provedene transformacije vremenske serije u REFII model, nad tako transformiranom vremenskom serijom moguće je izgraditi rješenja za specifične problemske prostore, ali je isto tako moguće na model primijeniti tradicionalne *data mining* metode. Na ovaj način je riješen problem objedinjavanja različitih metoda za analizu vremenskih serija, u okviru istog analitičkog postupka, odnosno moguće je otkrivene uzorke vremenske serije podvrći lančanom procesu analize u okviru istog analitičkog postupka.

Ukratko ovaj model anulira probleme proizašle iz tradicionalnog pristupa, spomenute na početku ovog poglavlja, te je otvoren za kreiranje novih algoritamskih postupaka za analizu vremenske serije ovisno o problemskom prostoru.

Slika 4.1. prikazuje temeljnu ulogu REFII modela u *data mining* analizi vremenskih serija.



SLIKA 4.1. TEMELJNA ULOGU REFII MODELA U DATA MINING ANALIZAMA VREMENSKIH SERIJA

Kao što je vidljivo iz slike 4.1. ovakva metodologija omogućava međusobnu interakciju metoda, što dovodi do sinergijskog efekta. Izlazni rezultati proizašli iz jedne metode mogu postati ulazni parametri za narednu metodu.

Kao što je vidljivo iz slike u okviru ovakve koncepcije moguće je ravnopravno koristiti metode *data mininga* nad vremenskom serijom sa algoritmima za rješavanje standardiziranih problema poput otkrivanja uzoraka u vremenskoj seriji, te algoritmima koji rješavaju nestandardizirane probleme u okviru vremenskih serija.

Naredna poglavlja prikazat će metodologiju konstrukcije algoritama na primjerima otkrivanja sezonskih oscilacija, otkrivanja uzoraka unutar vremenske serije te procjene sličnosti vremenskih serija.

4.2. Otkrivanje sezonskih oscilacija

4.2.1. Otkrivanja sezonskih oscilacija na temelju REFII modela

Sezonsku oscilaciju unutar vremenske serije možemo definirati kao očekivanu pravilnost s obzirom na period promatrane pojave.

Gledano sa perspektive otkrivanja uzoraka, sezonsku oscilaciju možemo definirati kao očekivani uzorak u točno određenom periodu vremena. Za razliku od problema otkrivanja uzoraka, kod sezonskih oscilacija unutar vremenske serije uzimamo u obzir i vrijeme pojave uzorka.

Sukladno temeljnoj ideji REFII modela, algoritam za otkrivanje sezonskih oscilacija kao temelj uzima vremensku seriju transformiranu u REFII notaciju, odnosno temelj za analizu mu je matrica transformacije.

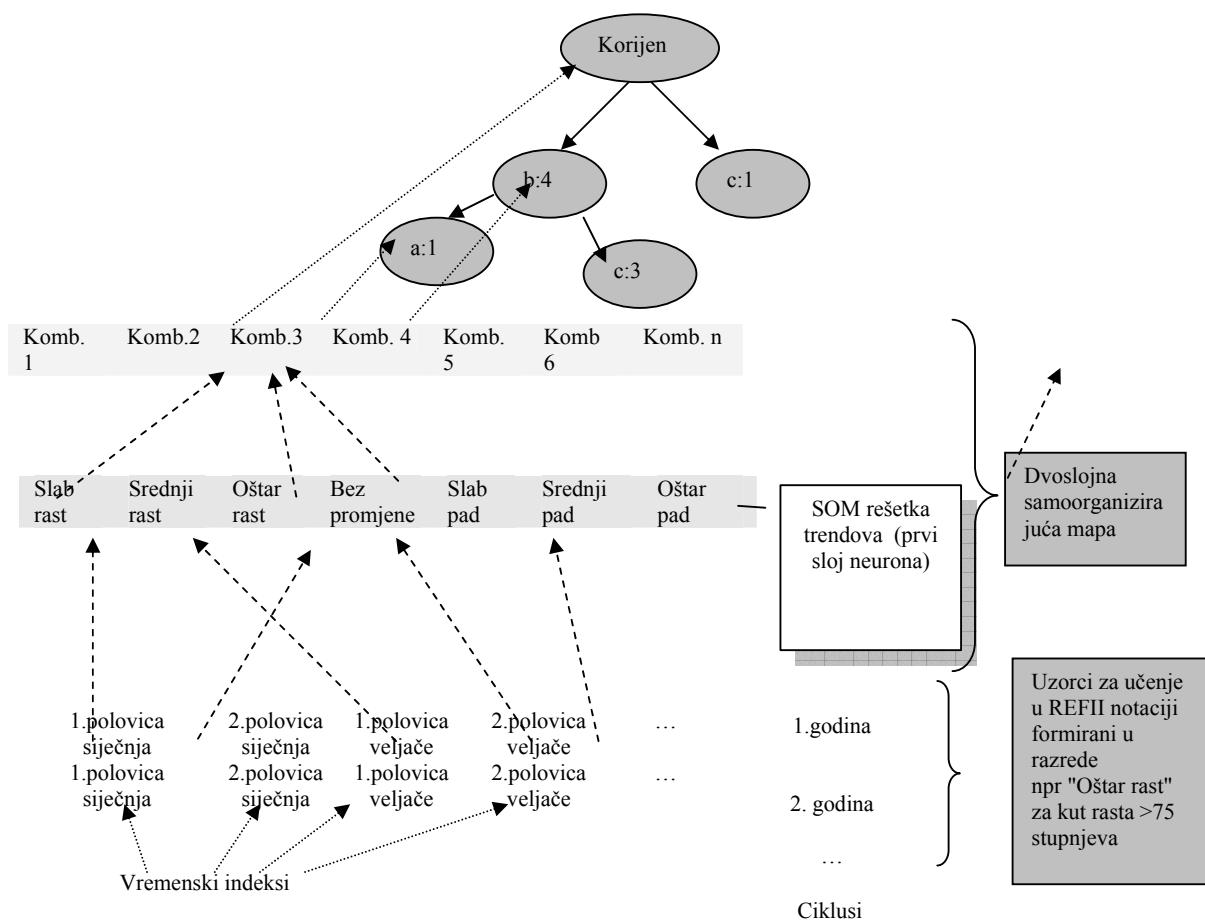
Kao temeljnu inspiraciju konstrukcije algoritma za otkrivanje sezonskih oscilacija i traženje uzoraka u vremenskoj seriji, korišteni su elementi samoorganizirajućih mapa. Sličan pristup, ali bez upotrebe samoorganizirajućih mapa i jedinstvenog modela transformacije vremenske serije vidljiv je u Pollyja i Wonga radu [Polly, 2001].

Osnovna ideja svodi se na formiranje vremenske rešetke za filtriranje uzoraka. U slučaju sezonskih oscilacija u obzir se uzima i vremenski indeks pojavnosti nekog uzorka, dok uzimanje vremenskog indeksa kao relevantnog faktora prilikom pretraživanja uzoraka ovisi o karakteru analiza.

Inicijalna dužina promatranog uzorka iznosi jednu vremensku jedinicu. Nakon proteka ciklusa, računaju se koeficijenti sigurnosti, te u naredni ciklus ulaze kandidati koji imaju zadovoljavajuće iznose koeficijenata sigurnosti. U narednom ciklusu dužina promatranog uzorka se povećava za jednu vremensku jedinicu. Postupak se ponavlja dokle god ima kandidata za ulazak u naredni ciklus, odnosno dok dužina uzorka ne postane jednaka dužini vremenske serije.

Ova problematika zahtijeva hibridnu dvoslojnu arhitekturu samoorganizirajućih mapa, te upotrebu stabla frekventnih uzoraka kojem je autor Han [Han, 2000]. Han osim za nesekvencijalne uzorke [Han, 2000] koristi stablo frekventnih uzoraka i za pretragu sekvencijalnih uzoraka [Han, 2004]. Stablo frekventnih uzoraka u modelu otkrivanja sezonskih oscilacija i otkrivanja uzoraka unutar vremenske serije na način koji je prikazan u ovome radu, prilagođeno je REFII modelu i konkretnim potrebama s obzirom na ciljeve analiza.

Shematski, ovaj postupak može se prikazati slikom 4.2.



SLIKA 4.2. OTKRIVANJE SEZONSKIH OSCILACIJA NA TEMELJU REFII MODELA

Kao što je vidljivo iz prikazane sheme temeljna strategija svodi se na korištenje intervalnih vrijednosti kao elemenata usporedbe uzoraka.

Za razliku od klasičnih rješenja za otkrivanje uzoraka, kod otkrivanja sezonskih oscilacija vrlo bitan element koji se uzima u obzir prilikom analize je i vrijeme otkrivanja događaja.

To vrijeme unutar REFII modela deklarira se kao vremenski indeks, i pojavnost uzoraka se promatra sa perspektive vremenskog indeksa, i ciklusa odvijanja.

Bitna mjera kod ove analize o kojoj ovise krajnji rezultati, je koeficijent pouzdanosti. Ona definira do koje frekvencije pojavnosti se neki uzorak uzima kao relevantan s obzirom na vremenski indeks. Primjerice ako je zahtijevani koeficijent pouzdanosti 0.95, a "Oštar rast" se pojavljuje u manje od 95% slučajeva s obzirom na vremenski indeksa, hipoteza o postojanju sezonske oscilacije za ovaj vremenski period će biti odbačena. U slučaju prihvatanja hipoteze o sezonskoj oscilaciji za ovaj vremenski period prihvaća se hipoteza o postojanju sezonske oscilacije za ovaj vremenski period, a izgenerirani uzorak ulazi kao kandidat u naredni proces obrade.

Razlog generiranja kandidata i provjeravanja relevantnosti uzoraka "2 i više", provodi se iz razloga što i uzorak "A" kod kojeg je prihvaćena hipoteza o postojanju sezonske oscilacije i uzorak "B" kod kojeg je prihvaćena hipoteza o postojanju sezonske

oscilacije ne znači da u paru odnosno kod dužih uzoraka skupno zadovoljavaju kriterij po pitanju koeficijenta pouzdanosti.

Na primjer ako uzorci "A" i "B" dužine 1, sa indeksima t i $t+1$, imaju koeficijent pouzdanosti 0.91 i 0.9, a minimalni zahtijevani koeficijent pouzdanosti je 0.9, tada treba izračunati koeficijent pouzdanosti para AB.

Ako je koeficijent pouzdanosti para A i B dužine 2 minimalno 0.9, tada se prihvaća hipoteza o postojanju sezonske oscilacije "AB" na intervalu $< t, t+1 >$.

Isto tako možemo računati vjerojatnost dešavanja uzorka B kao p ako se "desio" uzorak A prema formuli za uvjetnu vjerojatnost:

$$P(A|B) = p$$

Procjena uvjetnih vjerojatnosti ne samo da nam može pomoći kod postavljanja hipoteza o sezonskoj oscilaciji nad uzorcima na intervalima većima od jedne jedinice, nego nam može biti od koristi i kod otkrivanja potencijalno najfrekventnijih kandidata koje možemo očekivati u narednom ciklusu s obzirom na vremenski indeks.

Kako bi se optimiziralo "učenje" u algoritam baziran na REFII modelu može se ugraditi prihvaćanje hipoteze za uzorke dužine >1 uz uvjet da je izračunati koeficijent pouzdanosti jednak 1.

Daljnji kriterij za generiranje uzoraka dužine >1 je uvjet postojanja vremenskog kontinuiteta, odnosno vremenske neprekinutosti uzoraka čiji je koeficijent pouzdanosti jednak 1 kao što je to u primjeru prikazano narednom tablicom.

Tablica 4.1. Tablica vremenskih kontinuiteta sa pripadajućim koeficijentima pouzdanosti

Vremenski indeks 1	Vremenski indeks 2	Vremenski indeks 3	Vremenski indeks 4	Vremenski indeks n
R -koeficijent pouzdanosti=1	R -koeficijent pouzdanosti=0	R -koeficijent pouzdanosti=0.25	R -koeficijent pouzdanosti=1	R -koeficijent pouzdanosti=0
E -koeficijent pouzdanosti=0	E -koeficijent pouzdanosti=1	E -koeficijent pouzdanosti=0.25	E -koeficijent pouzdanosti=0	E -koeficijent pouzdanosti=1
F -koeficijent pouzdanosti=0	F -koeficijent pouzdanosti=0	F -koeficijent pouzdanosti=0.5	F -koeficijent pouzdanosti=0	F -koeficijent pouzdanosti=0

Na temelju tablice moguće je prihvatiti hipotezu o postojanju sezonske oscilacije R(vremenski indeks 1),E (vremenski indeks 1), odnosno "RE", te o postojanju sezonske oscilacije R na poziciji vremenskog indeksa 4.

Nadalje, moguće je definirati događaje [Manilla, 1997] na temelju otkrivenih uzoraka, te ih spojiti u kauzalne lance gdje elementi kauzalnog lanca ne moraju nužno biti samo uzorci iz vremenske serije.

Tip takve analize može se ilustrirati analizom modela ponašanja klijenta banke, gdje se analizira da li primjerice kontinuirani pad tečaja dolara u periodu od n i više jedinica vremena rezultira aktivnost konverzije valute na deviznom računu ako su ta sredstva u dolarima.

Ako se primjerice dokaže takav model ponašanja, tada banka u okviru CRM (eng. Customer Relationship Management) sustava može preko kontaktnog centra kontaktirati klijente i sugerirati im konverziju valute. Ovo je ilustrativni primjer kako analiza vremenskih serija može imati direktne implikacije na sustave poput CRM-a.

Kako bi REFII model trebao pružiti veću fleksibilnost od tradicionalnih modela između ostalog i za otkrivanje sezonskih oscilacija, on mora pružiti dodatne analitičke

mogućnosti po pitanju traženja pravilnosti i unutar vremenskih intervala koji nisu prihvaćeni kao intervali sa sezonskom pravilnošću. Tako primjerice, ako na intervalu vremenskog indeksa 2 (dužina intervala =1) koji označava period između siječnja i veljače postoji distribucija frekvencija uzoraka :

"Slab rast" – koeficijent pouzdanosti 0.3333333

"Srednji rast" – koeficijent pouzdanosti 0.3333333

"Visok rast" – koeficijent pouzdanosti 0.3333333

Pri čemu se neka pojava promatra od 1940-2000, moguće je pokušati tražiti pravilnost na način da li je unutar određenog kontinuiranog razdoblja postojao stalan trend, ili se je on možda izmjenjivao prema nekoj pravilnosti, je moguće tražiti sezonsku pravilnost i na razini samog uzorka.

Element površine ispod krivulje, može nam pomoći u otkrivanju trendova pojava. To znači da ako je na razini vremenskog indeksa 1 otkrivena sezonska oscilacija sa koeficijentom pouzdanosti 0.95, tada na temelju trendova površina ispod odsječka možemo procijeniti da li je trend opsega pojave primjerice u stalnom padu, stalnom rastu ili je ujednačen.

Analizu trenda opsega pojave možemo provesti po obrascu :

Trend rasta opsega sezonske pojave :

$$\text{ciklus}(i).\text{površina ispod odsječka}(j) < \text{ciklus}(i+1).\text{površina ispod odsječka}(j) < \text{ciklus}(i+2).\text{površina ispod odsječka}(j) < \dots < \text{ciklus}(i+n).\text{površina ispod odsječka}(j)$$

Trend pada opsega sezonske pojave :

$$\text{ciklus}(i).\text{površina ispod odsječka}(j) > \text{ciklus}(i+1).\text{površina ispod odsječka}(j) > \text{ciklus}(i+2).\text{površina ispod odsječka}(j) > \dots > \text{ciklus}(i+n).\text{površina ispod odsječka}(j)$$

Trend ujednačenosti opsega sezonske pojave :

$$\text{ciklus}(i).\text{površina ispod odsječka}(j) = \text{ciklus}(i+1).\text{površina ispod odsječka}(j) = \text{ciklus}(i+2).\text{površina ispod odsječka}(j) = \dots = \text{ciklus}(i+n).\text{površina ispod odsječka}(j)$$

Pri čemu i označava broj ciklusa a j frekventni promatrani vremenski odsječak s obzirom na promatrani ciklus. Objekt ciklus(i).površina ispod odsječka(j) se može izraziti preko intervalnih i implicitnih jednakosti.

U modelu procjene trenda možemo izraziti koeficijent trenda sezonskog odsječka/odsječaka kao omjer između najfrekventnije kategorije i ukupnog broja kategorija vezanih uz površinu ispod krivulje. Ako je koeficijent \geq zadanoj vrijednosti za prihvatanje hipoteze, tad prihvaćamo hipotezu o sezonskom trendu.

4.2.2. Algoritam rješenja

U nastavku je prikazan algoritam za otkrivanje sezonskih oscilacija posredstvom REFII modela

Transforimiraj vremensku seriju u REFII model

*Formiraj vremensku rešetku na temelju distribucije razreda otklona kutova i /ili površine
Definiraj koeficijente pouzdanosti*

Dužina uzorka=1

Lista kandidata= svi elementi vremenske serije u REFII notaciji

Dokle ima kandidata u listi kandidata

za element iz liste kandidata pronađi jednaki element u vremenskoj rešetki s obzirom na definiranu Dužinu uzorka

kreiraj listu frekvencija pojavnosti na način: ako element iz rešetke ne postoji u listi frekvencija kreiraj ga odnosno, ako postoji uvećaj ga za 1

Izdvoji sve one elemente iz liste frekvencije pojavnosti čiji je koeficijent pouzdanosti veći od zadanog i napuni ga u listu kandidata.

Dužina uzorka= Dužina uzorka+1

Uključi elemente uzorka čiji je koeficijent pouzdanosti veći od zadanog u stablo frekventnih uzoraka

Izdvoji elemente iz liste frekvencije pojavnosti čiji je izračunati koeficijent pouzdanosti =100 provjeri njihov kontinuitet i proglasi ih sezonskom oscilacijom

Postupak ponavlja dokle god ima kandidata

Procjeni trendove ujednačenosti opsega sezonskih pojava

Uđi u analizu pravilnosti za uzorke koji nemaju zadovoljavajući koeficijent pouzdanosti

Ovaj proces je moguće ilustrirati na narednom jednostavnom primjeru, gdje se vremenske serije u REFII notaciji nalaze u tablici 4.2.

Tablica 4.2. Tablica pretprocesiranih vremenskih serija u REFII notaciji za procjenu sezonskih oscilacija

	Ponedjeljak	Utorak	Srijeda	Četvrtak	Petak
1.tjedan	Nizak pad, Mala površina	Nizak rast, Mala površina	Nizak pad, Mala površina	Nizak rast, Mala površina	Nizak pad, Mala površina
2.tjedan	Nizak pad, Mala površina	Nizak rast, Mala površina	Nizak rast, Mala površina	Nizak rast, Mala površina	Nizak pad, Mala površina
3.tjedan	Nizak pad, Mala površina	Nizak rast, Mala površina	Jednako, Mala površina	Nizak rast, Mala površina	Nizak pad, Mala površina
4. tjedan	Nizak pad, Mala površina	Nizak rast, Mala površina	Visok rast, Srednja površina	Nizak rast, Mala površina	Nizak pad, Mala površina
5.tjedan	Nizak pad, Mala površina	Nizak rast, Mala površina	Srednji rast, Srednja površina	Nizak rast, Mala površina	Nizak pad, Mala površina

Prema prikazanom algoritmu formirana je rešetka (SOM rešetka) prikazana tablicom 4.3.

Tablica 4.3. Struktura SOM rešetke

	Visok pad	Srednji pad	Nizak pad	Jednako	Nizak rast	Srednji rast	Visok rast
Mala površina							
Srednja površina							
Velika površina							

U prvom koraku algoritam traži frekventne pojave na razini dana uz pomoć SOM rešetke. Pretpostavimo da je željeni stupanj pouzdanosti jednak 1. Tako se na temelju prikazanih podataka u prvom koraku formiraju frekventni uzorci dužine 1 za ponedjeljak (Nizak pad, Mala površina) 5/5, Utorak (Nizak rast, Mala površina) 5/5, Četvrtak (Nizak rast, Mala površina) 5/5, Petak (Nizak pad, mala površina) 5/5.

Svaki od frekventnih uzoraka dobiva svoj čvor bez poveznica u stablu frekventnih uzoraka. U drugom koraku traže se frekventni uzorci dužine 2. Kandidati za uzorke dužine 2 su frekventni uzorci sa glavama dužine jedan. U strukturi hijerarhijske samoorganizirajuće mape, koja se sastoji od dva sloja, gdje se u drugom sloju skladište frekventni kandidati, u svakom prolazu se nakon zadovoljavanja kriterija pouzdanosti pokušava povećati dužina uzorka sa frekventnim kandidatima. Taj se postupak ponavlja dokle ima novih kandidata.

U našem slučaju pojavljuju se uzorci sezonskih oscilacija ponedjeljak (Nizak pad, Mala površina), Utorak (Nizak rast, Mala površina), Četvrtak (Nizak rast, Mala površina), Petak (Nizak pad, mala površina).

Važno je napomenuti kako ovako formirana struktura vodi računa primjerice o tome ako je zadani koeficijent pouzdanosti 0.9, i parcijalno gledajući na razini svakog odsječka on može zadovoljavati taj kriterij, no ako se promatra uzorak dužine 2 usprkos tome pouzdanost ne mora biti 0.9. Isto se tako vodi računa, kao što je vidljivo iz ovog primjera, da u slučaju prekida kontinuiteta (srijeda) u sezonskim oscilacijama, prema postupku formiranja stabla frekventnih uzoraka [Han, 2000], u konačnici se pojavljuju dva stabla čiji su korijeni u ovom primjeru vrijednosti iz ponedjeljka i četvrtka.

4.2.3. Otkrivanje sezonskih oscilacija na empirijskim podacima

Podaci koji su korišteni za ovo istraživanje sezonskih oscilacija preuzeti su sa : <ftp://ftp.nist.gov/pub/dataplot/other/reference/MLCO2MON.DAT>.

Predstavljaju mjesečni prosjek koncentracije ugljičnog dioksida (CO₂) mjerenog infracrvenom spektrometrijom u razdoblju od 1974-1985 na Mauna opservatoriju.

Struktura podataka :

1. CO₂ koncentracija
2. Godina+dio godine
3. Godina
4. Mjesec

U prvom koraku podaci su pretprocesirani programom "Time explorer" razvijenim u Visual Fox pro 8.0 programskom jeziku. Analiza je provedena na Pentiumu 1.6 GHz, 512 MB RAM-a, proces obrade trajao je 1.5 sekundi za transformaciju u REFII model, dok je analiza sezonskih oscilacija trajala 2 sekunde.

U ovoj etapi formirana je matrica transformacije sa pripadajućim intervalnim vrijednostima otklona kutova na temelju formiranih razreda zadanih od strane korisnika, izračunate su vrijednosti površine ispod svakog odsječka, pridodijeljena je REF oznaka i kreirani su indeksi ciklusa i vremenski indeksi.

Zadani koeficijent pouzdanosti iznosi 0.95.

Razredi otklona kutova definirani su na slijedeći način:

Tablica 4.4. "Crisp " razredi otklona kutova

Razred	Donja granica	Gornja granica	REF
Nizak rast	0.000000000	0.300000000	R
Srednji rast	0.300000000	0.700000000	R
Visok rast	0.700000000	1.000000000	R
Nizak pad	0.000000000	0.300000000	F
Srednji pad	0.300000000	0.700000000	F
Oštar pad	0.700000000	1.000000000	F
Bez promjene	0.000000000	0.000000000	E

Tako pripremljeni podaci procesirani su modulom u "Time explorer-u" za otkrivanje sezonskih oscilacija čiji je algoritam predstavljen u prethodnom poglavlju. Kao rezultat prvog koraka analize dobivena je naredna tablica :

Tablica 4.5. Sezonske oscilacije

		Kutni otklon		
		Bez promjene	Nizak pad	Nizak rast
	1	,0%	,0%	100,0%
	2	,0%	,0%	100,0%
	3	,0%	,0%	100,0%
	4	,0%	,0%	100,0%
	5	,0%	7,7%	92,3%
Vremenski indeks	6	,0%	85,7%	14,3%
	7	,0%	100,0%	,0%
	8	,0%	100,0%	,0%
	9	,0%	100,0%	,0%
	10	7,7%	46,2%	46,2%
	11	,0%	,0%	100,0%
	12	,0%	,0%	100,0%

Oznake vremenskog indeksa npr. 1. označava period sa siječnja na veljaču, 2 označava period sa veljače na ožujak itd....

S obzirom na to da se nakon proračuna frekvencija aktivirao dio algoritma na temelju kojeg ako je koeficijent pouzdanosti 1 (100% u tablici) odsječak unutar vremenskog indeksa procjenjuje se kao sezonska oscilacija, te se provjerava kontinuitet na vremenskom intervalu, kao uzorci sezonske oscilacije izgenerirani su :

Tablica 4.6. Tablica sezonskih uzoraka

Period	Sezonska oscilacija
Siječanj-Veljača	Nizak rast
Veljača – Ožujak	Nizak rast
Ožujak - Travanj	Nizak rast
Travanj-Svibanj	Nizak rast
Srpanj-Kolovoz	Nizak pad
Kolovoz-Rujan	Nizak pad
Rujan-Listopad	Nizak pad
Studen-Prosinac	Nizak rast
Prosinac-Siječanj	Nizak rast

Zbog zadanog koeficijenta pouzdanosti od 0.95 u narednu etapu analize nisu ušli ostali kandidati.

Brz proračun frekvencija sezonskih uzoraka bez dublje analize međuzavisnosti, moguće je nad transformiranom vremenskom serijom u REFII model izvršiti naredbom SPSS skriptnog jezika :

```
CTABLES
/VLABELS VARIABLES=razred indeks DISPLAY=DEFAULT
/TABLE indeks [C][ROWPCT.COUNT PCT40.1] BY razred
/CATEGORIES VARIABLES=razred indeks ORDER=A KEY=VALUE EMPTY=EXCLUDE.
```

Ovakvim pristupom ne generira se stablo frekventnih uzoraka, i ne postoji mogućnost provođenja dubljih analiza.

Da bi se provela dublja analiza nad nefrekventnijim uzorcima, te se pokušala otkriti eventualna zakonitost njihovog ponašanja, jedna od mogućnosti je promatranje nefrekventnih uzoraka s obzirom na ciklus i modalitete pojavnosti kao što je to prikazano u narednoj tablici :

Tablica 4.7. Tablica sezonskih uzoraka s obzirom na cikluse

				Kutni otklon		
				Bez promjene	Nizak pad	Nizak rast
	5	Ciklus	1975	,0%	,0%	100,0%
			1976	,0%	100,0%	,0%
			1977	,0%	,0%	100,0%
			1978	,0%	,0%	100,0%
			1979	,0%	,0%	100,0%
			1980	,0%	,0%	100,0%
			1981	,0%	,0%	100,0%
			1982	,0%	,0%	100,0%
			1983	,0%	,0%	100,0%
			1984	,0%	,0%	100,0%
			1985	,0%	,0%	100,0%
			1986	,0%	,0%	100,0%
			1987	,0%	,0%	100,0%
	6	Ciklus	1974	,0%	100,0%	,0%
			1975	,0%	100,0%	,0%
			1976	,0%	100,0%	,0%
			1977	,0%	100,0%	,0%
			1978	,0%	,0%	100,0%
Vremenski indeks			1979	,0%	,0%	100,0%
			1980	,0%	100,0%	,0%
			1981	,0%	100,0%	,0%
			1982	,0%	100,0%	,0%
			1983	,0%	100,0%	,0%
			1984	,0%	100,0%	,0%
			1985	,0%	100,0%	,0%
			1986	,0%	100,0%	,0%
			1987	,0%	100,0%	,0%
	10	Ciklus	1974	,0%	100,0%	,0%
			1975	,0%	100,0%	,0%
			1976	,0%	100,0%	,0%
			1977	,0%	100,0%	,0%
			1978	,0%	100,0%	,0%
			1979	,0%	,0%	100,0%
			1980	100,0%	,0%	,0%
			1981	,0%	,0%	100,0%
			1982	,0%	,0%	100,0%
			1983	,0%	,0%	100,0%
			1984	,0%	,0%	100,0%
			1985	,0%	,0%	100,0%
			1986	,0%	100,0%	,0%

Na osnovu provedene analize, vidljivo je da vremenski indeks 5 zadržava trend niskog rasta, i da je trend niskog pada u tom periodu zabilježen na počecima perioda opservacije 1976 godine, te se na osnovu ovih podataka može ustvrditi daljnji trend niskog razdoblja u ovom razdoblju.

Kod vremenskog indeksa 6, može se prognozirati daljnji trend niskog pada, jer je nizak rast zabilježen samo 1978 i 1979.

Najrizičniji vremenski indeks za procjenu je indeks 10 gdje postoji stanovita pravilnost po pitanju trendova, ali je relativno teško na temelju provedene analize sa većom sigurnošću predviđati daljnje trendove.

Posljednji dio analize posvećen je trendovima opsega frekventnih odsječaka, koji se provodi na temelju površina ispod vremenskih odsječaka. Kao rezultat analize dobiveni su slijedeći rezultati:

Tablica 4.8. Tablica sezonskih uzoraka s obzirom na trendove

Period	Sezonska oscilacija
Siječanj-Veljača	Kontinuirani rast opsega, pouzdanost 1
Veljača – Ožujak	Kontinuirani rast opsega, pouzdanost 1
Ožujak - Travanj	Kontinuirani rast opsega, pouzdanost 1
Travanj-Svibanj	Kontinuirani rast opsega, pouzdanost 1
Srpanj-Kolovoz	Kontinuirani rast opsega, pouzdanost 1
Kolovoz-Rujan	Kontinuirani rast opsega, pouzdanost 1
Rujan-Listopad	Kontinuirani rast opsega, pouzdanost 1
Studen-Prosinac	Kontinuirani rast opsega, pouzdanost 1
Prosinac-Siječanj	Kontinuirani rast opsega, pouzdanost 1

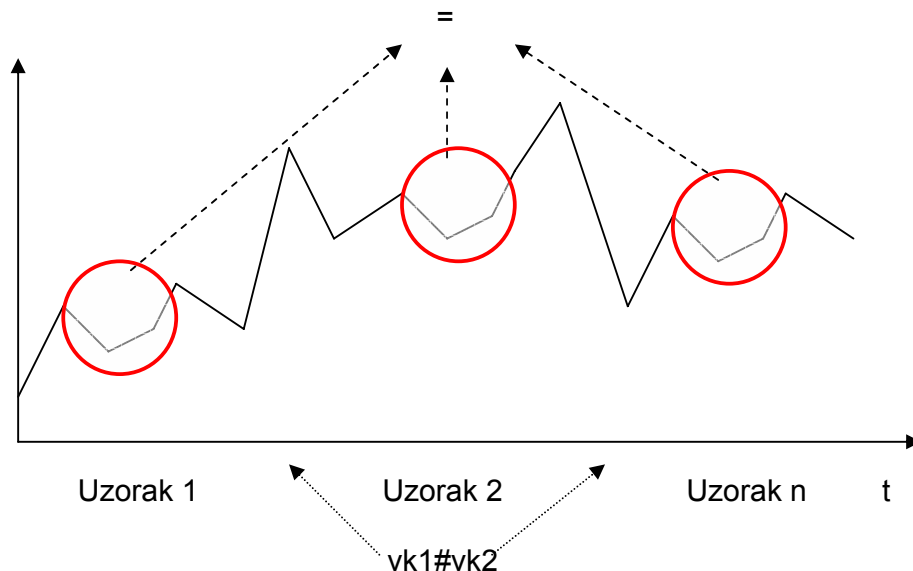
Generalno gledajući postoji sezonska oscilacija koncentracije mjenog ugljičnog dioksida na opservatoriju Mauna, te se bilježi trend rasta koncentracije ugljičnog dioksida sa protekom vremena (voluminoznost raste).

Iako je "na prvi pogled" moguće dobar dio ovih rezultata provesti vizualizacijom vremenske serije, naravno ne toliko precizno, i bez mogućnosti dubljih analiza, cilj je prezentirati način kako stroj može algoritamskim putem bez posredstva analitičara otkriti sezonske oscilacije na razini uzoraka i pri tome izvršiti dodatne analize.

U literaturi koja se između ostalog bavi i temporalnim rudarenjem podataka [Pyle, 2001], ističe se vizualizacija kao mogući oblik analize gdje čovjek ima glavnu ulogu. Ako želimo otkriti frekventne sezonske uzorke, na temelju njih primjerice formirati događaj [Manilla, 1997], i sa tako formiranim događajima krenuti u narednu etapu analize bilo sa atributima koji su vremenski zavisni odnosno nezavisni jedan od rješenja je ovakav pristup analizi.

4.3. Otkrivanje cikličkih oscilacija

Cikličke oscilacije u svjetlu vremenskih serija možemo definirati kao nejednoliko ponavljanje uzoraka unutar vremenskih serija. Za razliku od problematike sezonskih oscilacija, kod cikličkih oscilacija ne možemo na temelju vremenske distance (vremenska kompleksnost) temeljiti svoje procjene o ponavljanju uzoraka u vremenskoj seriji. Cikličke oscilacije karakteriziraju uzorci ponavljanja, a naglasak je upravo na otkrivanju uzoraka koji se ponavljaju u nejednakim vremenskim odmacima. Slika 4.3 prikazuje cikličke oscilacije u vremenskoj seriji, gdje oznaka vk, označava mjerilo vremenske kompleksnosti (ciklički indeks) između dva uzorka.



SLIKA 4.3. CIKLIČKE OSCILACIJE

Kod cikličkih oscilacija između uzoraka ne postoji jednaki vremenski razmak. On ponekad može biti uzrokovan određenim pravilnostima, često puta u korelaciji sa nekim tržišnim pojavama, ali se ne pojavljuje u predvidivim vremenskim razmacima. Cikličke oscilacije nam mogu biti interesantne prvenstveno iz nekoliko razloga:

- Nakon otkrivanja uzoraka pravilnosti unutar vremenske možemo istražiti uzroke tih pravilnosti jer nisu uzrokovane sezonskim oscilacijama
- Otkriti što u stvari za promatranu vremensku seriju znači ta pravilnost, da li je to model primjerice ponašanja klijenta u vremenskom prostoru, ili je to slijed uzročno posljedičnih veza uzrokovanih nizom faktora

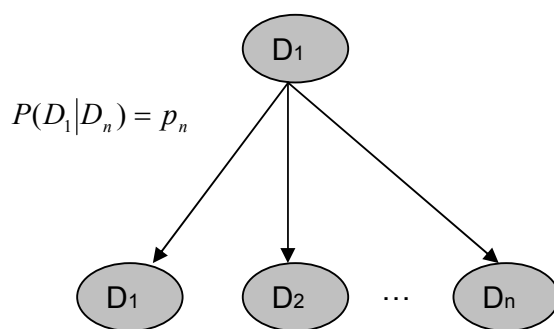
Ciklički uzorci u vremenskim serijama se također mogu promatrati sa perspektive vremenske kompleksnosti, gdje procjenjujemo da li postoji određena sličnost sa sezonskim oscilacijama. (Primjerice primanje osobnog dohotka je po definiciji sezonska pojava, ali kašnjenja plaće ovu pojavu mogu pretvoriti u cikličku pojavu uz uvjet da uzrokom – primanje osobnog dohotka raspolaganje novcem ima svojstvo jednoobraznog uzorka ponašanja)

Karakter i usmjerenost analize s obzirom na mehanizme REFII modela ovise o analitičaru. Da li će se analitičar koncentrirati na oblik funkcije te na osnovu toga donositi sud o cikličkoj oscilaciji, ili će krenuti i procjenu kvantitativnih elemenata analize, ovisi o konkretnom problemu.

Preduvjet provođenja analize cikličkih oscilacija jest otkrivanje uzoraka. Nakon otkrivanja uzoraka, moguće je promatrati dinamiku pojavnosti otkrivenih uzoraka što čini osnovnu ideju analize cikličkih oscilacija.

Pronađene frekventne uzorke možemo proglasiti događajima [Guralnik, 1999], te je u analizu moguće uključiti i epizode otkrivene u nekoj drugoj vremenskoj seriji posredstvom uvjetnih vjerojatnosti [Chiu, 2003]. U ovakav model rješenja posredstvom uvjetne vjerojatnosti moguće je uključiti i nevremenske attribute.

Ako analiziramo cikličke oscilacije i promatramo ih kao događaje, uz uvjet postojanja niza otkrivenih frekventnih uzoraka unutar vremenske serije moguće je kreirati model procjene odvijanja događaja D_n uz uvjet nastupa događaja D_1 kao što je vidljivo na slici 4.4.



SLIKA 4.4. MODEL UVJETNIH VJEROJATNOSTI PROCJENE DOGAĐAJA

Na prikazan način moguće je formirati Bayesovu mrežu događaja i procjenjivati pojavnosti događaja uz mogućnost integracije vremenske kompleksnosti kao elementa analize.

Probabilistički modeli u dosadašnjim znanstvenim radovima zanemaruju važnost jedinstvenog modela transformacije što ograničava širinu analize [Keogh, 1997], [Chiu, 2003].

Preporučena metodologija prilikom analize cikličkih oscilacija je korištenje intervalnih vrijednosti kutnih otklona i površine ispod krivulje.

U analizi cikličkih oscilacija možemo analizirati kako pojedinačne serije, tako i niz vremenskih serija, te tražiti kako cikličku oscilaciju, tako i korelatorne odnose među cikličkim uzorcima na različitim vremenskim serijama.

Generalno gledajući, REF model unutar REFII modela puno je grublja mjera za određivanje cikličkih oscilacija, ali može biti vrlo dobar brzo generirajući putokaz, koji će usmjeriti mehanizam za traženje uzoraka na temelju oblika krivulje u pravom smjeru. Nakon grubih procjena, algoritam za traženje uzoraka baziran na proračunu

kutnog odklona može dati vrlo preciznu sliku o stupnju "sličnosti" ili "jednakosti" uzoraka.

Ako želimo pronaći uzorke sa visokim stupnjem sličnosti unutar vremenskih serija koji se ciklički ponavljaju, tada uz smanjivanje intervalnih raspona odklona kutova kao elementa za procjenu jednakosti uzorka uvodimo i površinu ispod krivulje kao element analize, koji također može biti podijeljen u intervalne kategorije.

Ciklički indeks možemo definirati kao mjeru proizašlu iz mjerila vremenske kompleksnosti, a definira vremenske razmake između sličnih uzoraka u vremenskoj seriji.

Nad takvim indeksom možemo primijeniti metode deskriptivne statistike, kao što je mjera srednje vrijednosti, mjera disperzije, mod, medijan, te na taj način u slučaju ponavljanja istovrsnih uzoraka u cikličkim oscilacijama možemo raditi usporedne testove temeljene na cikličkim indeksima i elementima deskriptivne statistike. Ovdje još jednom dolazi do izražaja fleksibilnost i modularnost REFII modela. Osim osnovnih postupaka i elemenata koji služe u ovom konkretnom slučaju za otkrivanje uzoraka u vremenskoj seriji i koji spoznaju cikličke oscilacije, sustav je otvoren i za sofisticiranije analize koje se temelje na različitim analitičkim instrumentarijima.

4.4. Izravno otkrivanje pravila iz vremenske serije

4.4.1. Izravno otkrivanje pravila iz vremenskih serija posredstvom REFII modela

Jedna od karakteristika REFII modela je mogućnost direktnog otkrivanja pravila iz vremenskih serija, pri čemu elementi koji ulaze u analizu mogu, ali i ne moraju nužno imati temporalni karakter [Ohsaki, 2003].

Analitičara može interesirati odnos međuzavisnosti među pojavama koje su temporalnog i netemporalnog karaktera. Isto tako, za analizu nam može biti interesantno da li postoji pravilnost između pojave x koja se desila u vremenu t i pojave y koja se je desila u vremenu $t+1$, odnosno ($t+$ procijenjeni prag tolerancije).

Predmet istraživanja ne moraju nužno biti vremenski odsječci, već to mogu biti i vremenski uzorci, epizode, te nevremenske komponente.

Kada se govori o nevremenskim komponentama kao elementima koji ulaze u proces analize, tada u tu skupinu ulaze atributi poput regije, dobi, spola, pripadnosti određenoj grupi i slično.

Na taj način rezultati analize mogu dati informaciju da ako postoji rastući trend u nekom vremenskom periodu, da je najveći utjecaj na taj trend imala grupa atributa čija je vrijednost atributa dobi u nekom određenom razdoblju.

REFII model ne nudi svoje vlastito rješenje za direktno otkrivanje pravila iz vremenskih serija, već je jedno od mogućih rješenja u ostvarivanju tog cilja korištenje asocijativnih algoritama nad vremenskom serijom/serijama koje su transformirane u REFII notaciju.

Primjenom asocijativnih algoritama nad jednom transformiranom vremenskom serijom, koja u sebi sadrži vremenske indekse možemo otkrivati sezonske oscilacije. Ovo je alternativni način traženja sezonskih oscilacija unutar vremenskih serija transformiranih u REFII notaciju.

Ako u vremensku seriju uvedemo atribut koji nije zavisen u vremenu, poput regije, pripadnosti nekoj interesnoj skupini i slično, tada ti atributi postaju ravnopravni vremenskim atributima i sudjeluju u procesu analize.

Istovremeno možemo promatrati nekoliko vremenskih serija i među njima tražiti pravilnosti, što će biti predmet empirijskog istraživanja u narednom poglavlju.

4.4.2. Model rješenja izravnog otkrivanja pravila iz vremenskih serija

Kao predmet istraživanja uzeti su podaci sa :

http://www.stat.duke.edu/~mw/ts_data_sets.html

koji prikazuju indikatore proizvodnje na mjesečnoj razini u periodu od 1947-1993.

Izvor: Federal Reserve Statistical Release G.17

Struktura podataka : vremenske serije za period od 1947-1993

Analiza je provedena na Pentiumu 1.6 GHz, 512 MB RAM-a, proces obrade trajao je 4 sekunde za transformaciju u REFII model, dok je proces direktnog otkrivanja pravila trajao 3 sekunde.

YR = Godina

MN = Mjesec

IP = Indeks industrijske proizvodnje

MFG = Prerađivačka industrija

MFGD = Trajna dobra

MFGN = Netrajna dobra

MIN = Rudarstvo

UTIL = Javne usluge

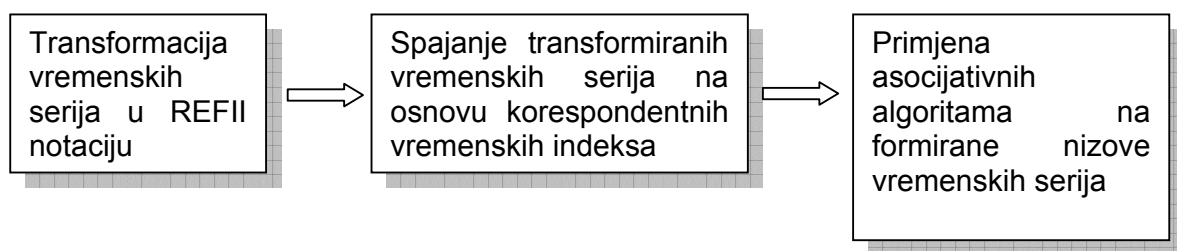
P = Proizvodi ukupno

MAT = Materijali

Svaka vremenska serija ima 564 elementa.

Cilj analize je otkriti pravilnosti u kakvoj su međuzavisnosti indeks industrijske proizvodnje, javne usluge, rudarstvo i proizvodnja materijala.

Idejni model rješenja prikazan je na slici 4.5.



SLIKA 4.5. IDEJNI MODEL RJEŠENJA IZRAVNOG OTKRIVNJA PRAVILA

Prilikom transformacije originalnih vrijednosti vremenske serije korištena je naredna tablica klasifikacije otklona kutova vremenskih odsječaka:

Tablica 4.9. Tablica klasifikacije otklona kutova

Razred	Donja granica	Gornja granica	REF
Nizak rast	0.000000000	0.300000000	R
Srednji rast	0.300000000	0.700000000	R
Visok rast	0.700000000	1.000000000	R
Nizak pad	0.000000000	0.300000000	F
Srednji pad	0.300000000	0.700000000	F
Oštar pad	0.700000000	1.000000000	F
Bez promjene	0.000000000	0.000000000	E

Nakon provedene transformacije četiri vremenske serije svaka duljine 562 elementa, vremenske serije su spojene primjenom SQL upita na temelju vremenskih indeksa, te je rezultat obrade eksportiran u tablicu formata :

Tablica 4.10. Tablica transformacije u REFII model za potrebe izravnog otkrivanja pravila

ciklus	indeks	ip	ut	min	mat
discrete	discrete	discrete	discrete	discrete	discrete
47	2	Nizak rast	Nizak rast	Nizak rast	Nizak rast
47	3	Nizak rast	Nizak rast	Nizak rast	Nizak rast
47	4	Nizak pad	Nizak rast	Nizak pad	Nizak pad
47	5	Nizak rast	Nizak rast	Nizak rast	Nizak rast
47	6	Bez promjene	Bez promjene	Nizak pad	Nizak pad
47	7	Nizak pad	Nizak rast	Nizak rast	Nizak pad
47	8	Nizak rast	Nizak rast	Nizak rast	Nizak pad
47	9	Nizak rast	Nizak rast	Nizak rast	Nizak rast
47	10	Nizak rast	Nizak rast	Nizak pad	Nizak rast
47	11	Nizak rast	Nizak rast	Nizak rast	Nizak rast
47	12	Bez promjene	Nizak rast	Nizak pad	Nizak pad
...

Ovaj format tablice je standardni format podataka u *data mining* modulu, *open source-Orange*, <http://magix.fri.uni-lj.si/orange/default.asp> za programski jezik Python www.python.org razvijenom na Ljubljanskem univerzitetu, katedra za umjetnu inteligenciju.

Na temelju tog modula kreiran je program za otkrivanje asocijativnih pravila u iz REFII modela. Nad transformiranim serijama u REFII modelu provela se je analiza pomoću asocijativnih algoritama.

U nastavku je prikazan izvorni kod pisan u programskom jeziku Python sa korištenjem Orange modula za otkrivanje asocijativnih pravila na temelju transformiranih vremenskih serija:

```
import orange, orngAssoc
data = orange.ExampleTable("serije.tab")

minSupport = 0.3
rules = orngAssoc.build(data, minSupport)
```

```
print "%i pravila sa support koeficijentom većim ili jednakim %5.3f .\n" % (len(rules),
minSupport)
```

```
subset = rules[0:100]
subset.printMeasures(['support','confidence'])
```

```
print
del subset[0:2]
subset.printMeasures(['support','confidence'])
```

4.4.3. Izravno otkrivanje pravila na empirijskim podacima

Uz zahtjev da support koeficijent bude veći ili jednak od 0.3 dobiveno je 17 pravila :

```
supp. conf. Pravilo
0.521 0.891 mat=Nizak rast -> ip=Nizak rast
0.521 0.880 ip=Nizak rast -> mat=Nizak rast
0.425 0.718 ip=Nizak rast -> ut=Nizak rast
0.425 0.664 ut=Nizak rast -> ip=Nizak rast
0.415 0.708 mat=Nizak rast -> ut=Nizak rast
0.415 0.647 ut=Nizak rast -> mat=Nizak rast
0.377 0.910 ut=Nizak rast mat=Nizak rast -> ip=Nizak rast
0.377 0.887 ut=Nizak rast ip=Nizak rast -> mat=Nizak rast
0.377 0.724 mat=Nizak rast ip=Nizak rast -> ut=Nizak rast
0.377 0.644 mat=Nizak rast -> ut=Nizak rast ip=Nizak rast
0.377 0.637 ip=Nizak rast -> ut=Nizak rast mat=Nizak rast
0.377 0.589 ut=Nizak rast -> mat=Nizak rast ip=Nizak rast
0.319 0.746 min=Nizak rast -> mat=Nizak rast
0.319 0.544 mat=Nizak rast -> min=Nizak rast
0.313 0.733 min=Nizak rast -> ip=Nizak rast
0.313 0.529 ip=Nizak rast -> min=Nizak rast
0.302 0.708 min=Nizak rast -> ut=Nizak rast
0.425 0.718 ip=Nizak rast -> ut=Nizak rast
0.425 0.664 ut=Nizak rast -> ip=Nizak rast
0.415 0.708 mat=Nizak rast -> ut=Nizak rast
0.415 0.647 ut=Nizak rast -> mat=Nizak rast
0.377 0.910 ut=Nizak rast mat=Nizak rast -> ip=Nizak rast
0.377 0.887 ut=Nizak rast ip=Nizak rast -> mat=Nizak rast
0.377 0.724 mat=Nizak rast ip=Nizak rast -> ut=Nizak rast
0.377 0.644 mat=Nizak rast -> ut=Nizak rast ip=Nizak rast
0.377 0.637 ip=Nizak rast -> ut=Nizak rast mat=Nizak rast
0.377 0.589 ut=Nizak rast -> mat=Nizak rast ip=Nizak rast
0.319 0.746 min=Nizak rast -> mat=Nizak rast
0.319 0.544 mat=Nizak rast -> min=Nizak rast
0.313 0.733 min=Nizak rast -> ip=Nizak rast
0.313 0.529 ip=Nizak rast -> min=Nizak rast
0.302 0.708 min=Nizak rast -> ut=Nizak rast
```

Primjerice pravilo ip=Nizak rast -> mat=Nizak rast , koje govori da ako je ukupan indeks industrijske proizvodnje u niskom rastu, tada je i proizvodnja materijala u

niskom rastu ima vrijednost support koeficijenta 0.521 i visoku pouzdanost od 0.891.

Iz ovog je primjera vidljiv način na koji REFII model otvara mogućnost primjene *data mining* algoritma nad elementima vremenske serije.

Na spomenuti set podataka moguće je djelovati i ostalim algoritmima poput stabla odlučivanja, ili pak procjenjivati relevantnosti atributa temeljem Gini indeksa s obzirom na ciljnu varijablu primjerice varijablu "godina". Asocijativni algoritmi su samo jedan od primjera kako iz vremenskih nakon transformacije u REFII notacije možemo otkrivati pravila primjenom *data mining* algoritama.

U fokusu analize ne moraju biti samo odsječci sa istim vremenskim indeksom. Ako želimo pratiti analizu utjecaja, tada možemo upariti pojavom na vremenskoj razini t , sa pojavom na vremenskoj razini $t+n$ za koju sumnjamo da bi mogle imati međuzavisni utjecaj sa vremenskim pomakom, te tada na njih djelujemo sa asocijativnim algoritmima.

Prednost ovakvog pristupa očituje se i u mogućnosti detaljnijih analiza nad izdvojenim skupom podataka, gdje obuhvaćamo vrijednosti atributa otklona sa zadovoljavajućim vrijednostima koeficijenata, te im pridružujemo i dodatne attribute s ciljem dodatne analize.

Primjer za to može biti izdvajanje dijela vremenske serije koji zadovoljava kriterije da je $lp="Nizak\ rast"$ i $mat="Nizak\ rast"$, pri čemu se izdvajaju i ostale vrijednosti za ut , min , $ciklus$ i $indeks$ i pokušavaju naći zakonitosti nad tim podskupom podataka. Kada je vremenska serija pretprocesirana na spomenuti način, tada je izdvajanje spomenutog podskupa moguće realizirati i primjenom jednostavnih SQL upita na vremensku seriju.

Daljnja metodologija koja se bazira na direktnom otkrivanju pravila može biti usmjerena ka primjeni Bayesovih mreža na nizove vremenskih serija koje su transformirane u REFII notaciju. Profesionalni alati poput Hugin (www.hugin.com) sposobni su temeljem metodologije "Batch" učenja [Jensen, 2001] riješiti probleme kombinatorne eksplozije procesa učenja Bayesove mreže. Ova spoznaja može biti vrlo korisna, jer na taj način matricu transformacije možemo iskoristiti kao podlogu za učenje Bayesove mreže.

Ovim pristupom moguće je proračunavati uvjetne vjerojatnosti kutnih otklona. Konceptcija Bayesovih mreža, kao i konceptcija primjene asocijativnih algoritama također omogućava spajanje atributa temporalnih svojstava sa netemporalnim atributima.

4.5. Otkrivanje epizoda u vremenskim serijama

Vrlo dobru konceptualnu razradu epizoda u vremenskim serijama dao je Heikki Manilla [Manilla, 1997], fokusirajući se pri tome na problematiku alarmnih sustava, i predviđanja određenih situacija na osnovu frekventnih uzoraka u vremenskoj seriji. U takvim sustavima pojam događaja u vremenskoj seriji odnosi se na primjerice zvonjavu alarma, aktiviranje protuprovalnog sustava, radnje provalnika i sl.

Cilj njegovih istraživanja odnosio se na promatranje frekventnih podskupova vremenskih serija koje se sastoje od niza događaja s ciljem predikcije događaja u određenim situacijama. Ovi elementi su temelj koncepcije koju naziva epizode u vremenskim serijama. U ovom poglavlju cilj mi je iskoristiti generalne ideje prezentirane u navedenom radu, te razraditi i proširiti te ideje u duhu REFII modela s ciljem proširivanja i nadopunjavanja spomenutih kategorija.

Epizode unutar vremenskih serija imaju vrlo veliko područje konceptualne primjene [Patel, 2002], [Guralnik, 1999]. Njihova osnovna karakteristika procjene faktora sigurnosti je računanje frekvencija pojavnosti niza događaja, koje onda proglašavamo epizodom, ako zadovoljava kriterij minimalne pojavnosti [Manilla, 1997].

Za razliku od ovih istraživanja koja se odnose na konkretnu problematiku, te na osnovu toga sužavaju problemski prostor, REFII model je opće orijentiran model, ali se generalna ideja ovog koncepta može primijeniti u REFII modelu.

Za razliku od opisanog sustava REFII model ne poznaje pojam "događaja". Događaje kao elemente vremenskih serija možemo konstruirati na osnovu komponenti REFII modela. Tako primjerice "događaj" može biti determiniran vrijednošću ili intervalnim rasponom vremenskog odsječka između kraka / krakova vremenskog odsječka i/ili površine/površina vremenskog odsječka, i / ili kao oznaka REF modela. Tako deklarirani događaj može imati pridruženu leksičku oznaku, kao što je primjerice "oštar rast", "nizak prihod", "oštar pad".

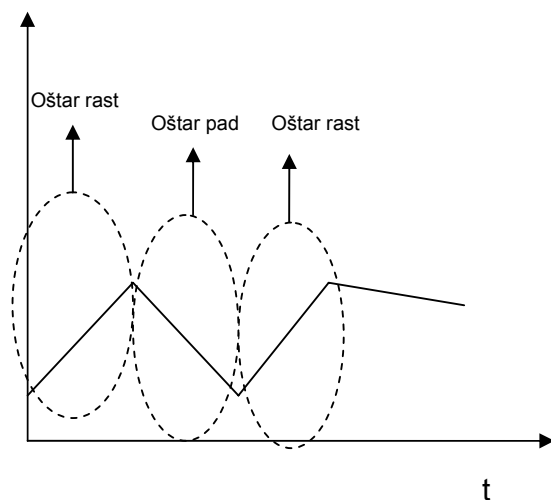
Kao primjer navodim hipotetsku transformaciju vremenskog odsječka u leksičku oznaku koja simbolizira događaj:

Tablica 4.11. Definicija leksičke oznake

Leksička oznaka: "Oštar rast"
Raspon koeficijenta kutnog odklona : $>0.7, <1.0$

Navedeni primjer oslanja se na transformaciju u leksičku oznaku temeljem nagiba između krakova vremenskog odsječka. Kao element transformacije možemo još uključiti i površinu ispod krivulje, i REF model. U konkretnom hipotetskom slučaju ovu leksičku oznaku moguće je formirati na osnovu vrijednosti koeficijenata kutnog odklona. Ovisno o situaciji za kreiranje događaja koji je deklariran pripadajućom leksičkom oznakom biramo elemente REFII modela koji sudjeluju u transformaciji.

U tako definiranom sustavu nikako ne smijemo zanemariti i mjerilo vremenske kompleksnosti, koje isto tako može imati utjecaja kod transformacije elemenata REFII modela u događaje. Slika 4.6 prikazuje transformaciju elemenata REFII modela u događaje.



SLIKA 4.6. TRANSFORMACIJA ELEMENATA REFII MODELA U DOGAĐAJE

Iz slike 4.6. vidljiva je osnovna metodologija transformacije, uz mjerilo vremenske kompleksnosti=1.

S obzirom na prirodu modela koji se obrađuje u radu, bilo bi besmisleno događaje deklarirati na temelju pojedinačnih vremenskih odsječaka. Mnogo je interesantnije promatrati frekventne uzorke te ih deklarirati kao događaje, promatrati njihove odnose kako u okviru jedne, tako i u međuzavisnosti kroz niz vremenskih serija.

Ako primjerice otkriveni frekventni uzorak karakterizira period kontinuiranog oštrog rasta u pet vremenskih jedinica, tada takav period možemo deklarirati kao epizodu kontinuiranog rasta.

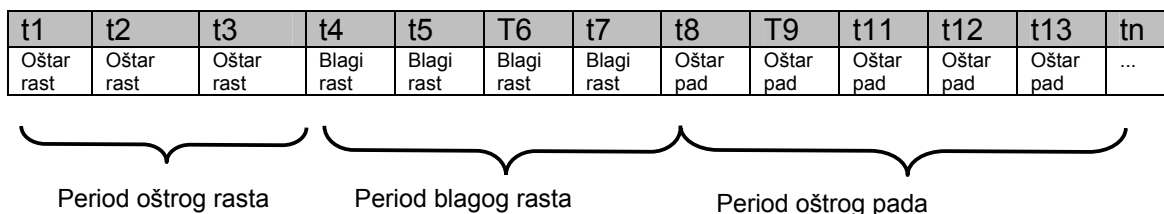
Postupak deklarativnog sažimanja uzorka u epizode prikazan je narednim pseudoalgoritmom:

Transformiraj sve elemente vremenske serije na temelju REFII modela u intervalne vrijednosti.

Provjeri postojanje pravila (npr. ako postoji niz intervalnih vrijednosti sa mjerilom kompleksnosti > 1 "oštar rast" u intervalu $t-t_n$ formiraj "epizodu" period oštrog rasta ili na temelju prethodno otkrivenih uzoraka)

Formiraj listu epizoda.

Ovaj postupak možemo prikazati shematski slikom 4.7. :



SLIKA 4.7. FORMIRANJE EPIZODA U VREMENSKOJ SERIJI

Iz sheme je vidljivo da se nakon transformacije elemenata u elementarne događaje mogu formirati pravila koja determiniraju nadkategorije, koje nužno ne moraju imati istu vremensku granularnost. Nadkategorije se mogu formirati i na još većoj razini granulacije, no to naravno ovisi o analitičaru, stupnjevima slobode, te o karakteru analize.

Sve prethodno spomenute konceptualne primjene REFII modela mogu se iskoristiti nakon ovakvog tipa transformacije.

Zamislimo situaciju gdje u vremenskoj seriji od n elemenata transformiranih u REFII notaciju tražimo zakonitosti. Nakon prve transformacije i formiranja intervalnih vrijednosti, još uvijek imamo isti broj elemenata, u narednoj etapi gdje sažimamo elemente prema unaprijed definiranim pravilima smanjujemo osnovnu populaciju, čime optimiziramo vrijeme pretrage, ali dobivamo elemente na višem stupnju apstrakcije. Sa perspektive donositelja odluka nakon provedene krajnje analize, činjenica da nakon *perioda oštrog rasta od prosječno 3 vremenske jedinice dolazi period blagog rasta od prosječno 4 vremenske jedinice, a nakon toga nastupa period oštrog pada od prosječno 5 vremenskih jedinica* sa vjerojatnošću od 0.7 može biti vrlo korisna.

Naravno ove pokazatelje možemo iskazati kvantitativnim pokazateljima, pa tada imamo potpuniju sliku.

Heikki Manilla [Manilla, 1997] u svom radu preporuča definiciju “vremenskih prozora”, koji dijele vremensku seriju na n dijelova, pri čemu vremenski prozori imaju pomak $n+1$ u vremenskoj seriji. Veličina vremenskih prozora može se definirati vrijednošću mjerila vremenske kompleksnosti, a konceptualno se svodi na maksimalno vrijeme u kome korisnik očekuje odvijanje određenog niza događaja. Sa perspektive istraživanja u radu, koncept događajnosti će biti prilagođen potrebama istraživanja na REFII modelu i to prvenstveno sa perspektive predikcije događaja u vremenskoj epizodi.

To znači da korisnik može primjerice postaviti upit : Da li, i za koliko nakon perioda/događaja oštrog pada, i blagog rasta ... nastupa oštar pad, unutar n vremenskih jedinica.

Sustav temeljen na REFII modelu i modelu epizoda u vremenskim serijama trebao bi dati odgovor na to pitanje sa stupnjem pouzdanosti. Isto tako korisnik može postaviti pitanje tipa: Što slijedi nakon perioda oštrog pada. Sustav na osnovu upita prolazi kroz sve definirane uzorke “oštrih padova”, i traži zakonitost jednoobraznog ponavljanja karakterističnog događaja koji slijedi nakon te pojave.

Heikki Manilla [Manilla, 1997] spominje tri osnovne vrste epizoda:

Linearne

Paralelne

Spojne

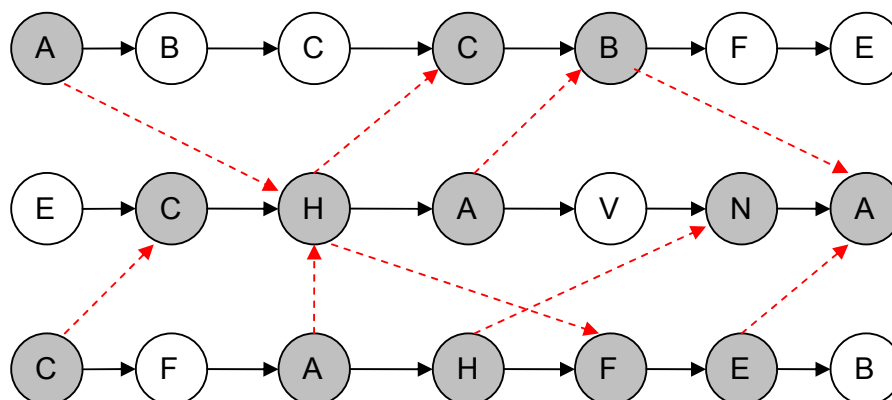
Linearne epizode uglavnom su obrađene u tekstu, a odnose se na situacije kada obrađujemo jednu vremensku seriju i u njoj pronalazimo uzorke ponavljajućih događaja.

Paralelne epizode moguće je analizirati u situacijama kada imamo dvije ili više vremenskih serija, gdje uz klasičnu analizu možemo provoditi i korelatorno –epizodne analize, poštujući pri tome parametre vremenske kompleksnosti.

Spojne analize provode se u situacijama kada imamo dvije ili više vremenskih serija, gdje u određenom vremenskom periodu svi promatrani uzorci poprimaju identične vrijednosti. Ovdje je također bitno uočiti važnost poštivanja zakonitosti koje proizlaze iz primjena mjerila vremenske kompleksnosti.

Vrlo složena analiza, zbog ekspanzije kombinatorne eksplozije odnosi se na korelatornu analizu događaja unutar niza paralelnih vremenskih serija, zanemarujući mjerilo vremenske kompleksnosti.

Ova situacija prikazana je na slici 4.8..



SLIKA 4.8. KORELATORNE ANALIZE DOGAĐAJA SA ZANEMARIVANJEM MJERILA VREMENSKE
KOMPLEKSNOŠTI

Zadatak ove analize jest traženje korelatornih odnosa između događaja u paralelnim serijama, bez obzira na mjerilo vremenske kompleksnosti. U ovom slučaju zbog optimizacije pretrage preporuča se korištenje vremenskih prozora, koje sustav može inicijalno intervalno omeđiti, dajući korisniku mogućnost samostalnog izbora intervala uz asistenciju sustava po pitanju sugestija oko trajanja obrade, ili ograničenosti resursa sustava zbog količine podataka za obradu. Ako pažljivije promotrimo sliku, tada su uočljivi elementi "link analize", koji u ovom slučaju čine nove "hibridne epizode". Važno je napomenuti da je u slučaju ovakve vrste analiza potrebno kao element indeksa navoditi vremensku seriju i vremenski indeks.

Za detaljniju analizu ovako povezanih događaja, možemo iskoristiti *Anacapa* metodu za analizu veza.

Nikako se ne smije smetnuti s uma da upravo korisnik definira događaj koristeći elemente REFII modela. Transformacija parametara u događaje nužan je postupak, kako bi se iskoristila idejna koncepcija epizoda u vremenskim serijama. Osim što analiza vremenskih serija primjenom epizoda ima karakter prediktivne analize, isto tako može poslužiti za otkrivanje znanja uz određeni stupanj pouzdanosti.

Ovu koncepciju u svjetlu REFII modela također možemo primijeniti i na koncepciju otkrivanja cikličkih oscilacija, sezonskih oscilacija, pri čemu onda analiza djelomično poprima karakter razrađenijeg REF modela. Nedostatak takvog pristupa očituje se u potrebi definiranja niza leksičkih oznaka, što je vrlo nepraktičan posao kada analiziramo kompleksne vremenske serije od n elemenata.

Koncepcija epizoda može dati vrlo dobre rezultate u situacijama kada otkrivamo uzorke uzročno posljedičnih veza, testiramo hipoteze o sljedovima događaja, otkrivamo unakrsne korelatorne odnose u uzorku vremenskih serija.

Postoji čitav niz modela koji su usvojili koncepciju epizoda kao temelj analize vremenskih serija. U svjetlu REFII modela, epizode proširuju mogućnosti analize i daju mu jednu novu dimenziju koja proizlazi iz kategorizacije niza pojava objedinjenih kako u elementarni vremenski odsječak, tako i u nadkategorije elementarnih vremenskih odsječaka. Nad tako formiranim kategorijama moguće je raditi analize s ciljem otkrivanja uzoraka.

Jedan od problema koji se pojavljuje u deklariranju leksičkih oznaka na osnovu REFII modela je krutost prilikom njihove intervalne definicije. Primjerice leksičku oznaku "oštar rast", mogu intervalno karakterizirati koeficijent kutnog odklona >0.8 . U ovu skupinu ne spada koeficijent kutnog odklona od 0.79. Ova oštra granica, i potreba za definicijom kategorija, koje su često puta ključni korak u pripremi analiza značajno utječu na krajnji ishod analiza. Ovakva koncepcija oštrog razgraničavanja kategorija i formiranje leksičkih oznaka u velikom broju slučajeva je prihvatljiva za analize, no ponekad ako se želi biti "fleksibilniji" u definiciju događaja moguće je uključiti elemente *fuzzy logike*. Jedan od primjera primjene *fuzzy logike* u sličnoj situaciji vezanog uz analizu vremenskih serija posredstvom neuronskih mreža nadograđenu elementima *fuzzy logike* dao je Cheng [Cheng, 1997]. Ugradnjom ovih elementa u osnovni model dobivamo fleksibilniji sustav. Vrlo je važno na osnovu konkretnog problema kojeg rješavamo procijeniti prednosti uvođenja *fuzzy logike* preko parametara REFII modela u koncept otkrivanja epizoda u vremenskim serijama, jer ima situacija u kojima *fuzzy logika* ne nosi nužno poboljšanje u izvedbenoj koncepciji modela.

4.6. Teorija “spavača”

Teorijom “spavača” objašnjavamo istovrsne pojave u vremenskoj seriji koje se dešavaju u određenom vremenskom razmaku, a koje su nelogične ili neočekivane sa perspektive poslovne prakse.

Najočitiji primjer gdje ova teorija može dati rezultate su primjerice korištenje bankovnih stambenih kredita.

Osoba koja je korisnik stambenog kredita, nakon realiziranog ugovora postaje relativno “neinteresantna” banci sa perspektive korištenja ove vrste usluge iz razloga što se pretpostavlja da ta osoba više neće koristiti ponovno ovakvu vrstu usluge, odnosno da je neće koristiti u skorije vrijeme.

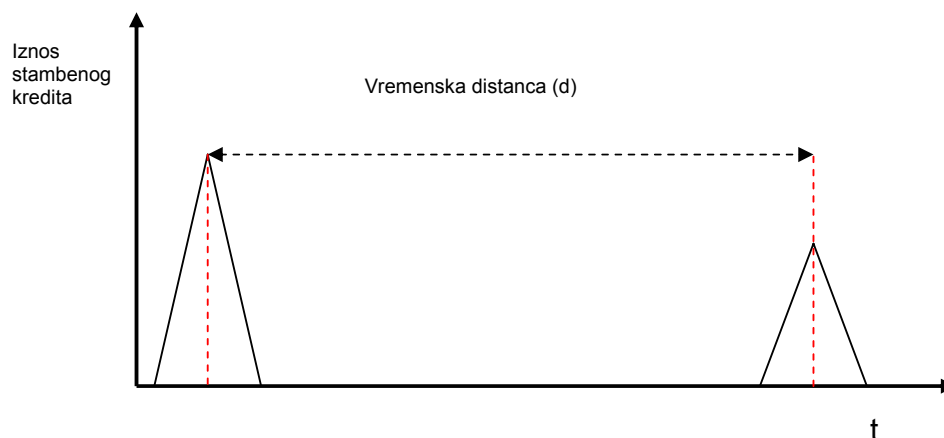
Praksa pokazuje da je to djelomično točno, određeni broj korisnika ovakvih vrsta kredita uistinu više neće podnositi zahtjev za stambenim kreditom, ali isto tako postoji određeni tržišni segment, koji odskaka od stereotipa. Razlog tome mogu primjerice biti klijenti koji imaju djecu, i određenog su imovinskog statusa, te nakon određenog broja godina ponovo traže stambeni kredit.

U ovom slučaju postavljaju se dva osnovna pitanja :

Koje su karakteristike ove ciljne skupine

Koje je prosječno vrijeme između dva zahtjeva

Na prvo pitanje odgovor može dati primjena klasičnih *data mining* metoda nad uzorcima podataka kao što su stabla odlučivanja. Odgovor na drugo pitanje moramo tražiti analizirajući vremenske serije, koja za slučaj ponovljenog traženja kredita izgleda kao na slici 4.9.



SLIKA 4.9. GRAFIČKA INTERPRETACIJA TEORIJE SPAVAČA PREKO VREMENSKE SERIJE

Iz slike je vidljivo da postoje dva ključna elementa za analizu, a to su : vremenska distanca, i iznos. Vremensku distancu mjerimo mjerilom vremenske kompleksnosti, a iznos je “mjerljiv” kutnim otklonom krakova, ili originalnim vrijednostima iz baze.

Gledajući ovu problematiku kroz prizmu događaja i epizoda, u vremenskoj seriji imamo dva događaja u periodu d .

Algoritam za pronalaženje zakonitosti nakon otkrivene karakteristike uzorka, fokusira se na očekivano vrijeme ponovnog zahtjeva kredita i očekivani iznos kredita.

Za otkrivanje ovih zakonitosti ponekad su nam dovoljni i elementi deskriptivne statistike u koje uvrštavamo elemente REFII modela. Osnovni preduvjet uspješne analize je izdvajanje podskupa klijenta iz ukupne populacije, koji zadovoljavaju kriterije koji čine "teoriju spavača".

Ono što se nikako ne smije zaboraviti, da se navedena teorija prvenstveno temelji na izvornim *data mining* metodama, a REFII je ovdje spomenut kao nadogradnja za preciznije određivanje karakteristika prvenstveno po pitanju mjerila vremenske kompleksnosti i iznosa koji se oslanja na koeficijent kutnog odklona.

REFII model može biti od velike koristi prilikom formiranja klastera na osnovu mjerila vremenske kompleksnosti i iznosa zahtijevanih kredita.

Teorija "spavača", je u biti vrlo jednostavna koncepcija za čiju implementaciju nije nužno potreban REFII model. Razlog njenog uvrštavanja kao konceptualne primjene REFII modela proizlazi iz dva osnovna razloga. Prvi se odnosi na velike skrivene potencijale primjene ove teorije u financijskim institucijama, gdje se ovaj model može primijeniti ne samo na problematiku kredita, nego i na kartično poslovanje, stambenu štednju i niz sličnih situacija. Iz tog razloga smatram da ga je bitno spomenuti kao koncepciju koja pridonosi cjelovitosti pogleda na problematiku vremenskih serija i njihovoj primjeni u bankarskim sustavima. Drugi bitan razlog uvrštavanja ove teorije kao konceptualne primjene REFII modela odnosi se na ilustraciju univerzalnosti REFII modela, koji se vrlo lako može prilagoditi različitim vrstama koncepcija koje proizlaze iz problematike vremenskih serija.

Ovaj teorijski koncept karakterizira relativno jednostavna algoritamska interpretacija. Problemi prilikom primjene ovog modela mogu se pojaviti zbog zahtijevanog opsega podataka nad kojima se vrši obrada. S jedne strane postoji problem nesveobuhvatnosti podataka, nužnih za provođenje ove vrste analize, a s druge strane javlja se problem potencijalne kombinatorne eksplozije, upravo iz razloga potrebne sveobuhvatnosti podataka. Razlog uključivanja povijesnih podataka unatrag više godina, proizlazi iz činjenice što klijenti koje se uklapaju u teorijski okvir "spavača" mogu imati relativno veliki iznos vremenske kompleksnosti. Analitičar određuje granična vremena, uzimajući u obzir pojam profitabilnosti, te se na osnovu raspoloživih podataka uz otkrivanje karakteristika segmenta tržišta bavi i proračunima očekivanih vremena reaktivacije, iznosa, kao i formiranje klastera na osnovu ovih vrijednosti.

Pomoću teorije spavača moguće je prepoznati potentne tržišne klastere do kojih se dolazi sinergijom primjene REFII modela i tradicionalnih *data mining* metoda.

4.7. Grupiranje vremenskih odsječaka

4.7.1. Pojam grupiranja vremenskih odsječaka i REFII model

Grupiranje vremenskih odsječaka usko je vezano uz pojam sličnosti. Osnovna ideja njihova grupiranja proizlazi iz potrebe za dijagnosticiranjem i deklariranjem vremenskih odsječaka koji imaju ista ili slična svojstva.

Autori [Lin, 2001],[Lin, 2002] su uglavnom skloni problematiku grupiranja rješavati primjenom algoritama za klasteriranje nad vremenskim odsječcima.

Klasteriranjem se određuju najbližije vremenske serije/odsječci, te se na osnovu toga vrši grupiranje prema kriteriju sličnosti.

Nedostatak ovakvog pristupa očituje se u inicijalnom postavljanju broja klastera.

U svjetlu REFII modela grupiranje je posljedica otkrivene sličnosti, pri čemu se sličnost definira preko intervalne jednakosti.

Prilikom korištenja Hartiganovog algoritma za procjenu sličnosti vremenske serije koriste se diskretne numeričke vrijednosti podesne za obradu sa funkcijama za procjenu distanci kao što je Euklidska ili Manhattan distanca [Jain, 1988]. Ovakve vrste algoritama se oslanjaju na klasično klasteriranje i uvode ga u vremenske serije, bez osvrta na model transformacije, te su nepodesna za rad sa intervalnim jednakostima i lingvističkim varijablama za procjenu sličnosti.

Fisherov algoritam za pronalaženje sličnosti svodi se na traženje optimalnih presjeka za određivanje segmenata vremenskih serija, pri čemu je prisutan problem kombinatorne eksplozije.

Ova dva pristupa iako efikasna i usko usmjerena na problemski prostor sličnosti nepodesna su za rješavanje problema sličnosti putem REFII modela, jer prvenstveno ne postoji kompatibilnost na razini modela transformacije. Sličnost algoritamski definirana na način kako je to prikazano na narednom ilustrativnom primjeru, povezana sa samoorganizirajućim mapama kao što je to učinjeno na empirijskim istraživanjima za otkrivanje uzoraka rješava probleme preklapanja uzoraka u vremenskim serijama i kombinatorne eksplozije te se bazira na procjenama sličnosti segmenata vremenske serije jednake dužine.

U takvom sustavu, na dvije ili više vremenskih serija transformiranih putem REFII modela moguće je provoditi ispitivanje sličnosti na način usporedbe uzoraka uz pomoć intervalnih jednakosti.

Samoorganizirajuće mape pomažu u redukciji kombinatorne eksplozije na način da uz pomoć njih možemo izgraditi stablo sličnih uzoraka, a u narednu etapu obrade ulaze svi uzorci dužine $n+1$ koji imaju slične uzorke dužina n .

Isti mehanizam za redukciju kombinatorne eksplozije korišten je kod algoritma za otkrivanja uzoraka, s tom razlikom što se preko intervalnih jednakosti ne uspoređuju uzorci iz dvije vremenske serije nego uzorci nastali na temelju jedne vremenske serije. Naredni empirijski primjer ilustrira metodologiju procjene sličnosti vremenskih serija preko intervalnih jednakosti, a konstruiran je s namjerom ubrzavanja i optimiziranja procesa pretrage u situacijama kada otkrivamo i grupiramo slične odsječke na n dugih vremenskih serija.

Za dva odsječka kod kojih postoji intervalna jednakost kažemo da su slična, odnosno možemo im dodijeliti pripadnost istoj grupi.

Za niz odsječaka kod kojih postoji intervalna jednakost među svim odsječcima kažemo da su slični, te im također možemo pridodijeliti pripadnost istoj grupi [Lee, 2000], [Park, 1999], [Park, 2001].

Ovo je temeljni postulat na osnovu kojeg je izgrađen algoritam za pronalaženje sličnosti unutar vremenskih serija.

Naredni prikazan postupak otkrivanja sličnosti unutar vremenskih serija oslanja se na proračun udaljenosti među vremenskim odsječcima na temelju funkcija distanci. Ovakav pristup je inspiriran nizom znanstvenih radova na temu sličnosti koji koriste funkcije udaljenosti s ciljem procjene sličnosti [Lin, 2001], [Lin, 2002].

4.7.2. Model grupiranja vremenskih odsječaka

Osnovna ideja prvog predloženog postupka procjene sličnosti proizlazi iz činjenice da slični odsjecci ulaze u istu grupu pripadnosti [Park, 2001], [Keogh, 1998].

Jedan od zadataka grupiranja može se odnositi na traženje sličnosti između niza vremenskih serija na intervalu od $1..n$ pri čemu interval od $1..n$ predstavlja cjelokupni interval vremenske serije [Rafiei, 1998], [Popivanov, 2002], [Manilla, 2001].

Pojednostavljeno rečeno, m vremenskih serija na cijelom intervalu od 1..n spadaju u istu grupu uz uvjet da postoji intervalna jednakost među vremenskim odsječcima na cjelokupnom intervalu vremenske serije.

Alternativni pristup može se odnositi na traženje sličnosti između niza vremenskih serija i to na podintervalu intervala 1..n.

Kada imamo temeljnu definiciju sličnosti kao što je postavljena u definiciji sličnosti, tada na temelju te definicije možemo izvesti čitav niz različitih tipova sličnosti, s obzirom na cilj analize. Tako primjerice i otkrivanje motiva (eng. motifs) [Lin, 2001], ili epizoda [Manilla, 1997] može isto biti promatrano kroz prizmu sličnosti.

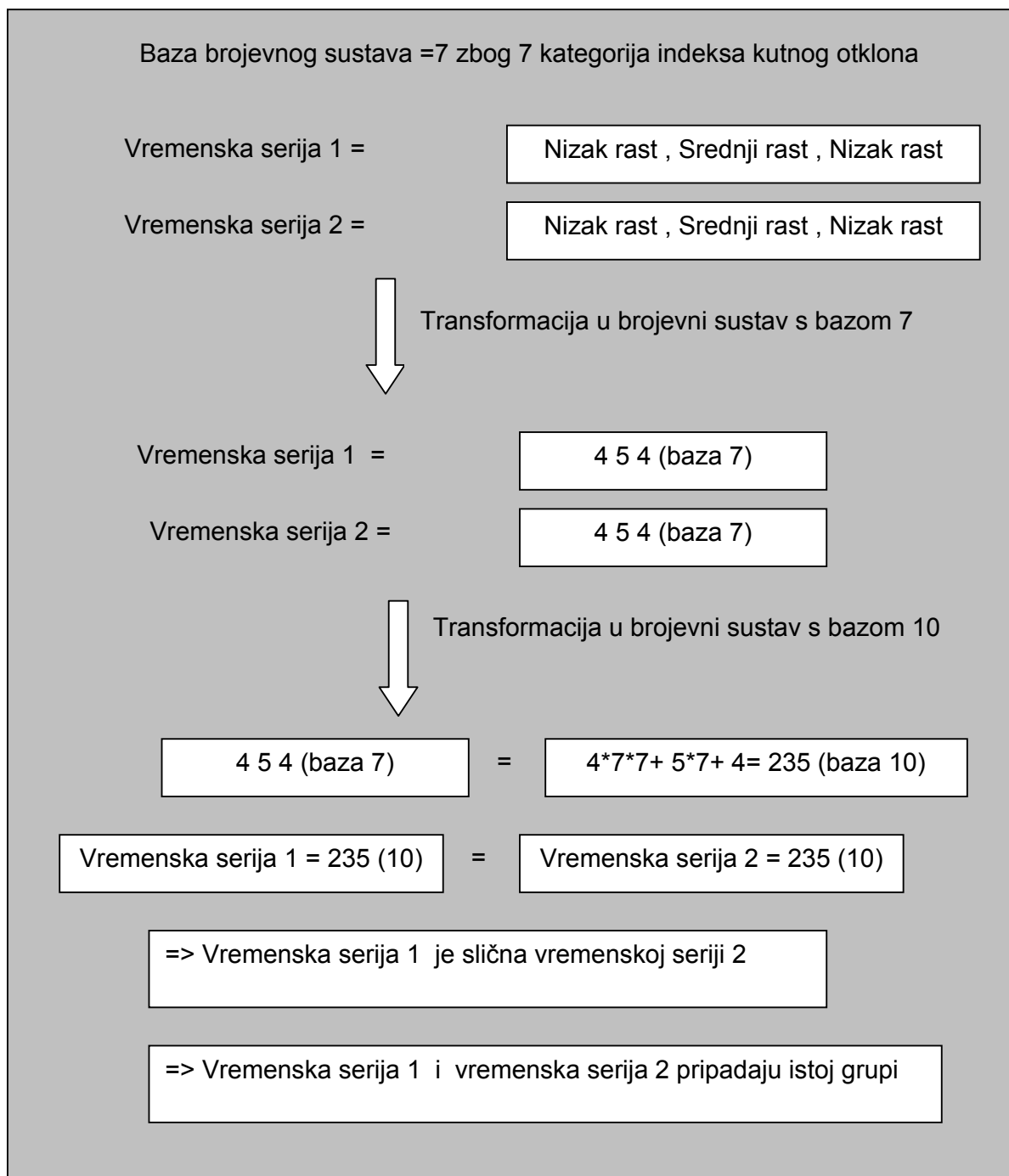
U svjetlu REFII modela temeljna definicija sličnosti izražena preko intervalnih jednakosti polazišna je osnova za različite vrste grupiranja bilo da se radi o grupiranju na temelju sličnosti čitavog intervala vremenske serije ili njenih dijelova.

Prvi korak kod operacije grupiranja svodi se na temeljnu transformaciju vremenske serije / vremenskih serija u REFII model.

Ovisno o cilju analize koji se može odnositi na grupiranje vremenskih odsječaka ili vremenskih serija na cijelom intervalu provodi se analiza sličnosti.

Osnovna ideja kod procjene sličnosti putem vremenske serije transformirane u REFII notaciju svodi se na transformaciju kategorija u numeričke vrijednosti, te pridruživanje tih numeričkih vrijednosti s obzirom na određenu pripadnost na vremenskim odsječcima REFII modela. Broj kategorija jednak je bazi brojevnog sustava, te se kasnije s obzirom na promatrani interval računa adekvatna numerička vrijednost sa bazom 10. Ako su te krajnje vrijednosti jednake za promatrane segmente ili vremenske serije, tada za njih kažemo da su slične, odnosno možemo ih grupirati u istu skupinu.

Ovaj postupak predložen je na slici 4.10. :



SLIKA 4.10. OTKRIVANJE SLIČNOSTI

Na osnovu metodologije prikazane na slici 4.10. moguće je tražiti sličnosti kako između dijelova vremenske serije, tako i između više vremenskih serija ili određenih intervala vremenskih serija te na osnovu sličnosti grupirati vremenske serije i odsječke. Ovaj je model primjenjiv u slučajevima kada očekujemo potpunu jednakost za sve odsječke na temelju intervalnih vrijednosti. Osnovna namjena ovakvog pristupa je davanje odgovora na pitanje da li su dva ili više nizova u potpunosti identični uzimajući u obzir

intervalne vrijednosti nizova u REFII modelu. Nedostatak ove metode očituje se u njenoj prevelikoj osjetljivosti prilikom pojave razlika intervalnih vrijednosti unutar vremenskih serija. Zbog toga se ovaj postupak koristi kako bi se spoznala sličnost među vremenskim serijama, pri čemu se očekuje potpuna jednakost na razini svih intervalnih vrijednosti vremenske serije, kako bi se donio sud o sličnosti između vremenskih serija.

Naredni jednostavni primjer ilustrira postupak procjene sličnosti između dvije vremenske serije. Transformacije po bazi 7 dobivene su na temelju tablice 4.12.

Tablica 4.12. Tablica sekundarne transformacije

Razred	Donja granica	Gornja granica	REF	Znamenke brojevnog sustava po bazi 7
Nizak rast	0.000000000	0.300000000	R	4
Srednji rast	0.300000000	0.700000000	R	5
Visok rast	0.700000000	1.000000000	R	6
Nizak pad	0.000000000	0.300000000	F	2
Srednji pad	0.300000000	0.700000000	F	1
Oštar pad	0.700000000	1.000000000	F	0
Bez promjene	0.000000000	0.000000000	E	3

Naredna tablica prikazuje način proračuna sličnosti

Tablica 4.13. Procjena sličnosti na temelju dvije vremenske serije

Seriya 1 u REFII notaciji	Seriya 2 u REFII notaciji	Transformacija serije 1 po bazi 7	Transformacija serije 2 po bazi 7
Nizak rast	Nizak rast	4	4
Nizak rast	Nizak rast	4	4
Nizak rast	Nizak rast	4	4
Nizak rast	Nizak rast	4	4
Nizak pad	Nizak pad	2	2
Nizak rast	Nizak rast	4	4
Nizak rast	Nizak rast	4	4
srednji rast	Srednji rast	5	5
Vrijednosti po bazi 10		729103	729103

Vrijednosti po bazi 10 dobivene su na temelju formule

$$4+4*7+4*7^2+4*7^3+2*7^4+4*7^5+4*7^6+5*7^6=729103.$$

Kako oba niza imaju identične vrijednosti, možemo zaključiti kako su ta dva niza slična. Što je različita vrijednost bliže kraju procjenjivane vremenske serije, to će rezultati ove metode pokazati veća odstupanja. Upravo iz tog razloga, ovu metodu koristimo samo za procjene sličnosti od kojih se očekuje potpuna jednakost na razini svih intervalnih vrijednosti vremenskih serija. Prednost ovakvog pristupa očituje se u brzom mehanizmu za procjenu sličnosti između vremenskih serija, koji je posebno pogodan za pretragu motiva u vremenskoj seriji [Chiu, 2003], [Lin, 2001], [Patel, 2002], [Tanaka, 2003], odnosno pretraži sličnog zadanog uzorka dužine kn u vremenskoj seriji dužine m.

Metodologija koja kod procjene sličnosti nije toliko osjetljiva na razlike između vrijednosti vremenskih odsječaka, oslanja se na proračun udaljenosti između svakog

vremenskog odsječka. Udaljenost možemo računati na temelju formule za Euklidsku distancu [Han, 2001] :

$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2}$$

gdje su i i j vremenske serije $i=(x_{i1}, x_{i2}, \dots, x_{ip})$, $j=(x_{j1}, x_{j2}, \dots, x_{jp})$

Naredni primjer ilustrira procjenu sličnosti između vremenskih serija u REFII modelu pomoću primjene funkcije udaljenosti. U narednom jednostavnom primjeru za procjenu sličnosti koristit će se vrijednosti kutnog otklona u izvornom obliku. Osim vrijednosti kutnog otklona, sukladno potrebama analize, moguće je istovremeno sličnost procjenjivati i na temelju površine ispod krivulje.

Korištenje funkcija udaljenosti u svrhu pronalaženja sličnosti vremenskih serija vrlo je popularna u znanstvenim radovima [Lin, 2001], no funkcije udaljenosti se primjenjuju uglavnom na izvornim vrijednostima vremenskih serija.

U narednom ilustrativnom primjeru, prikazanom u tablici 4.14., gdje procjenjujemo da li je vremenska serija A sličnija vremenskoj seriji B ili C koristimo funkcije udaljenosti na vrijednostima kutnih otklona, sastavnim dijelom REFII metoda.

Tablica 4.14. Procjena sličnosti na temelju dvije vremenske serije

Koeficijenti kutnog otklona vremenske serije A	Koeficijenti kutnog otklona vremenske serije B	Koeficijenti kutnog otklona vremenske serije C	A->B	A->C
0,33	0,43	0,43	0,177764	0,12041595
0,6	0,7	0,66		
0,5	0,5	0,5		
0,1	0,2	0,1		
0,3	0,34	0,33		

Procjene sličnosti u tablici 4.14. izračunate su na temelju formule Euklidskih distanci:

$$A \rightarrow B = (\text{SQRT}((0,33-0,43)^2 + (0,6-0,7)^2 + (0,5-0,5)^2 + (0,1-0,2)^2 + (0,3-0,34)^2)) = 0,177764.$$

Na isti način izračunat je odnos sličnosti između vremenskih serija A i C. Kako je distanca između serija A i C manja, zaključujemo da su serije A i C sličnije nego serije A i B.

Ovo je jednostavan primjer, koji dobiva na kompleksnosti kada primjerice imamo N vremenskih serija i želimo grupirati vremenske serije na temelju sličnosti. Isto tako ova metodologija nam može biti korisna kod pretrage motiva u vremenskoj seriji.

Između ostalog ovaj primjer pokazuje zašto je u strukturama koje čuvaju transformirane vrijednosti u REFII notaciji potrebno pohraniti kako intervalne vrijednosti, tako i izvorne vrijednosti kutnih otklona i površine, jer različite analize zahtijevaju različite skupove ulaznih vrijednosti.

Ako želimo znati, primjerice u kojim segmentima se dvije vremenske serije razlikuju, potrebno je proračunavati parcijalne udaljenosti, prvo za dužinu uzorka 1, pa nakon toga uzorci koji pokazuju sličnost ulaze kao kandidati u daljnju analizu, ako što je to prikazano u algoritmu za otkrivanje sezonskih oscilacija, te u algoritmu za otkrivanje frekventnih uzoraka. U tom slučaju ove algoritme moramo osnažiti samoorganizirajućim mapama i stablima frekventnih uzoraka.

4.7.3. Rezultati otkrivanja sličnih grupa na empirijskim podacima

Podaci za analizu sličnosti vremenskih odsječaka uzeti su sa :

http://www.stat.duke.edu/data-sets/mw/ts_data/gdp i predstavljaju bruto nacionalni proizvod za zemlje : Austrija, Kanada, Francuska, Njemačka, Grčka, Italija, Švedska, UK, SAD u periodu od 1950-1983.

Cilj istraživanja je otkrivanje sličnih zemalja, odnosno deklariranje grupa s obzirom na trend kretanja Bruto nacionalnog proizvoda s obzirom na zadani kontinuirani vremenski raspon.

Cilj istraživanja inspiriran je na temelju rada [Lin, 2001], koji se bavi otkrivanjem motiva i grupiranjem vremenskih serija na temelju otkrivenih motiva.

Alternativni pristup analizi koji je moguće primijeniti s obzirom na mogućnosti REFII modela je otkrivanje uzoraka sličnosti unutar jedne vremenske serije i na taj način provesti grupiranje, kao i provesti grupiranje vremenskih serija s obzirom na generalno zadan postulat da vremenska serija pripada grupi "A" *uz uvjet da postoji potpuna sličnost sa ostalim vremenskim serijama u grupi "A"*.

To bi značilo da vremenska serija pripada grupi "A", ako postoji intervalan jednakost svih vremenskih odsječaka sa odsječcima vremenskih serija koje pripadaju grupi "A".

Pojam sličnosti može biti i fleksibilniji pojam od navedenog [Cheng, 1997] , [Prat, 2001], što bi značilo da se u REFII model sličnosti može uvesti i pojam tolerancije, što bi značilo da analitičar dopušta do određene granice razlike po pitanju intervalnih jednakosti.

Na taj način možemo definirati stupanj tolerancije od primjerice 1%, što bi značilo da vremenske serije T_1 i T_2 možemo deklarirati sličnima i grupirati ih u istu skupinu ako intervalna jednakost nije registrirana do maksimalno 1% vremenskih odsječaka.

Cilj analize na spomenutom skupu podataka bit će usredotočen na otkrivanje sličnosti vremenskih serija na zadanom intervalu s ciljem provođenja grupiranja.

Za potrebe navedene analize definirani su razredi otklona kutova i pripadajuće znamenke brojevnog sustava po bazi 7 kao što je vidljivo u narednoj tablici 4.12.

Nakon transformacije podataka o bruto nacionalnom proizvodu u REFII notaciju za oblik krivulje dobivamo matricu narednog oblika:

Tablica 4.15. Tablica primarne transformacije oblika krivulje dijela podataka za analizu

ciklus	aust	can	fra	ger	gree	ital	swe	uk	usa
1951	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast
1952	Nizak pad	Nizak rast	Nizak rast	Nizak rast	Nizak pad	Nizak rast	Nizak pad	Nizak rast	Nizak rast
1953	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast
1954	Nizak rast	Nizak pad	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak pad
1955	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast
1956	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast
1957	Nizak rast	Nizak pad	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast
1958	Nizak rast	Nizak pad	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak pad	Nizak pad
1959	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast
1960	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Bez promjene	Nizak rast
1961	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Nizak rast	Bez promjene
...

Kako bi se mogle provoditi analize sličnosti a samim time i grupiranja potrebno je prema predloženom algoritmu izvršiti transformaciju prikazanih vrijednosti u znamenke brojeva po bazi 7.

Kao rezultat takve transformacije dobiva se tablica:

Tablica 4.16. Tablica sekundarne transformacije oblika krivulje dijela podataka za analizu

ciklus	t_aut	t_can	t_fra	t_germ	t_gre	t_ital	t_swe	t_uk	t_usa
1951	4	4	4	4	4	4	4	4	4
1952	2	4	4	4	2	4	2	4	4
1953	4	4	4	4	4	4	4	4	4
1954	4	2	4	4	4	4	4	4	2
1955	4	4	4	4	4	4	4	4	4
1956	4	4	4	4	4	4	4	4	4
1957	4	2	4	4	4	4	4	4	4
1958	4	2	4	4	4	4	4	2	2
1959	4	4	4	4	4	4	4	4	4
1960	4	4	4	4	4	4	4	3	4
1961	4	4	4	4	4	4	4	4	3
...

Ovako formirana matrica sličnosti temelj je za provođenje analize sličnosti i grupiranje vremenskih odsječaka.

Da bi se spoznala sličnost na određenom intervalu, dovoljno je taj interval kojeg predstavlja niz znamenaka po bazi 7 pretvoriti u broj po bazi 10, te procijeniti da li postoji ili ne postoji sličnost na zadanom intervalu.

Istraženo je da li postoji sličnost između bruto nacionalnog proizvoda u periodu 1961-1970

za zemlje Austrija, Kanada, Francuska, Njemačka, Grčka, Italija, Švedska, UK i SAD.

Tablica 4.17. Tablica kategorija sličnosti

Zemlja –(1961-1970)	Vrijednost po bazi 10	Grupa
Austrija	3910428	1
Kanada	3910428	1
Francuska	3910428	1
Njemačka	3876814	2
Grčka	3910428	1
Italija	3910428	1
Švedska	3910428	1
UK	3910428	1
USA	2263342	3

Na temelju provedene analize vidljivo je da jedino USA i Njemačka u danom periodu ne spadaju u istu grupu sa ostalim zemljama. Grupe u koju padaju USA i Njemačka

nemaju analitičku težinu, zbog spomenutih ograničenja ove metode. Grupa koju karakteriziraju članovi sa identičnim vrijednostima po bazi 10 smatra se relevantnom, jer članovi te grupe pokazuju sličnost u oblicima vremenske serije. Ta sličnost se očituje u potpunoj jednakosti intervalnih vrijednosti svih odsječaka.

Daljnijim analizama gdje se je željelo doći do odgovora do kojeg razdoblja su Njemačka i USA spadale u istu grupu sa ostalim zemljama, otkriveno je da je Njemačka slična ostalim zemljama u periodu od 1961-1966, a USA do 1969.

Algoritam na temelju kojeg su dobiveni ovi rezultati sekvencijalno računa vrijednosti po bazi 10 od 1961 godine za svaku zemlju, te nakon svake iteracije u naredni krug izračuna uzimaju se kandidati koji su u prethodnoj iteraciji imali jednake vrijednosti.

Njemačka je imala različitu vrijednost sa trendom na 1967, a SAD su imale različitu vrijednost sa trendom na 1970 godinu.

Vremenski raspon koji je uzet kao vremenski okvir analize u znanstvenim radovima se nazivaju vremenski prozori (eng. time window) [Manilla, 1997], [Keogh, 2001]. Vremenski prozori imaju širok spektar primjene kod upitnih jezika na vremenske serije (eng time series SQL) [Sardi, 2001].

Kod preciznijeg određivanja grupa, pa tako i sličnosti u obzir se mogu uzeti i površine ispod krivulje, pa primjerice prema tom kriteriju Austrija i Švedska zasigurno ne bi pripale istoj grupi zbog male površine ispod krivulje koja pripada Austriji i velike površine ispod Švedske u odnosu na Austriju.

O analitičaru ovisi fokusiranost na oblik krivulje i/ili na voluminoznost pojave, te deklaracije pripadnosti grupi s obzirom na te dvije karakteristike.

Sličnost, kakva je ovdje definirana i upotrijebljena za grupiranje, može se iskoristiti i u kreiranju upitnog jezika za vremenske serije [Perng, 2002]. Takav upitni jezik može davati odgovore na pitanja tipa :

- Da li postoji sličnost između vremenske serije T_1 i T_2 u vremenskom rasponu od V_1 do V_2 (Da li pripadaju istoj grupi) [Tanaka, 2003], [Lin, 2001], [Yi, 1998]
- Da li u vremenskoj seriji T_1 postoji motiv M , [Patel, 2002]
- Kolika je frekvencija pojavnosti motiva M u vremenskoj seriji T_1
- Da li su vremenske serije $T_1, T_2, T_3, \dots, T_n$ slične [Popivanov, 2002], [Rafiei, 1998]
- Da li postoji međuzavisnost pojave motiva M_1 i motiva M_2
- Da li postoji korelatorna međuzavisnost trendova dvije vremenske serije i koja

Ovo su samo neka od pitanja na koje možemo dobiti odgovor primjenjujući definiciju sličnosti kroz upitni jezik na vremenske serije [Huang, 1999], [Perng, 2002].

Preciznost i detaljnost analize ovisi o cilju, a preciznost i detaljnost određuje analitičar izborom kako elemenata koji ulaze u analizu tako i distribucijom razreda za koeficijente kutnog otklona i površine ispod krivulje.

Slični segmenti mogu biti deklarirani kao epizode, te se nad njima mogu vršiti daljnje analize poput izravnog otkrivanja pravila, ili proračuna uvjetnih vjerojatnosti.

4.8. Formalna logika, ekspertni sustavi i REFII model

Konceptualnom primjenom REFII modela otkrivamo pravila o promatranoj vremenskoj seriji. Analize koje se odnose na otkrivanje cikličkih oscilacija, sezonskih oscilacija, epizoda, potencijalnih spavača, te dijagnostiku grupe vremenskih odsječaka kao rezultat obrade nude pravila o promatranoj vremenskoj seriji. Niz takvih pravila daje znanje o skrivenim pravilima unutar vremenske serije. Ova pravila mogu egzistirati

samostalno kao pojedinačne činjenice, ali se isto tako mogu formalizirati u vidu pravila u formatu AKO-ONDA. Prednost ovakvog pohranjivanja znanja o vremenskim serijama je višestruka [Ohsaki, 2003]. U prvom redu niz periodičnih analiza na istom setu podataka (vremenskoj seriji) može biti usporediv sa perspektive vremenske dimenzije. Daljnja prednost proizlazi iz činjenice da ovako strukturirana pravila možemo ulančavati i otkrivati nova, izvedena znanja uz pomoć algoritama za ulančavanje.

Vrlo značajna prednost ovakve koncepcije očituje se i u mogućnosti objedinjavanja rezultata različitog konceptualnog karaktera, gdje formalna logika postaje vezni element za daljnju produkciju znanja.

Tako primjerice u bazi činjenica možemo pohraniti činjenice vezane uz analize temeljene na REFII modelu, deskriptivnoj statistici, klasičnim statističkim analizama vremenskih serija i sl.

Ljuska ekspertnog sustava tako postaje snažan alata za otkrivanje novog znanja o vremenskim serijama na temelju analitičkog znanja.

Zamislamo situaciju u kojoj pratimo ciljne skupine okarakterizirane kao vrijednost atributa, ili niza atributa u vremenskoj seriji unutar određenog vremenskog odsječka.

Analiza ove vremenske serije može rezultirati nizom pravila proizašlih iz te vremenske serije. Ovakve analize provodimo periodično, i ti periodični rezultati analiza pune se u bazu činjenica. Niz ovako provedenih analiza u različitim vremenskim periodima nad vremenskim odsječcima, pruža nam mogućnost komparativne analize tržišnih trendova s obzirom na ciljne segmente tržišta.

Na temelju ovog primjera možemo analizirati :

Kontinuitete i sezonske oscilacije otkrivenih epizoda

Stabilnost sezonskih oscilacija

Stabilnost i mutacije cikličkih oscilacija

Korelatorne odnose između različitih pojava unutar vremenskih serija

Ovdje odsječke vremenskih serija promatramo kao niz pravila u formi AKO –ONDA, i koristimo mehanizme ulančavanja prema naprijed i prema natrag za otkrivanje znanja. U ovakav sustav možemo uvesti i pojam meta pravila, odnosno pravila o pravilima, što znači da pojam pravila možemo uzdići na višu razinu.

Ovakva koncepcija vrlo je bliska sustavima za prikupljanje obradu i generiranje znanja, s tom razlikom što je ovakav sustav prvenstveno orijentiran na manipulaciju znanjem o vremenskim serijama.

Ovakav pristup omogućava efikasnu pohranu izgeneriranog znanja na temelju periodičnih analiza, te dugoročno gledajući nakon izvršenog niza periodičnih analiza, ova baza znanja može se iskoristiti za otkrivanje zakonitosti koje proizlaze iz otkrivenih pravilnosti na razini vrlo širokog vremenskog opsega.

Formalna logika, je najviši stupanj transformacije koji u sebi sadrži znanje o vremenskoj seriji. Ako krenemo analizirati stupnjeve transformacije kroz prizmu REFII modela od najnižeg prema najvišem onda ih možemo klasificirati na slijedeći način:

Transformacija vremenskih odsječaka vremenske serije u vremenske odsječke višeg stupnja sa mjerilom vremenske kompleksnosti =1

Transformacija vremenskih odsječaka iz točke 1 u REFII notaciju

Transformacija elemenata iz točke 2 u elemente "čitljive" klasičnim *data mining* algoritmima

Transformacija znanja proizašlih kao rezultat analize u AKO –ONDA pravila

Transformacija AKO-ONDA pravila na temelju meta-pravila

Što je veći stupanj transformacije (1-5) to je veći stupanj apstrakcije. Svaki od ovih stupnjeva transformacije možemo shvatiti kao pretprocesiranje podataka, kako bismo mogli izvršiti određeni proces obrade nad vremenskim serijama.

Primjena formalne logike pokazuje mogućnost strukturiranja i obrade analitičkog znanja proizašlog iz REFII modela metodologijom podudaranja uzoraka.

U sustavu koji ima povratnu vezu nakon što se otkriju pravilnosti u nizu ciklusu analiza i prihvate kao činjenice, ove se činjenice (radi optimizacije vremena obrade) mogu testirati kao hipoteze.

Primjerice, ako otkrijemo sezonsku oscilaciju prvog tjedna u mjesecu i ta se oscilacija pojavljuje u n analitičkih perioda, u analitičkom periodu $n+1$ tražit ćemo potvrdu hipoteze o sezonskoj oscilaciji prvog tjedna u mjesecu.

Alternativni pristup kreiranju ekspertnih sustava odnosi se na korištenje fuzzy logike.

Uzmimo primjerice problematiku dijagnostike bolesti kod pacijenata.

Na primjeru promatramo varijable, kao što su :

- Temperatura tijela
- Jačina boli u želucu
- Mučnina
- Trend razine bijelih krvnih zrnaca
- Oteknuće limfnih čvorova
- ...

Kvantizacija koja se provodi za vremenski uvjetovane varijable (REFII model) može se provesti na nekoliko načina:

1. Analizom rječnika eksperata kojim oni opisuju pojedine veličine, te određivanje kvantuma odnosno tipičnih leksičkih varijabli i njihove definicije kao neizrazitih veličina
2. Kvantizacijom kao npr. kod tehnika kompresije podataka.

Tako možemo definirati temperaturu tijela kao:

- nisku
- normalnu
- povišenu
- visoku

Jačinu boli u želucu možemo skalirati u intervalu od 0-1 kao :

- Blagu
- Srednju
- Jaku
- Izrazito jaku

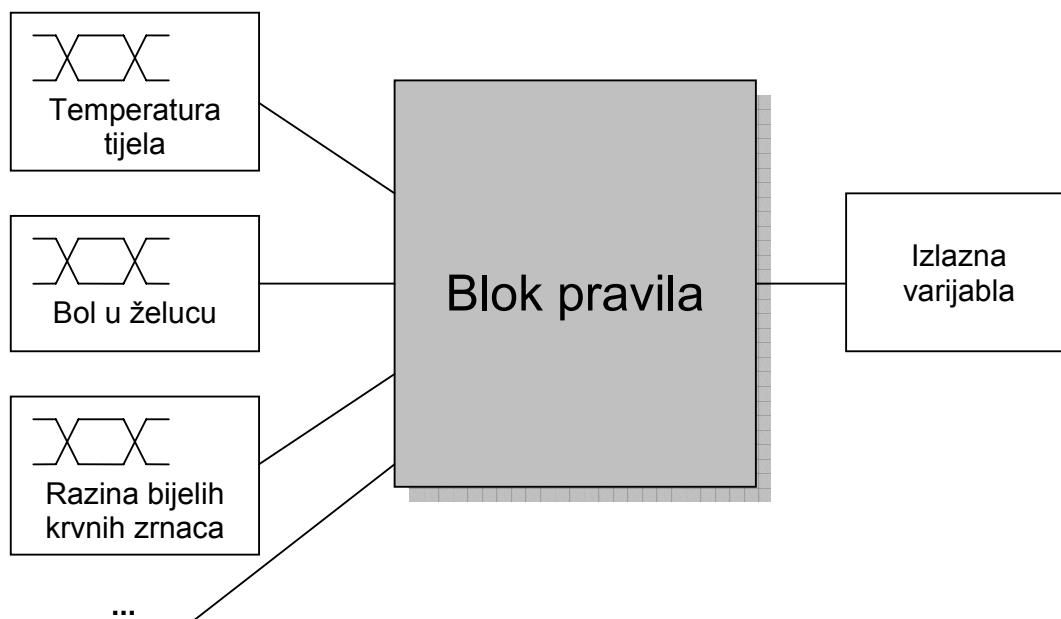
Trend razine bijelih krvnih zrnaca možemo deklarirati uz pomoć REFII modela kao

- Jaki rast
- Jaki pad
- ...

Tako redom možemo kreirati vrijednosti svih lingvističkih varijabli.

Ozbiljniji medicinski sustavi mogu imati pobrojane sve moguće simptome i oni se mogu međusobno povezivati u grumene znanja s ciljem dijagnostike bolesti. Ovo je ilustrativan primjer koji samo pokazuje osnovne principe izgradnje *fuzzy* medicinskih ekspertnih sustava koji u sebi objedinjavaju i vremensku komponentu temeljenu na *fuzzy* logici.

Nakon što formiramo lingvističke varijable, možemo ih povezivati u blokove pravila (eng. Rule block) kao što je to prikazano na slici 4.11.



SLIKA 4.11.. LEKSIČKE VARIJABLE I BLOKOVİ PRAVILA

U izlaznoj varijabli definiramo moguće bolesti, a blokovi pravila u sebi sadrže mehanizme koji na temelju *fuzzy* logike daju krajnje vrijednosti.

Strukturu blokova pravila možemo prikazati na slijedeći način:

Tablica 4.18. Blokovi pravila

Temperatura tijela	Bolovi u želucu	Mučnina	Razina bijelih krvnih zrnaca	Oteknuća limfnih čvorova	...	Tip bolesti
Visoka	Blagi	Jaka	Jaki rast	Slaba	...	Bolest 1
Normalna	Izrazito jaka	Nema	Jaki pad	Nema	...	Bolest 2
Povišena	Nema	Nema	Blagi rast	Jaka	...	Bolest 3
...

Gdje je tip bolesti predstavlja izlaznu varijablu.

Na prikazanom primjeru nije vidljiva sva kompleksnost *fuzzy* ekspertnih sustava, jer postoji mogućnost definicije i virtualnih varijabli, kombinacije izlaznih varijabli koje postaju ulazne varijable u neki novi blok pravila. Tako primjerice sustavi mogu biti

sposobni otkrivati kako bolesti, tako i potencijalne opasnosti od drugih oboljenja koje su vezane uz tu bolest ili grupu bolesti s obzirom na postojeće simptome.

5. Metodologija primjene REFII modela

5.1. Važnost temeljnog pretprocesiranja izvornih podataka iz vremenskih serija

5.1.1. Primarna transformacija vremenske serije u REFII model

Kao što je već rečeno, tradicionalni pristup *data mining* analizi vremenskih serija orijentiran je provođenje parcijalnih nepovezanih postupaka analize, te se uglavnom oslanja na vizualizaciju vremenskih serija [Pyle, 2001], [Han, 2000].

Postoji čitava paleta *data mining* metoda za analizu vremenskih serija, koje su međusobno nepovezane i nekompatibilne, te zbog toga ne postižu sinergijski efekt analize, kao što je to moguće postići za podatke sa nevremenskom komponentom.

Važnost i pokušaji rješenja ovog problema vidljivi su u znanstvenim radovima koji pokušavaju objediniti nekoliko tipova analiza te nude parcijalno rješenje fokusirajući se na model transformacije [Lin, 2003], [Pratt, 2001], [Han, 2000].

Osnovni problem ovakvih rješenja proizlazi iz njihove vezanosti za rješavanje problema za koji su konstruirani.

Idejno najbliži koncepciji REFII modela su modeli koji vremenske odsječke transformiraju u niz znakovna [Pratt, 2001]. Njihov je nedostatak u nefleksibilnosti, i nejednoznačnosti transformacije, te nemogućnosti vezivanja modela transformacije sa tradicionalnim metodama *data mininga*.

Transformaciju izvorne, granulirane vremenske serije u REFII notaciju nazivamo primarnom transformacijom.

Primarna transformacija, detaljno obrađena u 2. poglavlju je temelj sustava i analitičkih postupaka koji su bazirani na REFII modelu.

Ovaj model pruža fleksibilnost analitičaru na način da može analizu učiniti preciznijom/globalnijom na način :

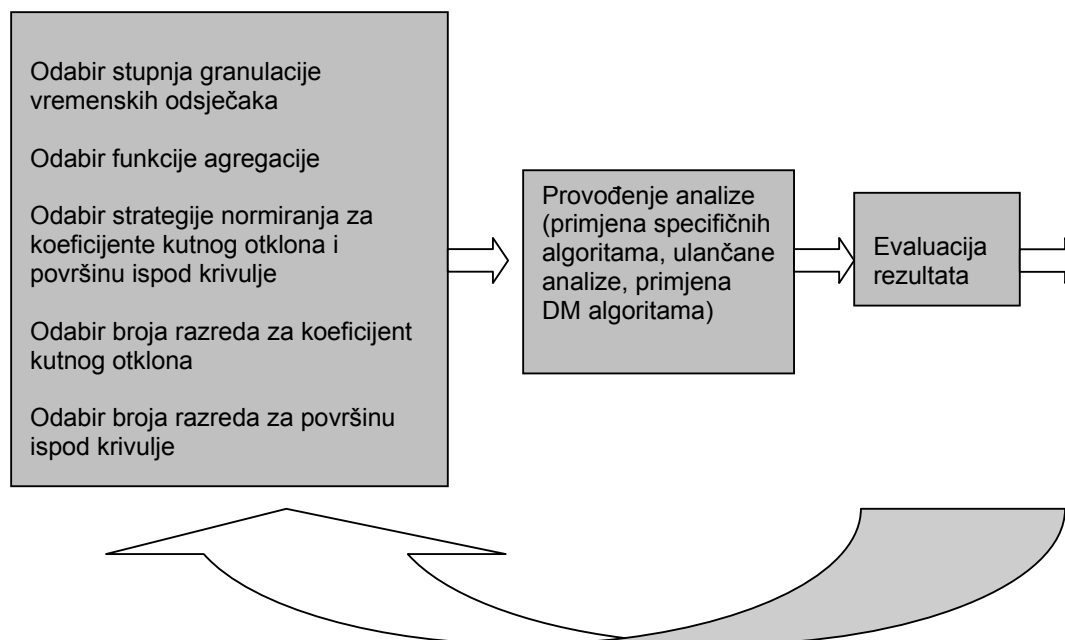
- mijenjanja stupnja granulacije vremenskih odsječaka
- mijenjanje broja razreda koeficijenta kutnog odklona
- mijenjanje broja razreda površine ispod krivulje
- promjene funkcija sažimanja (prosjeak, suma, mod...)
- promjena strategije normiranja vrijednosti unutar vremenske serije

Što se tiče normiranja, s obzirom na cilj analize minimalna i maksimalna vrijednost mogu biti preuzete iz empirijskih podataka, ili pak mogu biti izabrane na temelju ekspertnog znanja analitičara.

U slučaju analize niza vremenskih serija kao minimalne i maksimalne vrijednosti moguće je uzeti minimalnu i maksimalnu vrijednost svake pojedine vremenske serije, ili je moguće uzeti globalni minimum, odnosno maksimum na razini svih vremenskih serija.

Sa ovih pet osnovnih instrumenata analitičar može modelirati stupanj preciznosti analize posredstvom REFII modela.

REFII model uzima u obzir spiralni pristup u razvoju modela *data mininga* [Berry, 1997], te omogućava povratak na određeni korak tijekom analize s ciljem postizanja što boljih rezultata analize kao što je to prikazano na slici 5.1.



SLIKA 5.1. SPIRALNI PRISTUP ANALIZE I REFII MODEL

Iz slike 5.1. vidljivo je da analitičar može analizirati vremensku seriju na temeljima spiralnog pristupa, te može utjecati na tijek same analize i traženje zakonitosti unutar vremenskih serija.

Za razliku od tradicionalne *data mining* analize sa podacima netemporalnog karaktera kada primjerice koristimo stablo odlučivanja, te particioniramo osnovnu populaciju koja ulazi u analizu na uzorak za treniranje i uzorak za testiranje, te je taj proces dio spiralnog pristupa analize kod REFII modela postoji stanovita razlika. Ta se razlika očituje u drugačijem karakteru temporalnih podataka i samoj koncepciji REFII modela. Prilikom traženja zakonitosti u vremenskim serijama promjenom inicijalnih vrijednosti parametara kao što je granulacija, ili broj razreda kutnog otklona, pojava koja nije pokazivala zakonitosti na nižim razinama granulacije, može pokazivati zakonitost na višim razinama granulacije.

Ako primjerice želimo analizirati sezonske oscilacije, postoji mogućnost da vremenska serija na razini vremenske granulacije –dani u tjednu i broju razreda 7 ne pokazuje zakonitosti u sezonskim oscilacijama. Ako promijenimo stupanj granulacije na tjedne vremenska serija može pokazivati zakonitosti.

Daljnja karakteristika ovakvog pristupa, svodi se na moguću promjenu strategije u primjeni analitičkih postupaka. Tako se u narednom ciklusu mogu ulančati novi niz metoda, ili promijeniti analitičke metode s ciljem evaluacije dobivenih rezultata.

5.1.2. Sekundarna transformacija REFII modela

Kako je temeljna ambicija REFII modela postizanje kompatibilnosti između različitih metodologija i koncepcija analize vremenskih serija, kao i pružanje temelja za izgradnju nestandardnih analitičkih rješenja, sekundarna transformacija unutar samog modela služi za premošćivanje različitih metodoloških koncepcija.

Primjeri sekundarnih transformacija su :

- Transformacija u epizode
- Transformacija u motive
- Transformacija u nizove sličnosti

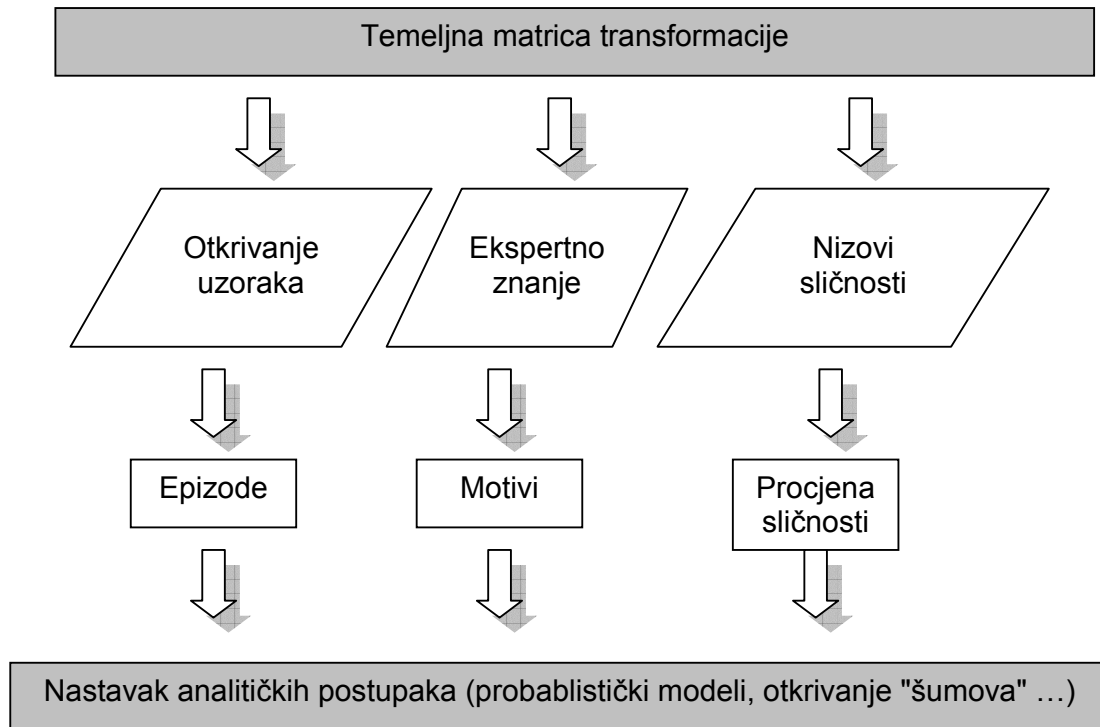
Elementi sekundarne transformacije su nadkategorije koje su formirane ne temelju elemenata proizašlih iz primarne transformacije. Sekundarna transformacija može se provesti na temelju algoritamske obrade elemenata iz primarne transformacije (matrice transformacije) [Manilla, 1997] , na temelju ekspertnog znanja [Patel, 2002], [Lin, 2001], ili pak za potrebe specifičnih vrsta analize (nizovi sličnosti).

Kada govorimo o sekundarnoj transformaciji temeljenoj na algoritamskoj obradi, tada uzimajući u obzir spomenute sekundarne transformacije prvenstveno se podrazumijeva otkrivanje uzoraka koji se mogu proglasiti epizodama, te se mogu kao epizode uključiti u tijekove daljnje analitičke obrade.

Kada govorimo o sekundarnoj transformaciji temeljenoj na ekspertnom znanju, tada uzimajući u obzir sekundarne transformacije spomenute u radu prvenstveno se podrazumijeva otkrivanje unaprijed zadanih uzoraka (motiva).

Sekundarna transformacija u nizove sličnosti podrazumijeva transformaciju kategoriju otklona kutova θ i / ili površina ispod krivulja u niz brojevnih vrijednosti sustava sa brojevnom bazom koja se određuje na osnovu broja kategorija razreda. Nakon ovog postupka slijedi transformacija u dekadski brojevni sustav i primjena euklidskih distanci s ciljem procjene sličnosti, kao što je to prikazano u poglavlju 4.7.

Metodologija provođenja sekundarne iz primarne transformacije prikazana je na slici 5.2.



SLIKA 5.2. METODOLOGIJA PROVOĐENJA SEKUNDARNE IZ PRIMARNE TRANSFORMACIJE

Iz slike 5.2., vidljivo je da sekundarna transformacija omogućava nastavak analitičkih postupaka nad vremenskom serijom. Kada je primjerice cilj otkriti uzorke, tada nakon uspješno obavljenog zadatka, ne moramo otkrivene uzorke deklarirati kao epizode. No ako se žele spoznati međusobni utjecaji otkrivenih uzorka, tada je potrebno otkrivene uzorke definirati kao epizode, te ih je moguće uklopiti u probablistički model Bayesove mreže [Jensen, 2001] s ciljem promatranja međuzavisnosti. Motive [Patel, 2002], [Lin, 2001] je moguće konstruirati direktno na temelju matrice transformacije, te se na osnovu definiranih motiva nastavljaju analitički procesi. Kod procjene sličnosti, s obzirom na posebnost područja, matrica koja je formirana na temelju primarne transformacije, služi kao predložak za sekundarnu transformaciju, gdje se koristi nestandardno rješenje za procjenu sličnosti i grupiranje. Metodologija sekundarne transformacije zadovoljava kriterij ulančavanja analiza vremenskih serija, te može biti ravnopravno uključena u proces analize s obzirom na očekivani cilj.

5.1.3. Građenje stabla vremenskih indeksa

Vrlo važna komponenta REFII modela integrirana u matricu transformacije su vremenski indksi, predstavljeni indeksima vremenskog odsječka u matrici. U kompleksnim sustavima ova komponenta mora biti vrlo fleksibilna, na način da s obzirom na željeni stupanj granulacije vremenskih serija za potrebe analiza u ovisnosti o selekciji vremenskih indeksa sustav vrši primarnu i sekundarnu transformaciju vremenske serije u REFII model.

Poštujući načelo fleksibilnosti, koje analitičaru omogućuje *data mining* analize vremenske serije posredstvom spiralnog pristupa, gdje je unaprijed nemoguće odrediti

kako potrebne analize, tako i potrebni stupanj granularacije kako bi se došlo do pravilnosti unutar vremenskih serija.

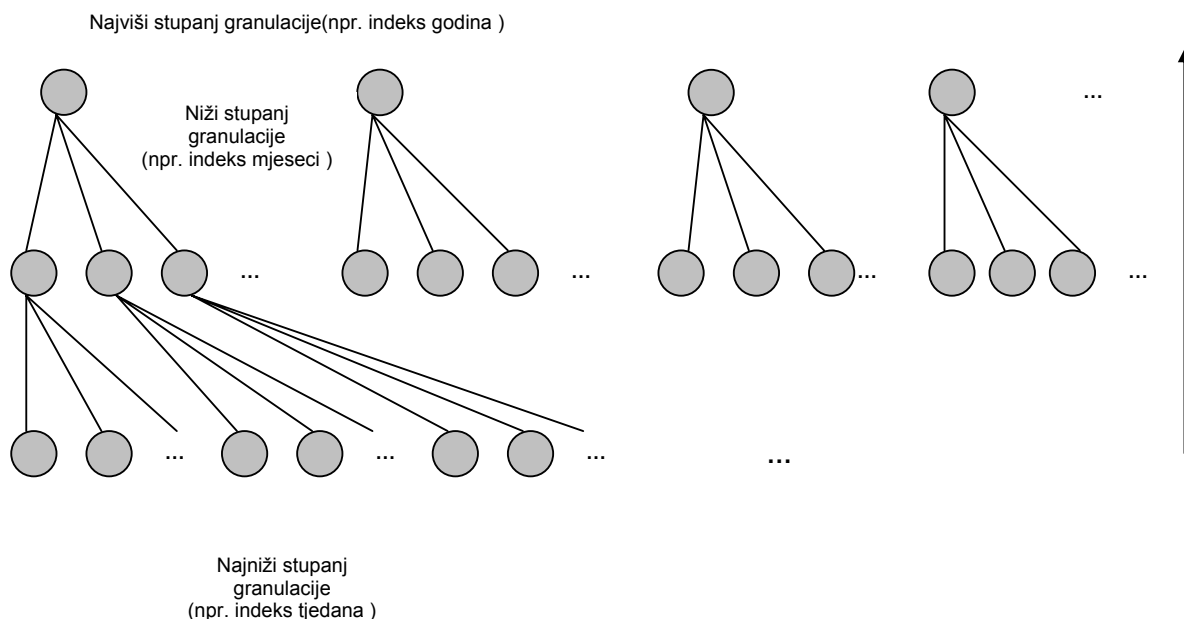
Ako se primjerice žele otkriti sezonske oscilacije, tada u inicijalnoj etapi možemo analizirati postojanje sezonskih oscilacija na najnižem stupnju granularacije. Ovisno o postojanju/ nepostojanju sezonskih oscilacija možemo povećavati stupanj granularacije vremenske serije shodno tome sustav vrši transformaciju u REFII notaciju s obzirom na zadani granularitet.

Ovaj postupak se može ponavljati sve dok se ne pronađe sezonska pravilnost na nekom od stupnja granularacije, ili dok se ne dođe do maksimalnog stupnja granularacije.

Pomoć prilikom realizacije analize vremenskih serija posredstvom REFII modela, prvenstveno prateći metodologiju spiralnog sustava analize su vremenski indeksi.

Vremenski indeksi nemaju samo utjecaja u primarnim procesima transformacije, nego isto tako sudjeluju kao gradivni elementi sustava u procesima sekundarne transformacije.

Slika 5.3. na prikazuje primjeru metodologije izgradnje vremenskih indeksa u REFII modelu.



SLIKA 5.3. PRIMJER METODOLOGIJE IZGRADNJE VREMENSKIH INDEKSA

Stablasta struktura indeksa kakva je prikazana na slici 5.3. omogućava spiralni pristup analize vremenskih serija. Prilikom provođenja analiza u analitičku obradu se uzima kategorija i potkategorija indeksa.

Ako analiziramo sezonske oscilacije, tada se prilikom analize mogu u obzir uzeti godine kao kategorija i mjeseci kao potkategorija, što je vidljivo u poglavlju 4.2.3.

Kompleksniji sustavi na ovaj način mogu vršiti analize uzimajući u obzir elemente vremenske granularacije.

Postoje slučajevi kada u analitičkom procesu ne moraju nužno sudjelovati samo kategorija indeksa i njegova potkategorija, već i nadkategorije sa većim stupnjem granularacije i u kombinaciji sa bilo kojom kategorijom bez obzira na stupanj granularacije.

Primjer za to je otkrivanja uzoraka, gdje su podaci granulirani na razini dana. Ako se primjerice otkrije postojanje uzorka na razini dana, a postoji još n nadkategorija (tjedni,

mjeseci kvartali, godine), moguće je provesti direktnu analizu pojavnosti uzoraka s obzirom na najvišu kategoriju – godine.

Osnovno pravilo građenja stabla vremenskih indeksa s obzirom na podatke u izvornim vremenskim serijama je da je najniži mogući stupanj granulacije vremenskog indeksa jednak stupnju granulacije izvorne vremenske serije.

Naravno u svakom je slučaju potrebno uzeti u obzir (prema poglavlju 2.6.) obavljanje vremenske interpolacije nad izvornom vremenskom serijom/serijama, kako bi se nad tako kreiranom vremenskom serijom mogli vršiti viši stupnjevi granulacije, odnosno kreirati vremenski indeksi više hijerarhijske razine.

Osnovni fokus analize vremenskih serija ne moraju nužno biti smo i isključivo elementi matrice transformacije, već u analizu ravnopravno mogu ući i vremenski indeksi.

Na taj način možemo u postupak analize uz elemente matrice transformacije koji se odnose na vrijednosti proizašle iz vremenske serije uključiti i vremenske indekse.

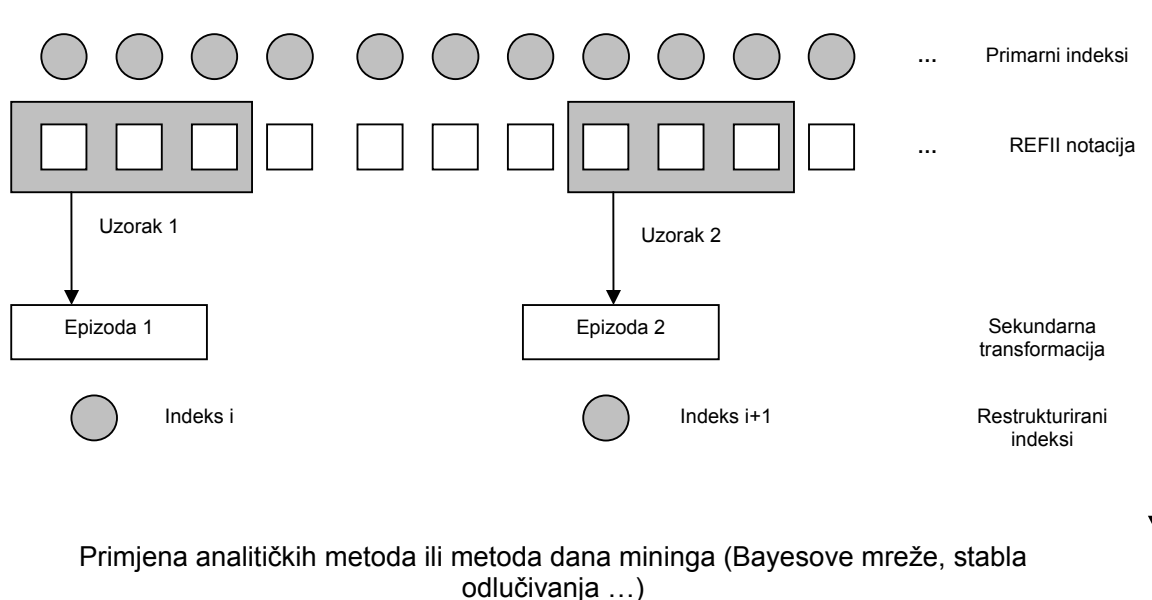
Primjer takve analize može biti zavisnost trenda kretanja pojave s obzirom na vremenski indeks, što se može realizirati primjenom asocijativnih algoritama na vrijednostima proizašlima iz vremenske serije i vremenskih indeksa svih hijerarhijskih razina.

Restrukturiranje vremenskih indeksa provodi se prilikom sekundarne transformacije nad REFII vrijednostima proizašlima iz REFII modela.

Prilikom sekundarne transformacije restrukturiranje indeksa se provodi s obzirom na karakter same analize.

To restrukturiranje može biti nadgradnja na postojeću strukturu indeksa te se nastoji uspostaviti logička veza između novostrukturiranih indeksa i indeksa proizašlih i primarne transformacije.

Restrukturirani indeksi proizašli iz sekundarne transformacije ne moraju nužno pratiti klasičnu vremensku komponentu (dan, tjedan, mjesec...), već to može biti i redoslijedna komponenta kao što je to prikazano na slici 5.4.



SLIKA 5.4. KREIRANJE RESTRUKTURIRANIH INDEKSA

Iz slike 5.4. vidljivo je da je kreiranje restrukturiranih indeksa u vezi sa sekundarnom transformacijom vrijednosti REFII modela. Prilikom procesa analize, ovisno o analitičkom cilju u model se mogu uključiti kako primarni, tako i restrukturirani indeksi istovremeno.

Naravno, moguće je u prilikom procesa analize pažnju usredotočiti samo na restrukturirane indekse i elemente sekundarne transformacije.

Primjer za to je analiza epizoda posredstvom Bayesovih mreža, gdje za izgradnju modela mogu biti dostatni elementi sekundarne transformacije i transformirani indeksi.

5.2. Integracija metoda *data mininga* u vremenske serije

Trenutna znanstvena aktivnost prilikom integracija metoda *data mininga* i vremenskih serija, svodi se na rješavanje specifičnih problema, kao što je to spomenuto u 1. poglavlju.

Pri tome se model transformacije vremenske serije izgrađuje s ciljem rješavanja specifičnog problema i korištenja jedne ciljane metode *data mininga* [Lin, 2002], [Keogh, 1998], [Keogh, 1997].

Trenutni pristup korištenja metoda *data mininga* u vremenskih serijama svodi se prvenstveno na primjenu klasteriranja i uvjetnih vjerojatnosti. Temeljni problem ovakvih rješenja koji je spomenut u 1. poglavlju svodi se na međusobnu nekompatibilnost, i rješavanje uskog specijaliziranog problema, pri čemu je vremenska serija transformirana s ciljem primjene specifične metode *data mininga* kako bi se postigao cilj poput grupiranja (klasteriranje).

REFII model nudi otvorenost prema metodama *data mininga*, pri čemu nije bitno unaprijed znati analitički cilj, već se kombiniranjem i kontingencijskim pristupom s obzirom na krajnji cilj dolazi do željenog rješenja.

Naredna tablica prikazuje neke od metoda *data mininga* i mogućnosti analize s obzirom na raspoložive elemente REFII modela.

Tablica 5.1. Metode rudarenja podataka s obzirom na mogućnost primjene REFII modela

	REFII				
	Primarna transformacija			Sekundarna transformacija	
	Primarni indeksi	Površina ispod odsječka (vrijednost/razred)	Koeficijent kutnog otklona (vrijednost/razred)	Restrukturirani indeks	Elementi sekundarne transformacije (epizode, motivi, nizovi sličnosti)
Standardna statistika	*	*	*	*	*
Asocijativni algoritmi	*	*	*	*	*
Klasteriranje		*	*		*
Link analiza		*	*		*
Bayesove mreže	*	*	*	*	*
Stabla odlučivanja	*	*	*	*	*
SOM		*	*	*	*

U prikazanoj tablici polja označena sa "*" predstavljaju poveznicu između metode *data mininga* i elementa REFII modela.

Tako se primjerice standardne statističke metode mogu primijeniti kako na koeficijentima kutnog otklona, površini, tako i na intervalnim vrijednostima modela.

Primjer integracije asocijativnih algoritama i REFII modela prikazan je u poglavlju 4.4., gdje je prikazan način izravnog otkrivanja pravila iz vremenskih serija.

Klasteriranje se može primijeniti nad koeficijentima kutnog otklona, površinama ispod krivulje, što je i prikazano u poglavlju 6.7.4., gdje je primijenjeno k-mean klasteriranje nad koeficijentima kutnog otklona.

Isto tako, u poglavlju 6.7.3. prikazana je metodologija primjene stabla odlučivanja nad vremenskom serijom transformiranom u REFII model, gdje se je proveo postupak vremenske ekspanzije. SOM, su sastavni dio prikazanih algoritama za otkrivanje frekventnih uzoraka (poglavlje 6.1.), dok je primjena Bayesovih mreža u kombinaciji sa REFII modela prikazana u poglavlju 7.3. koje se referira na primjenu REFII modela u medicini.

Razlika između dosadašnjeg pristupa i pristupa koji nudi REFII model odnosi se na mogućnost izbora *data mining* metode s obzirom na željeni cilj analize, pri čemu se nastoji podatke temporalnog karaktera izjednačiti sa podacima netemporalnog karaktera po pitanju mogućnosti korištenja metoda *data mininga*.

Ako na primjer želimo segmentirati klijente koji su odgovorili na našu ponudu, i one koji to nisu s obzirom na dob, spol, teritorij i slične atribute, tada ne postoje nikakva prepreka primjene stabla odlučivanja kako bismo segmentirali populaciju.

Ako želimo znati koji su glavni faktori koji utječu na formiranje određene otkrivene epizode u vremenskoj seriji, te želimo u model uključiti kako trendove nakon i trendove prije odvijanja spomenute epizode, uzimajući u obzir i netemporalne podatke poput primjerice doba, spola, regije i sličnih karakteristika subjekata koji su sudjelovali u

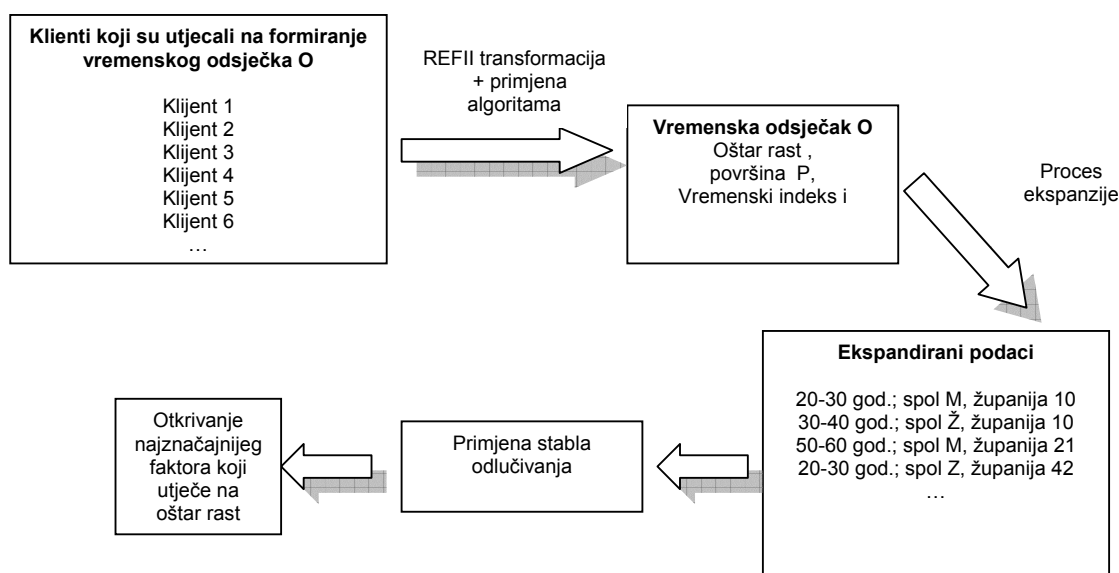
formiranju otkrivene epizode, tada bi klasičnim putem morali formirati poseban specijalistički model za rješavanja ovog problema.

U slučaju korištenja REFII modela, nakon ekstrakcije epizoda u vremenskoj seriji, da bismo spoznali glavne faktore utjecaja dovoljno je primijeniti stablo odlučivanja na elemente REFII modela.

Ono što je bitno napomenuti u slučaju korištenja stabla odlučivanja nad vremenskom serijom i integracijom netemporalnih elemenata u model, potrebno je izvršiti ekspanziju vremenskog odsječka odnosno epizode.

Ekspanziju vremenskog odsječka možemo definirati kao pridruživanje originalnih atributskih vrijednosti svakom pojedinačnom slučaju koji je utjecao na trend sa pridruživanjem originalnih vrijednosti trenda.

Ilustracija ekspanzije prikazana je na slici 5.5.



SLIKA 5.5. EKSPANZIJA VREMENSKOG ODSJEČKA NETEMPORALNIM ATRIBUTIMA

Kao što je to slučaj sa prikazanim primjerom stabla odlučivanja REFII model teži ka fleksibilnosti i rješavanju problema prema okvirnom predlošku prikazanom na slici 5.5. bez potrebe kreiranja posebnih rješenja da bi se na vremensku seriju mogle primijeniti *data mining* metode.

Cilj je imati otvoren sustav koji je sposoban apsorbirati kako specifične algoritme primjenjive na REFII model s ciljem otkrivanja zakonitosti u vremenskim serijama, ali isto tako se teži ka otvorenosti prema klasičnim *data mining* algoritmima.

Na sličan način kao što je primijenjena metoda stabla odlučivanja moguće je primijeniti i metodu Bayesovih mreža i to na elemente primarne i sekundarne transformacije vremenske serije, gdje ovisno o potrebama i cilju analize u analitički proces mogu ući kako razredi vrijednosti, tako i indeksi nastali u procesima primarne i sekundarne transformacije.

Ostale metode i njihova mogućnost primjene na vremenskim serijama vidljive su u prikazanoj tablici.

Teško je nabrojati sve mogućnosti integracije i sinergijskih efekata između modela *data mininga* i REFI modela. Oni kao što je već spomenuto ovise o tijeku analize i problemskom prostoru kojeg treba riješiti, a ovaj pristup omogućava postizanje sinergijskog efekta na temelju sagrađenih mostova između dviju koncepcija koje kao krajnju mogućnost nude povezivanje vremenskih serija sa *data mining* metodama bez potrebe stvaranja zasebnih specijalističkih i međusobno nekompatibilnih modela.

5.3. Metodologija integracije različitih koncepcija analize

Karakteristike znanstvenih radova koje se bave problematikom vremenskih serija su kao što je već mnogo puta spomenuto međusobna nekompatibilnost i usredotočivanje na uski segment problemskog prostora.

Rezultat takvog pristupa očituje se i u činjenici da i pojedini autori koji se bave sa dva različita problemska prostora u domeni vremenskih serija ne daju rješenje za moguće povezivanje vlastitih modela, jer su previše usredotočeni na rješavanje određenog problema.

Očiti primjer takvog pristupa je rješavanje pronalaženja uzoraka unutar vremenskih serija [Han, 2003] i rješavanje problema otkrivanja dugih uzoraka u šumovitom prostoru (dugi uzorci s mutacijama) [Han, 2002].

Ovaj problem je još izraženiji ako pokušavamo naći metodološke veze za iste tipove problema pri čemu su inače efikasna rješenja predložena od različitih autora [Wang, 2001], [Brazma, 1998].

Ako pak želimo iskoristiti istraživanja bazirana na probabilističkim pristupima, [Chiu, 2003], [Keogh, 1997], otkrivanju uzoraka [Brazma, 1998], [Geurts, 2001], [Han, 2003], [Lam, 1998], otkrivanju mutacija u dugim uzorcima [Wang, 2001], [Han, 2002], otkrivanju epizoda [Manilla, 1997], otkrivanja sličnosti unutar vremenskih serija [Caraça-Valente, 2000], [Das, 1997], [Faloutsos, 1997], [Lee, 2000], [Manilla, 2001], [Park, 1999], [Popivanov, 2002], [Rafiei, 1998], [Yi, 1998], traženje motiva unutar vremenskih serija [Chiu, 2003], [Lin, 2001], [Patel, 2002], [Tanaka, 2003], kako bismo izgradili jedinstven sustav analize vremenskih sustava, tada je gotovo nemoguće zbog nepostojanja međusobne podatkovne poveznice (jedinstvenog modela transformacije).

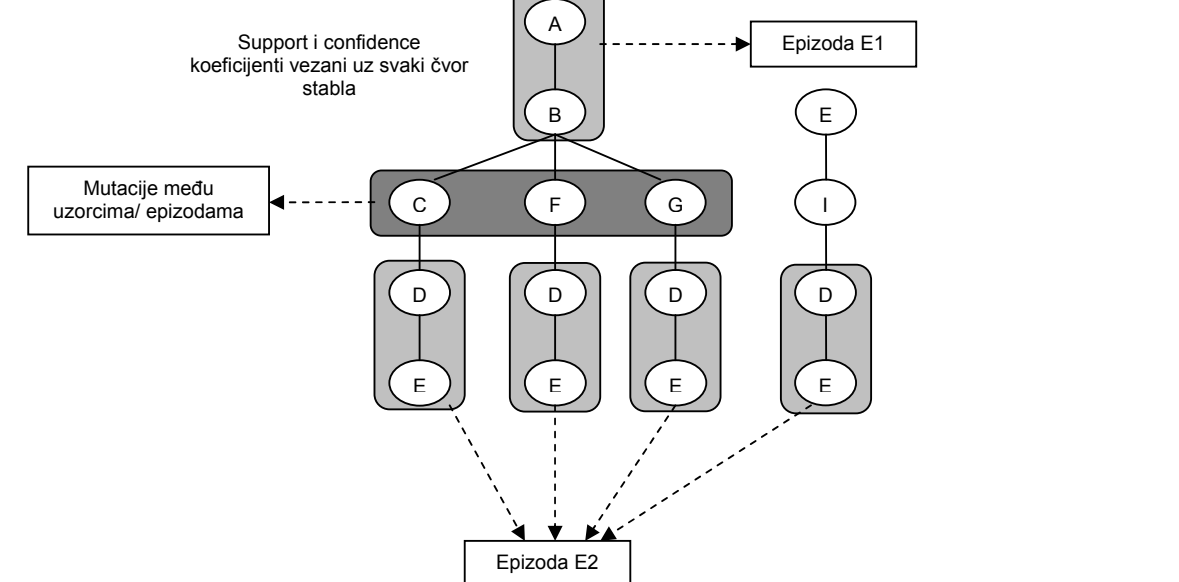
Važnost sinteze metodoloških postupaka proizlazi iz potreba kompleksnijih analiza. Ako na primjer otkrijemo uzorke unutar vremenskih serija, te želimo znati da li među uzorcima postoji vezni element koji je sklon mutacijama, te želimo znati probabilističke odnose između uzorka koje želimo promatrati kao epizode, tada bi trebalo integrirati rezultate istraživanja iz niza znanstvenih radova, gdje postoji usredotočenost na otkrivanje uzoraka, analizu epizoda, probabilističke odnose među odsječcima i rezultate istraživanja otkrivanja dugih uzoraka u šumovitom prostoru.

Kako među tim znanstvenim radovima ne postoji poveznica u smislu mogućnosti jednostavne integracije metoda tako zadan problem zahtijevao bi generiranje novog pristupa analizi koji rješava upravo jedan ovako kompleksan model, i tako redom.

Za svaki kompleksi zahtjev iz domene analize vremenskih serija bilo bi potrebno izgrađivati kompleksan model rješenja.

REFI model nastoji optimizirati procese u rješavanju kompleksnih problema u domeni analize vremenskih serija, te kao odgovor na pitanje kako integrirati niz različitih koncepata analize koji mogu biti korisni u rješavanju kompleksnih tipova analize nudi

Kada bi se spomenuti problem otkrivanja uzoraka unutar vremenskih serija, te spoznavanja postojanja veznog elementa koji je sklon mutacijama, sa probabilističkim odnosima između uzorka koje želimo promatrati kao epizode tada je taj problem elegantno riješiv primjenom REFII modela.



Slika 5.6. daje rješenje u okviru REFII modela. Nakon generiranja stabla frekventnih uzoraka, epizode možemo definirati preko uzoraka, a međuelementi (C,F,G) između uzoraka (epizode E1 i epizode E2) možemo promatrati kao mutacije, odnosno uzorke u šumovitom prostoru (eng. noisy environment).

Tako primjerice možemo računati uvjetne vjerojatnosti pojavnosti mutacija (C,F,G) nakon odvijanja epizode E1 na slijedeći način :

$$P(E_1|F) = p_F$$

$$P(E_1|G) = p_G$$

Pri čemu se mutacije C, F i G mogu odnositi na trendove (nizak rast, visok rast i slično), ili pak na događaje (period niskog rasta, sastavljen od niza trendova dužine n , koje karakterizira nizak rast). Epizodu E1, također možemo definirati na isti način.

Ako unutar iste analize želimo otkrivati motive u vremenskim serijama, tada je po uzoru na poglavlje 4.7.2. potrebno izvršiti sekundarnu transformaciju i na osnovu nje po matrici transformacije REFII modela izvršiti pretragu motiva.

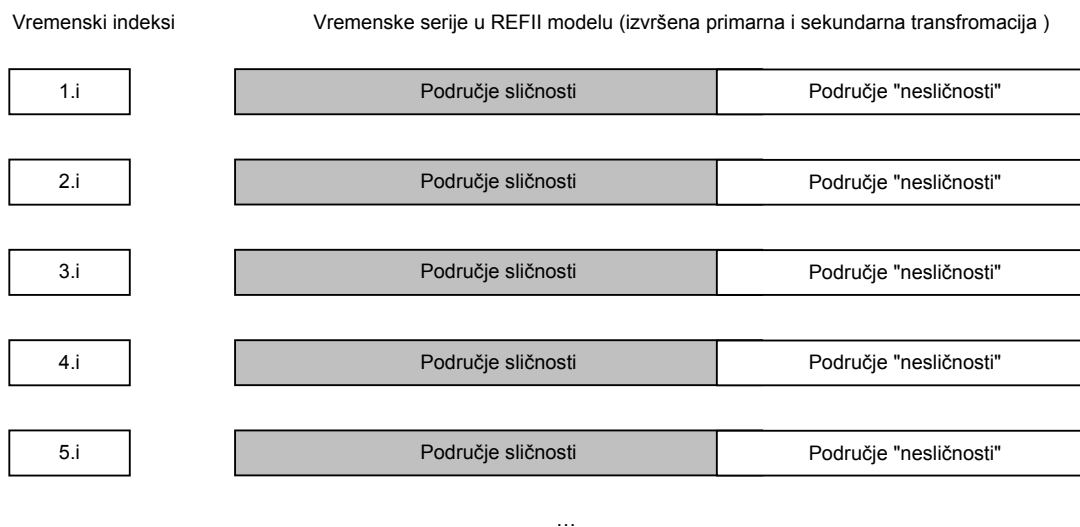
Isto tako kao što je to prikazano u poglavlju 4.4.2, moguće je izravno tražiti pravila na vremenskoj seriji u okviru iste analize, što je ujedno i primjer integracije *data mining* analiza unutar vremenskih serija.

Iz ovog je primjera vidljiva fleksibilnost i mogućnost integracije različitih koncepcija analize vremenskih analiza u okviru istog sustava, bez osmišljavanja pojedinačnih analitičkih modela specijaliziranih za točno određenu vrstu kompleksne analize.

Daljnja značajna stvar u okviru kompleksnih analiza posredstvom REFII modela je mogućnost izbora različitih načina za postizanje istih analitičkih ciljeva.

Sezonske oscilacije u okviru REFII modela možemo otkrivati na temelju algoritma prikazanog u poglavlju 4.2.2. Alternativni pristup otkrivanja sezonskih oscilacija moguće je provesti temeljem definicije sezonskih oscilacija preko definicije sličnosti prikazane u poglavlju 4.7.2.

Postojanje sezonske oscilacije preko sličnosti možemo definirati kao postojane sličnosti s obzirom na unaprijed definirane pozicije s obzirom na vremenske indekse, kao što je to prikazano na slici 5.7.



SLIKA 5.7. DEFINICIJA SEZONSKIH OSCILACIJA PREKO SLIČNOSTI

Iz slike 5.7. vidljiva je osnovna ideja definicije sezonskih oscilacija preko definicije sličnosti. Na osnovu ove koncepcije ako postoji sličnost na korespondentnim vremenskim odsječcima definiranim kroz vremenske indekse, tada postoji i sezonska oscilacija. Pod korespondentnim vremenskim odsječcima definiranim kroz vremenske odsječke podrazumijevaju se dani, tjedni, mjeseci, kvartali...

Kao i u metodološkom rješenju prikazanom u poglavlju 4.2.2. , i ovo rješenje nudi mogućnost definicije koeficijenata pouzdanosti.

5.4. Konstrukcija nestandardnih algoritamskih rješenja unutar REFII modela

S obzirom na trenutnu situaciju u domeni analiza vremenskih serija, gdje ne postoji ujednačen metodološki koncept analize, već se analize svode na primjenu niza različitih koncepcija i pristupa, gdje je svaki analitički postupak zatvorena cjelina gotovo je teško ili nemoguće integrirati nestandardne modele rješenja bez postojanja jedinstvenog modela transformacije vremenske serije.

REFII model konstruiran je upravo s ciljem brze prilagodbe i mogućnosti prilagodbe za razvoj nestandardnih algoritamskih rješenja u domeni analiza vremenskih serija.

Integracija različitih koncepcija analize u analitički sustav, može se također promatrati kao konstrukcija nestandardnih algoritamskih rješenja analize.

Snaga ovakvog pristupa dolazi do izražaja kod rješavanja problema poput povezivanja vremenskih serija sa stablima odlučivanja koje je prikazano u poglavlju 5.2, gdje je izvršena ekspanzija vremenskog odsječka netemporalnim atributima za potrebe analize posredstvom stabla odlučivanja.

Ako je riječ o povezivanju vremenskih serija sa Bayesovim mrežama, ili algoritmima za klasteriranje tada su postupci povezivanja bitno drugačiji.

Ilustrativan primjer fleksibilne podatkovne strukture uzoraka je stablo frekventnih uzoraka, koje može poslužiti kao polazišna osnova za primjenu i razvoj različitih vrsta algoritama.

Tako se primjerice na stablo frekventnih uzoraka mogu primijeniti algoritmi za pretraživanje poput *A**, *Breadth first*, ili *Deep first* [Chabris,1987] kako bi se otkrila potencijalna povezanost vremenskih odsječaka ili epizoda.

Daljnja strategija korištenja ovakve podatkovne strukture može se svesti na otkrivanje mutacija na dugim otkrivenim uzorcima kao što je to vidljivo iz poglavlja 5.3. primjenom osnovnih koncepcija uvjetne vjerojatnosti i epizoda.

Jedna od mogućih strategija analize je reprezentacija vremenskih segmenata u obliku nizova sličnosti kao elemenata strukture stabla.

Iz svega nabrojanog, vidljivo je da REFII pruža solidnu osnovu za izgradnju novih analitičkih rješenja. U domeni tržišnih i *data mining* analiza, gdje je vrlo bitna mogućnost integracije i ulančavanja različitih koncepcija analize, REFII omogućuje ostvarivanje fleksibilnosti i kreiranja novih rješenja, kao što je to slučaj i sa atributima netemporalnog karaktera.

Osim toga ova koncepcija otvara mogućnost promatranja određenih podatkovnih struktura poput tekstualnih datoteka ili "web log" datoteka kroz prizmu vremenskih serija s ciljem provođenja daljnjih analitičkih postupaka nad njima.

Ideja je postojeće podatkovne strukture prikazati u REFII notaciji, jer se one mogu promatrati u korelaciji sa vremenskom komponentom, te nakon toga primijeniti raspoložive algoritamske postupke na matricu transformacije s ciljem otkrivanja zakonitosti. Tekstualne datoteke, kao i web logove moguće je promatrati kao kontinuirani niz vrijednosti ovisan o vremenskoj komponenti. Ciljevi analize ovakvog pristupa svode se na otkrivanje rečeničnih uzoraka u tekstu, odnosno pronalaženje

sličnosti u tekstovima, odnosno na analizu “nizova klikova” (eng. Click stream) u slučaju web logova.

Na temelju analize nizova klikova moguće je segmentirati posjetitelje na temelju sličnosti (vidi poglavlje 4.7.), procjenjivati ergonomiju Web stranica, ili pak generirati vjerojatnost naredne aktivnosti posredstvom uvjetne vjerojatnosti, ili direktnog otkrivanja pravila iz vremenske serije (vidi poglavlje 4.3. i 4.4.).

Moguće analize vezane uz tekstualne datoteke odnose se na razvoj algoritama za prepoznavanje sličnosti prema kontekstu, kao i prema rečeničnim uzorcima.

5.5.Određivanje stupnja pouzdanosti analize

Karakteristika REFII modela očituje se u mogućnosti definicije stupnja pouzdanosti analize od strane samog analitičara. Mehanizmi koji analitičaru stoje na raspolaganju prilikom određivanja stupnja pouzdanosti svode se uglavnom na manipulaciju veličinom razreda kutnih odklona i površine ispod krivulje, kao i definicije određivanja stupnja pouzdanosti analize (vidi poglavlje 4.2.2. i 4.4.2).

Sam postupak određivanja stupnja pouzdanosti može se odvijati na način da analitičar isprva zada vrlo visoku toleranciju prilikom provođenja analitičkog postupka, te nakon toga može povećavati kriterije za pouzdanost i to prvenstveno na segmentima gdje je otkrivena pravilnost uz niže zahtijevani stupanj pouzdanosti.

REFII model pruža slobodu po pitanju deklariranja stupnja pouzdanosti, jer analitički ciljevi diktiraju stupanj očekivane preciznosti.

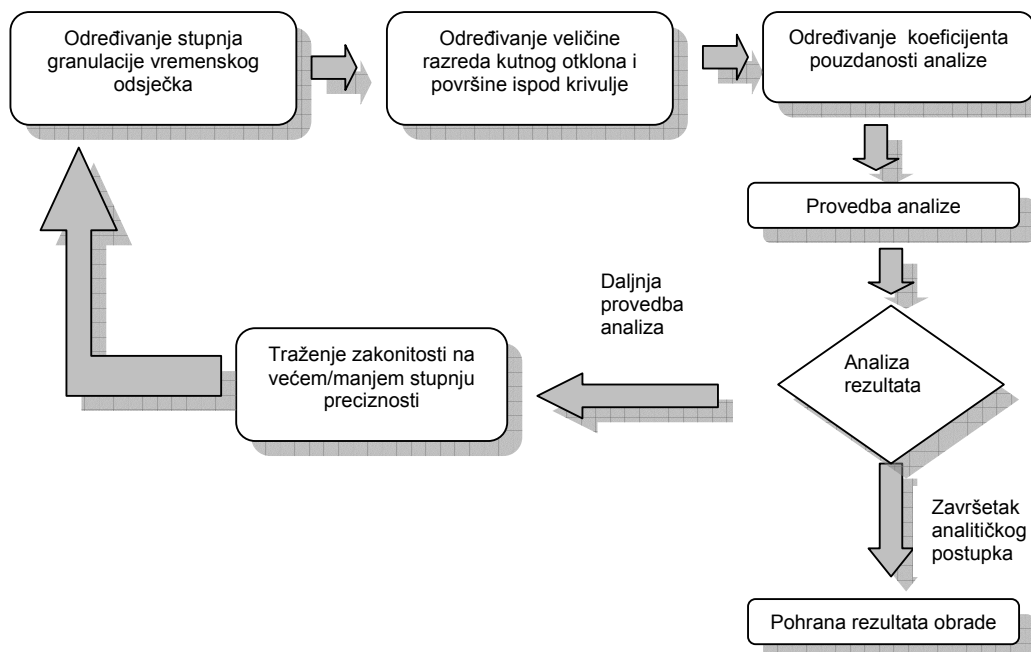
U situacijama kada želimo znati da li je očekivani trend neke pojave padajući ili rastući, tada u kategoriju rastućeg trenda spada i slab rast, kao i oštar rast. Na isti način se može promatrati i postojanje sezonske oscilacije, kao i postojanje sličnosti.

Druga je krajnost insistiranje na krajnjoj preciznosti, gdje su odkloni kutnih koeficijenata svrstani u grupe od primjerice 0.001.

Ovi mehanizmi direktno utječu na procjene intervalne jednakosti, što je osnovi mehanizam u procjeni kategorizacije po pitanju pripadnosti grupe.

Koeficijenti pouzdanosti kao faktori određivanja stupnja pouzdanosti, drugi su puno izravniji element o kome ovisi tijek analize. Kao što je već prikazano u poglavljima 4.2.2. i 4.4.2, gdje su se za potrebe analize sezonskih oscilacija i direktnog otkrivanja pravila koristili faktori pouzdanosti, ista metodologija primjenjiva je i u analitičkim postupcima poput otkrivanja uzoraka unutar vremenskih serija, te sličnosti unutar vremenskih serija.

Spiralni proces određivanja željenog stupnja analize prikazan je na slici 5.8.



SLIKA 5.8. SPIRALNI PRISTUP U ODREĐIVANJU ŽELJENOG STUPNJA POUZDANOSTI

Slika 5.8. sadrži i granulaciju kao instrumentarij za određivanje stupnja pouzdanosti. U temeljnom modelu, fokus je bio usmjeren na veličinu razreda kutnih otklona i površine ispod krivulje, kao i definicije određivanja stupnja pouzdanosti analize. Stupanj granulacije vremenskog odsječka više spada u kategoriju uvjetovanu ciljem analize, ali se u širem smislu može promatrati i kao instrument za određivanje stupnja pouzdanosti.

Kada je samim ciljem analize uvjetovana granulacija, (npr. otkrivanje mjesečnih oscilacija za neku pojavu), tada se taj element zanemaruje kao instrument za povećanje/smanjivanje stupnja pouzdanosti.

Gledajući sa šireg aspekta, stupanj granulacije nije instrument za regulaciju stupnja pouzdanosti, no on se može promatrati i na taj način ovisno o karakteru i cilju same provedbe analize.

Slučaj u kome se stupanj granulacije vremenskih odsječaka može promatrati kao instrument određivanja stupnja pouzdanosti odnosi se primjerice na otkrivenu sezonsku oscilaciju na razini mjeseca. Ako otkrijemo dane u mjesecu koji imaju najveći utjecaj na formiranje generalnog trenda vezanog uz mjesečnu oscilaciju, tada možemo govoriti o ulozi granulacije kao instrumenta za regulaciju stupnja pouzdanosti analize.

Prikazane instrumente za određivanje stupnja pouzdanosti, aktivira i koristi analitičar, ovisno o problemskom prostoru iz kojeg potiče problem. Bitno je shvatiti interakciju između dva osnovna instrumenta za utvrđivanje stupnja pouzdanosti te interakciju opcionalnog uključivanja granulacije kao trećeg elementa kod modeliranja i dizajniranja rješenja koje zahtijeva određeni stupanj pouzdanosti analize.

5.6. Provođenje analiza

REFII model prvenstveno je zamišljen kao sustav koji analitičaru omogućava kontingencijski pristup u analizi vremenskih serija. To znači da analitičar pristupajući analizi vremenske serije može, ali i ne mora unaprijed imati definiran analitički zadatak. Za razliku od tradicionalnog pristupa analizi vremenskih serija, gdje se unaprijed zadaje analitički cilj koji je ujedno i jedini cilj analize, prilikom primjene REFII modela može se također krenuti u realizaciju unaprijed zadanog analitičkog zadatka, te se proces analize dalje razvija u ovisnosti o prethodno dobivenim rezultatima analize (vidi poglavlja 5.2.i 5.3). omogućavajući pri tome i primjenu metoda *data mininga* na vremenske serije, kao što je to prikazano u poglavlju 5.4..

Provođenje analiza možemo promatrati sa dva osnovna aspekta :

- Kao *ad hoc* provođenje analiza s unaprijed definiranim analitičkim ciljem što je bliže tradicionalnom konceptu provođenja *data mining* analiza nad vremenskim serijama
- Sustavno provođenje analiza s ciljem otkrivanja zakonitosti na populaciji vremenskih odsječaka bez unaprijed definiranog, ili vrlo fleksibilno definiranog analitičkog cilja

U prvom slučaju analize se provode s ciljem otkrivanja primjerice sezonskih oscilacija, motiva, ili sličnosti u vremenskim serijama, gdje analiza mora dati nedvosmislen odgovor na postavljeni zahtjev.

Iako je ovaj pristup u osnovi bliže ideji bliže tradicionalnog koncepta provođenja *data mining* analiza nad vremenskim serijama, ne smijemo ga stavljati u istu kategoriju kada je riječ o provođenju analiza vremenskih serija posredstvom REFII modela.

Razlog tome leži u činjenici, što ovakav pristup usprkos tradicionalno zadanom cilju analize možemo provesti koristeći sve prednosti REFII modela.

Ta prednost se očituje u većoj količini informacija koju je analiza potpomognuta REFII modelom u stanju pružiti analitičaru.

Tu prednost je moguće ilustrirati situacijama gdje je osnovni cilj otkriti određene motive u vremenskoj seriji.

Tradicionalni pristupi koji se bave pronalaženju motiva u vremenskim serijama, daju informaciju o postojanju/ nepostojanju traženog motiva te njegovoj frekvenciji pojavnosti [Lin, 2001], [Patel, 2002], za primjerice ekstrahirana dodatne informacije o uzročno posljedičnim vezama traženih motiva, potrebno je izvršiti novu analizu [Chiu, 2003], te pokušati pronaći mogućnost spajanja koncepcija u okviru navedenih znanstvenih radova.

Ako pak želimo otkriti da li postoji šira sličnost od one zadane motivima za primjerice $t+1$; $t-1$ tada je potrebno uključiti i algoritam za otkrivanje sličnosti, i to njegovu posebnu izvedenicu s obzirom na zadane motive. Slično je i sa pokušajem otkrivanja uzročno – posljedičnih veza u motivima s obzirom na sezonski faktor.

Dakle tradicionalni pristup u analizi vremenskih serija je poprilično demotivirajući i kompliciran kada želimo saznati nešto više u odnosu na zadani analitički cilj.

Sa perspektive REFII modela, nakon otkrivanja motiva, na temelju istog modela transformacije vremenske serije i definicije motiva kao epizoda, moguće je tražiti uzročno posljedične veze između motiva posredstvom uvjetnih vjerojatnosti primijenjene na epizode. Za otkrivanje postojanja šire sličnosti od one zadane motivima za dužinu u odnosu na motiv $t+1$; $t-1$ potrebno je kreirati i uspoređivati stringove sličnosti koji sadrže ,motive plus dodatne komponente. Otkrivanja uzročno –

posljedičnih veza u motivima s obzirom na sezonski faktor .moguće je provesti na temelju pozicije motiva u vremenskoj seriji s obzirom na pripadajuće vremenske indekse.

REFII modela omogućava sustavno provođenje analiza s ciljem otkrivanja zakonitosti bez unaprijed definiranog, ili vrlo fleksibilno definiranog analitičkog cilja.

Ako primjerice imamo vremenske serije iz domene osiguranja koje se odnose na provale u prostorije osiguranih poduzeća, tada možemo analizirati vremensku seriju s ciljem otkrivanja potencijalnih zakonitosti, bez jasno definiranog cilja analize. Gledano sa perspektive tradicionalnog pristupa, analitički zadaci nad ovakvom vremenskom serijom mogle bi se svasti na primjerice otkrivanje sezonskih oscilacija, analizom utjecaja na trendove s obzirom na regije i slično.

Svaka od tih analiza provodila bi se na tradicionalni način bez mogućnosti kontingencijskog pristupa.

Kontingencijski pristup, koji omogućava REFII model, omogućava provođenje daljnjih analitičkih aktivnosti s obzirom na prethodno otkrivene pravilnosti.

Ako na spomenutoj vremenskoj seriji provala u osigurana poduzeća otkrijemo primjerice vrlo frekventne uzorke, tada te uzorke možemo promatrati s obzirom na sezonsku komponentu. Otkrivene uzorke možemo ekspanirati nevremenskim atributima kao što je prikazano u poglavlju 5.2. koji se mogu odnositi primjerice na djelatnost poduzeća, broj zaposlenih i slično.

Nadalje moguće je promatrati sličnost pojedinih vremenskih perioda i klasterirati vremenske periode provala prema sličnosti.

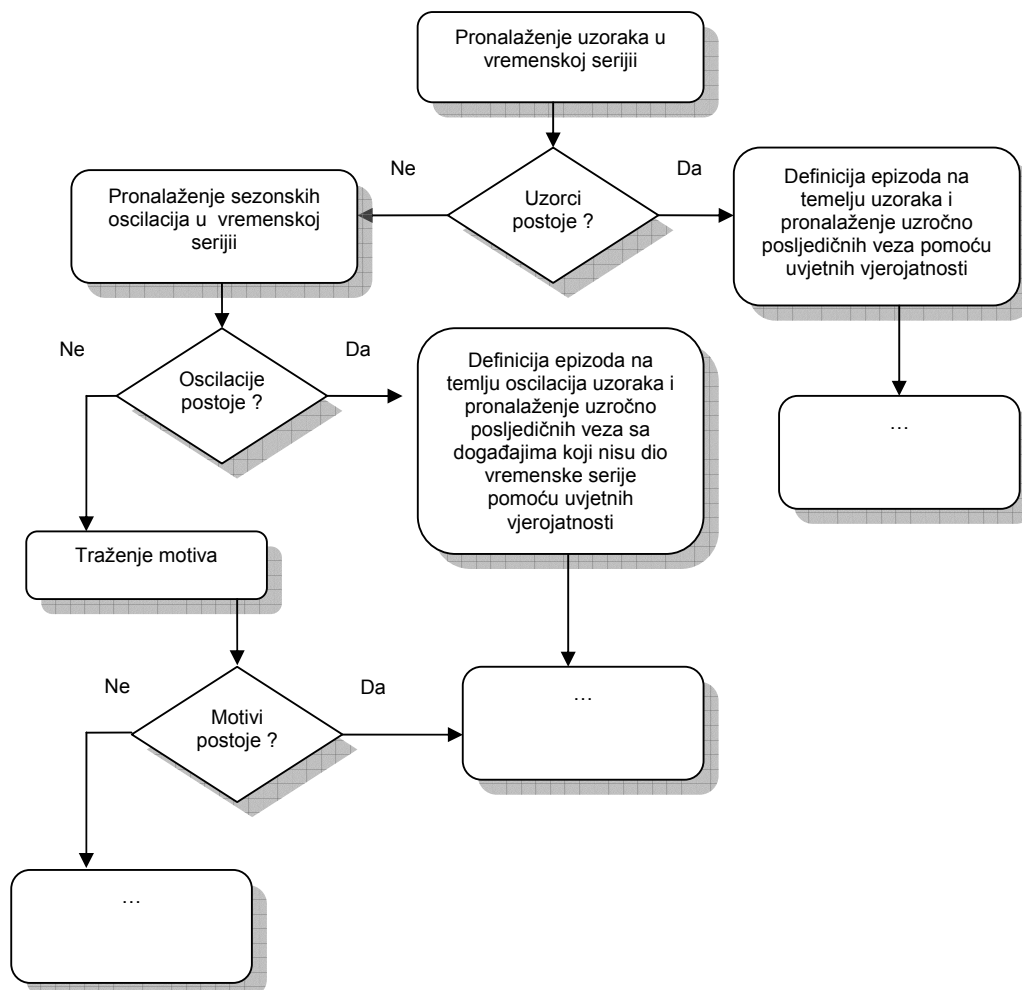
Alternativni pristup analizi mogao bi se svesti na deklariranje pronađenih vremenskih uzoraka epizodama, te tražiti uzročno posljedične veze s obzirom na događaje koji nužno ne moraju biti elementi promatrane vremenske serije (državni praznici, vjerski praznici i slično.)

Kao što je iz navedenog primjera vidljivo, za tražene zakonitosti u vremenskim serijama puno je pogodniji REFII model, koji je puno fleksibilniji u provođenju *data mining* analiza vremenskih serija.

Kod projekata poput otkrivanja pravilnosti u genetičkim lancima, vremenskim serijama upotrebe bankomata, vremenskim serijama provođenja gotovinskih transakcija na bankovnim šalterima, otkrivanju zakonitosti u vremenskim serijama o prometnim nezgodama, kontingencijski pristup analizi vremenskih serija može biti vrlo efikasan, te može rezultirati otkrivanjem neočekivanih skrivenih zakonitosti.

Za razliku od tradicionalnog pristupa gdje se uglavnom testiraju hipoteze o postojanju različitih tipova pravilnosti u vremenskim serijama, sustavno provođenje analiza nad vremenskim serijama samo usmjerava daljnje tijekove analize poput putokaza koje vode ka cilju.

Metodologija sustavske analize pomoću REFII modela vidljiva je iz slike 5.9.



SLIKA 5.9. METODOLOGIJA SUSTAVNOG PRISTUPA ANALIZE VREMENSKIH SERIJA POMOĆU REFII MODELA

Slika 5.9. prikazuje svu kompleksnost sustavnog pristupa analize vremenskih serija posredstvom REFII modela, koja je sposobna ulančavati metodološke postupke analize vremenskih serija bez obzira da li je riječ o tradicionalnoj analizi vremenskih serija, ili je riječ o novim algoritmima proizašlih na temelju REFII modela, ili je pak riječ o tradicionalnim *data mining* metodama koje se pomoću REFII modela mogu primijeniti na vremenske serije.

Osnovna prednost ovakvog pristupa proizlazi iz činjenice da analitičar nije pasivni sudionik procesa analize, već njen glavni moderator. Hipotetski dijagram tijekom analize prikazan na slici 5.9, moguće je grditi i tijekom analize, a ne prije nje.

Naravno, moguće je i unaprijed definirati dijagram tijeka analize, ali se analitički postupci mogu određivati i s obzirom na dobivene rezultate u prethodnom koraku.

Jedno od bitnih pitanja u provođenju sustavnom pristupu je pitanja inicijalne analize, odnosno kojim tipom analize započeti analitički proces, sa ciljem kasnijeg razgranavanja analitičkih postupaka.

Moguće je započeti sa analizom kako kod tradicionalnog pristupa, gdje je unaprijed zadan cilj, pa u ovisnosti o rezultatima krenuti u drugu etapu analize, ili pak inicijalno izvršiti nekoliko intuitivno najpodesnijih analiza, pa rezultate najuspješnije uzeti kao putokaz za nastavak tijeka analitičkih procesa.

Ovdje postoji čitav niz mogućnosti, koje velikim dijelom ovise i o problemskom prostoru koji se obrađuje.

5.7. Prikaz i primjena rezultata

Rezultati proizašli temeljem analiza vremenskih serija, osim što mogu poslužiti kao parametri u ulančanim procesima analize, mogu se skladištiti kao formalizirano znanje u okviru ekspertnih sustava.

Najjednostavniji oblik informacije koji može proizaći na temelju analize vremenskih serija je informacija koja služi kao ulazni parametar u procesu odlučivanja. Takva je primjerice informacija o tome da tržišni segment A i tržišni segment B imaju isti model ponašanja. Ova tvrdnja daje određenu informaciju koja se može iskoristiti u procesima donošenja odluka.

U okviru složenih inteligentnih sustava, niz pravila možemo povezati u sustav pravila. To možemo postići ili klasičnim ekspertnim sustavima, ili *fuzzy* ekspertnim sustavima.

Osnovna ideja primjene REFII modela je rušenje barijere koja područje vremenskih promatra kao zasebni svijet analize, ne pružajući dovoljno mogućnosti povezivanja i rezultata analiza koja nemaju svoje uporište u vremenskoj komponenti.

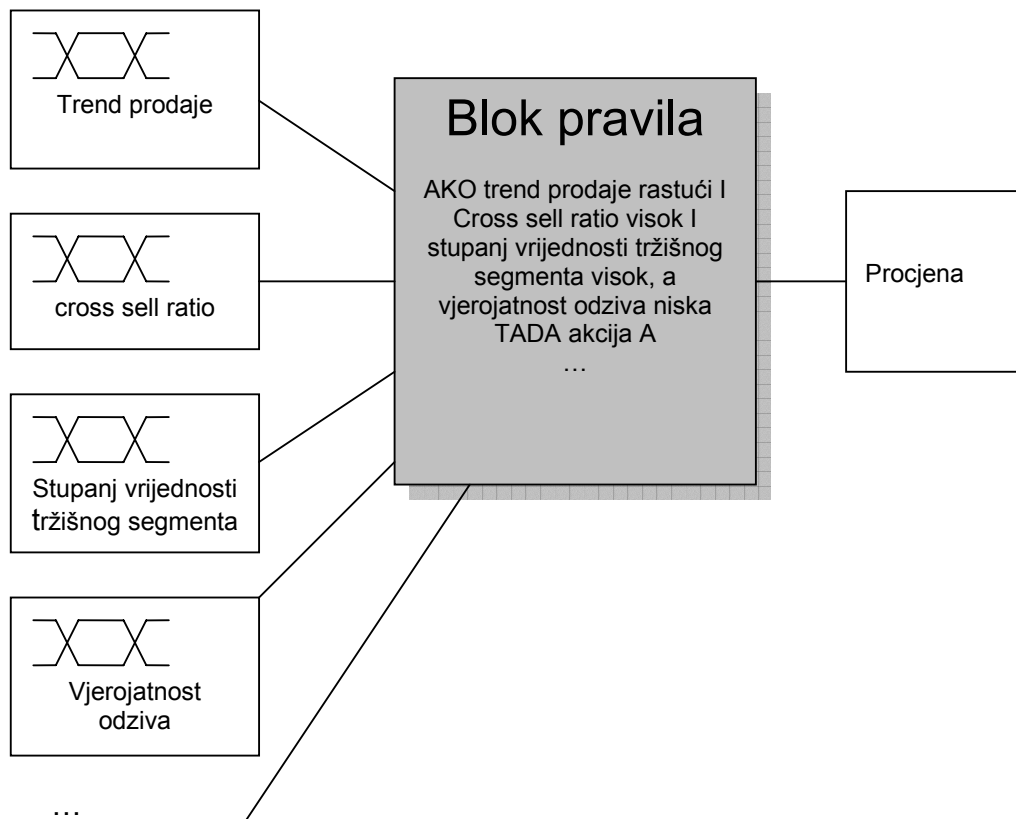
REFII model se za razliku od koncepcija koje favoriziraju vizualizaciju vremenskih serija, oslanjajući se na ljudske perceptivne mehanizme u otkrivanju znanja, oslanja na matematički instrumentarij koji na temelju strojnog učenja otkriva pravilnosti unutar vremenskih serija.

Vodeći se idejom o kreiranju jednog cjelovitog sustava za analizu vremenskih serija koji bi objedinio dosadašnju analitičku praksu u domeni vremenskih serija, te sva buduća istraživanja u domeni vremenskih serija, potrebno je i osmisliti koncepciju adekvatne pohrane takvih informacija u inteligentne sustave koje su u stanju procesirati informacije i nuditi solucije za određene tipove problema.

Taj bi sustav između ostalog morao objединiti pohranu znanja proizašlu na temelju analiza vremenskih serija i ostalih vrsta analiza, u vidu sustava pravila.

Na ovaj način moguće je vršiti preciznu segmentaciju tržišta, kombinirajući analitička znanja iz domene vremenskih serija, izvan domene vremenskih serija, te kombinirajući ekspertna znanja stručnjaka.

Ilustrativni dio takvog sustava prikazan je na slici 5.10., te predstavlja hipotetski model segmentacije baziran na *fuzzy* ekspertnom sustavu koji kombinira znanja iz različitih izvora s ciljem prijedloga konkretnih akcija za svakog klijenta pojedinačno.



SLIKA 5.10. PRIJEDLOG OBJEDINJAVANJA ZNANJA U FUZZY EKSPERTNOM SUSTAVU

Naravno, nisu samo *fuzzy* ekspertni sustavi iskoristivi za skladištenje, obradu znanja i predlaganje adekvatnih rješenja, tu se mogu iskoristiti i standardni ekspertni sustavi. Rezultati analiza proizašli na temelju primjene REFII modela, ne koriste se samo u strogo formaliziranom obliku. Opisani sustav ilustrativno prikazuje mogućnosti primjene rezultata analize u širem kontekstu od tradicionalnog poimanja rezultata analiza vremenskih serija.

Ovisno o problemskom prostoru prikaz i primjena rezultata proizašlih na temelju analiza vremenskih serija mogu varirati od pisanih analitičkih izvještaja, preko etapa pretprocesiranja podataka s ciljem iznalaženja temeljnih zakonitosti, pa sve do formiranja inteligentnih sustava koji sugeriraju skup rješenja za određenu promatranu problematiku.

Promatranje ove problematike u svjetlu inteligentnih sustava, trebalo bi zaokružiti teorijsku koncepciju na kojemu počiva REFII model kao jedinstveni model transformacije vremenske serije. Pomoću REFII modela vremenska serija ima svoju logističku potporu od trenutka temeljne transformacije, preko samih analitičkih procesa, kreiranja novih algoritamskih koncepcija, povezivanja sa tradicionalnim metodama *data mininga*, pa sve do izgradnje inteligentnih sustava koji u sebi sadrže integrirano znanje proizašlo na temelju analize pomoću REFII modela.

6. Otkrivanje tržišnih zakonitosti pomoću REF II modela na realnim podacima

6.1. Otkrivanje uzoraka iz vremenske serije

Otkrivanje uzoraka unutar vremenskih serija česta je tema znanstvenih radova iz domene temporalnog *data mininga* [Agrawal, 1995], [Brazma, 1998], [Ge, 2000], [Geurts, 2001], [Han, 2002], [Han, 2003], [Keogh, 1997], [Polly, 2001], [Pratt, 2001], [Wang, 2001], [Williams, 2002a].

Karakteristika svih navedenih radova, je da imaju autorski pristup ovoj problematici koja premalo pažnje pridaje jedinstvenom modelu transformacije vremenske serije, te su ti radovi fokusirani na rješavanje problematike otkrivanja uzoraka, bez velikih razmatranja kako dobivene rezultate dalje obrađivati primjerice tradicionalnim metodama *data mininga*. Izuzetak je jedino [Williams, 2002a], koji je uočio problem metodološke nepovezanosti, te se njegov rad [Williams, 2002] referencira i na ovaj rad prilikom teorijskog razmatranja *general hidden distribution based analysis theory for temporal data mining*.

Detaljni prikaz ovog teorijskog koncepta, te kritika ove teorije i formiranog modela dana je u poglavlju 1.1.

Ovo empirijsko istraživanje provedeno na REFII modelu prema teorijskim okvirima danih u prethodnim poglavljima, pokazuje kako REFII model rješava probleme otkrivanja uzoraka u vremenskim serijama.

Do sada je pokazano kako se na temelju REFII modela mogu izgraditi rješenja za otkrivanje sezonskih oscilacija, direktnog otkrivanja pravila iz vremenskih serija, te otkrivanje sličnosti. Na temelju jedinstvenog skupa podataka vremenske serije uz pomoć REFII modela, moguće je lančano provesti niz analiza, tako da izlaz iz jedne analize postaje ulaz u drugu analizu.

Ovo poglavlje pokazuje način kako provesti otkrivanje uzoraka na temelju REFII modela, pri čemu je dano izvorno algoritamsko rješenje za tu namjenu.

6.1.1. Izvor i struktura podataka za analizu

Podaci koji su se koristili u analizi izvorni su podaci jednog osiguravajućeg društva i predstavljaju štetnike automobilske police osiguranja.

Tablica 6.1. Struktura tablice štetnika

Podružnica
Registarska oznaka
Broj šasije
Oznaka police
Datum isteka police
Oznaka premije
Ime

Prezime
JMBG/MBO
Oznaka zapisnika
Datum štete
Datum prijave
Mjesto štete
Šifra štete

Tablica sadrži 46,653 slogova za godine odvijanja šteta dijela šteta 2000, te štete u 2001. godini za pravne i fizičke subjekte.

Analitički cilj je otkriti uzorke u vremenskoj seriji prometnih nesreća u 2001. godini i to za fizičke osobe.

U prvom koraku pročišćeni su podaci primjenom SQL upita koji je selektirao samo štete počinjene od strane fizičkih osoba u 2001. godini.

```
SELECT * from stetnici WHERE YEAR(datum_stet)=2001 AND (LEN(stetn)=13 OR LEFT(stetn,2)<>"00") INTO TABLE c:\doktoratstetovi\prva
```

Kao rezultat SQL upita dobiveno je 39,139 slogova

Da bi se pripremili podaci za pretprocesiranje primjenom, REFII modela, potrebno je agregirati vremensku seriju na razinu dana, pri čemu je korištena funkcija agregacije SUM() po danima.

```
SELECT datum_stet,count(*) as broj FROM prva GROUP BY datum_stet ORDER BY datum_stet DESC INTO TABLE c:\doktoratstetovi\druga
```

Na ovaj su način pripremljeni podaci za interpolaciju i naknadno granuliranje.

6.1.2. Vremenska interpolacija, i vremenska granulacija

Vremenska interpolacija i vremenska granulacija sastavni su dio aplikacije za transformaciju vremenske serije u REFII model (*Time explorer*) pisane u programskom jeziku Visual Fox Pro 8.0 (vidi u prilogu Uvodna maska "Time explorera"), a objekt koji vrši vremensku interpolaciju i vremensku granulaciju *granulacija.command1.click()* opisan je u pseudo-algoritmom u poglavlju 2.6.

U okviru istog algoritamskog postupka opisana je i transformacija vremenske serije.

Listing objekta, koji je dio algoritma iz poglavlja 2.6. nalazi se u prilogu "Program za vremensku interpolaciju i vremensku granulaciju". Program *Time explorer* korišten je već u 4. poglavlju kod empirijskih istraživanja prilikom transformacije vremenske serije. Vremenska granulacija ostvarena je na dnevnoj razini.

6.1.3. Transformacija u REFII notaciju

Pročišćene interpolirane i granulirane vremenske serije procesiraju se dalje posredstvom objekta *transform.command1.click()*. Listing objekata prikazan je u prilogu transformacija u REFII model.

Ovaj objekt poziva objekte :

izvrsi.ref(v1,v2):

Objekt za transformaciju u REF oznaku
 $izvrsi.koeficijent(izvrsi.normiraj(v1,mini,maks),izvrsi.normiraj(v2,mini,maks))$
 Objekt za proračun koeficijenta kutnog otklona
 $izvrsi.povrsina(v1,v2):$
 Objekt za proračun površine ispod krivulje,

Program *svrstaj.prg* :

Vrši sekundarnu transformaciju u razrede zadane od strane analitičara

Razredi kutnih otklona su definirani na slijedeći način :

Tablica 6.2. Razredi kutnih otklona

Razred	Donja granica	Gornja granica	REF
Nizak rast	0.000000000	0.300000000	R
Srednji rast	0.300000000	0.700000000	R
Visok rast	0.700000000	1.000000000	R
Nizak pad	0.000000000	0.300000000	F
Srednji pad	0.300000000	0.700000000	F
Oštar pad	0.700000000	1.000000000	F
Bez promjene	0.000000000	0.000000000	E

Ova skupina pomoćnih programa prikazana je u prilogu: pomoćni objekti transformacije, a rađena je na temelju algoritma iz poglavlja 2.6.

Odsječak rezultata obrade prikazan je u narednoj tablici :

Tablica 6.3. Odsječak rezultata obrade

ciklus	indeks	ref	povrsina	razred	koef
1	3	R	65,5	Nizak rast	0,14893617
1	4	R	88,5	Nizak rast	0,046808511
1	5	R	94,5	Nizak rast	0,004255319
1	6	R	101,5	Nizak rast	0,055319149
1	7	F	80,5	Nizak pad	0,234042553
2	1	R	65,5	Nizak rast	0,106382979
2	2	R	88	Nizak rast	0,085106383
2	3	F	92	Nizak pad	0,05106383
2	4	F	82,5	Nizak pad	0,029787234
2	5	F	77,5	Nizak pad	0,012765957
2	6	R	93	Nizak rast	0,144680851
2	7	F	87	Nizak pad	0,195744681
3	1	F	55	Nizak pad	0,076595745
3	2	R	68	Nizak rast	0,187234043
3	3	R	98	Nizak rast	0,068085106
3	4	F	90	Nizak pad	0,136170213
3	5	R	86,5	Nizak rast	0,106382979
3	6	R	105,5	Nizak rast	0,055319149
3	7	F	90,5	Nizak pad	0,182978723
4	1	R	85,5	Nizak rast	0,140425532
4	2	R	107,5	Nizak rast	0,046808511
4	3	E	113	Bez promjene	0
4	4	F	95	Nizak pad	0,153191489
...

Indeks "ciklus" predstavlja tjedne u godini, a indeks "indeks" predstavlja dane u tjednu. Kako je cilj pronaći uzorke unutar vremenske serije, a imajući u vidu korištenje programskog jezika Python te optimizaciju obrade putem lista, izvršena je sekundarna transformacija razreda prema narednoj tablici:

Tablica 6.4. Sekundarna transformacija razreda

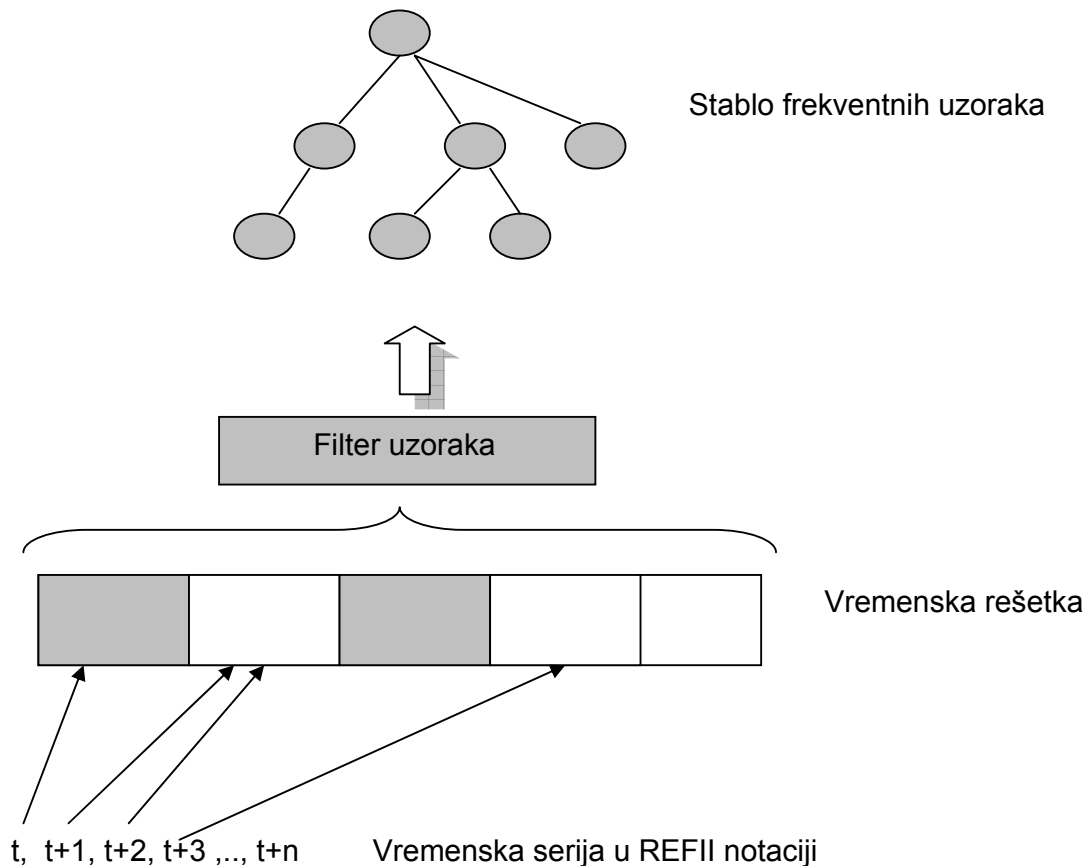
Razred	Oznaka
Nizak rast	1
Srednji rast	2
Visok rast	3
Nizak pad	4
Srednji pad	5
Oštar pad	6
Bez promjene	7

Nakon izvršene transformacije pomoću SQL-a izdvojeno je polje nastalo na temelju sekundarne transformacije i naredbom : *COPY TO c:\doktoratpython\niz.txt DELIMITED* prebačeno u ASCII format.

6.1.4. Rješenje modela otkrivanja uzoraka

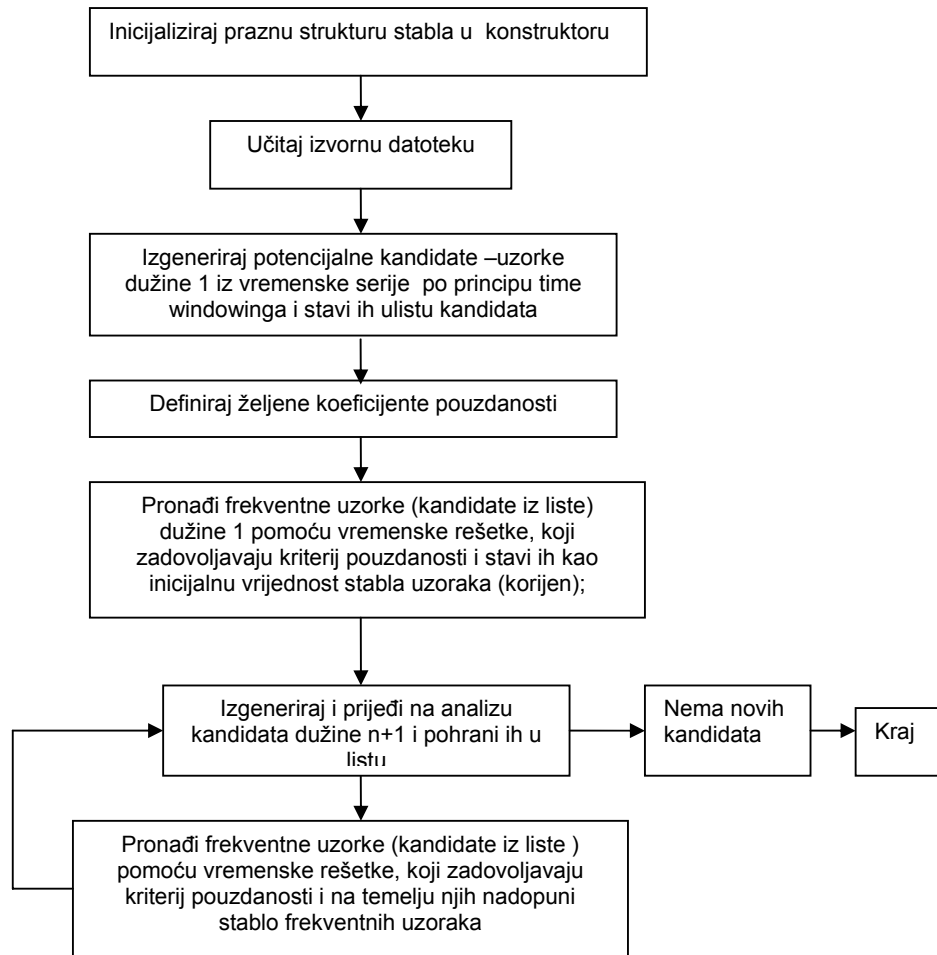
Slična ideja obrazložena je u poglavlju 4.2.1. na primjeru sezonskih oscilacija. Rješenje se svodi na primjeni tri osnovna elementa REFI modelu transformacije, vremenskoj rešetki čiji idejni korijeni potječu iz samoorganizirajućih mapa [Kohonen, 2001] i stabla frekventnih uzoraka [Han, 2000].

Ovaj model prikazan je na slici 6.1.



SLIKA 6.1. MODEL PRETRAGE UZORAKA U VREMENSKOJ SERIJI

Algoritam inicijalno generira kandidate principom vremenskih prozora (eng. time windowing) [Manilla, 1997] gdje se generiraju potencijalni kandidati - uzorci od dužine 1 do dužine $n-1$, gdje n predstavlja ukupan broj elemenata vremenske serije. Klase rješenja pisane u programskom jeziku Python dane su u prilogu: Modul za otkrivanje uzoraka u vremenskim serijama. Problematika pretrage uzoraka na temelju REFI modela riješena je kao što je prikazano na slici 6.2.



SLIKA 6.2. ALGORITAM PRETRAGE UZORKA U VREMENSKOJ SERIJI

Slika 6.2. predstavlja simplificirani algoritam rješenja.

Stablo je reprezentirano nizom slojeva listi na naredan način :

Sloj n :[glava uzorka, promatrani element uzorka, frekvencija pojavnosti uzorka , [Cijeli uzorak do sloja n]]]

Konstruktor u modulu *init*, postavlja inicijalne vrijednosti stabla frekventnih uzoraka, i pomoćne liste za proces obrade.

Klasa *datoteka* čita datoteku, i trpa je inicijalno u listu iz koje će se generirati uzorci.

Glavna klasa *prolaz* poziva klasu *kroz_listu* koja generira kandidate na temelju vremenskih prozora, poziva klasu *sito* koja generira frekventne uzorke za dužinu uzorka 1.

Unutar klase *kroz_listu* nadalje se provodi testiranje frekventnosti preko klase *brisi_nefrekventne*, i brišu se nefrekventni uzorci pozivom klase *filtriraj_listu*.

Klasa *kroz_listu* nadalje provodi otkrivanje uzoraka posredstvom klase *medjuslojno_sito*.

Klasa *medjuslojno_sito* na način ako postoji "glava" uzorka identična traženom uzorku, a ne postoji "rep" uzorka kao element stabla, formira se list stabla sa elementom repa i , brojač uzorka se inicijalizira na vrijednost 1. Ako postoji "glava" uzorka identična traženom uzorku, i postoji "rep" uzorka kao element stabla, tada se inkrementira brojač uzorka.

Program koji poziva modul sa klasama i traži uzorke dan je listingom :

```
import REFII
stablo=Uzorci()
stablo.datoteka("niz.txt")
stablo.prolaz()
print "cvor", stablo.sloj0
print "prvi sloj ", stablo.sloj1
print "drugi sloj ", stablo.sloj2
print "treći sloj ", stablo.sloj3
print "četvrti sloj ", stablo.sloj4
```

Nakon učitavanja modula naredbom import, i kreiranju objekata, te učitavanja datoteke, aktivira se objekt *stablo.prolaz()* koji traži uzorke. Nakon toga print naredbom prikazujemo slojeve stabla frekventnih uzoraka.

Proces traženja uzoraka ilustrativno se može prikazati narednim primjerom u kojem je zadana vremenska serija u REFII modelu, sa slijedećim elementima : (Nizak pad, Nizak rast, Nizak pad, Nizak rast, Oštar pad, Nizak pad).

U prvom koraku određuje se željeni stupanj pouzdanosti uzoraka. Nakon toga proračunavaju se frekvencije pojavnosti uzoraka dužine 1.

Frekvencije pojavnosti su u prikazanom slučaju za dužinu uzorka 1, *Nizak pad* $3/6=0,5$; za *Nizak rast* $2/6= 0.33333$; za *Oštar pad* $1/6=0,1666$.

U narednom koraku generiraju se uzorci dužine dva, te se dobiva lista kandidata (*Nizak pad – Nizak rast*, *Nizak rast-Nizak pad*, *Nizak pad –Nizak rast*, *Nizak rast-Oštar pad*, *Oštar pad – Nizak pad*).

Ako želimo da koeficijent pouzdanosti mora biti $\geq 0,3$, tada se formiraju korijeni stabala sa vrijednostima *Nizak pad* i *Nizak rast*.

Nakon što se u listi kandidata izoliraju samo kandidati koji u glavi uzorka imaju otkrivene frekventne uzorke dužine 1, proračunava se frekventnost uzoraka dužine 2 uz pomoć rešetke samoorganizirajuće mape. Samoorganizirajuća mapa osim kao što je slučaj u ovom primjeru može imati u dimenziji i kategorije površine.

Nakon proračuna frekvencija dužine dva dobije se naredna situacija : (*Nizak pad – Nizak rast* $2/5=0,4$, *Nizak rast-Nizak pad* $1/5=0,2$, *Nizak rast-Oštar pad* $1/5=0,2$).

Po kriteriju koeficijenta pouzdanosti jedini uzorak koji zadovoljava ovaj kriterij je uzorak *Nizak pad – Nizak rast* $2/5=0,4$, te on ulazi u naredni ciklus prilikom generiranja kandidata za uzorke dužine 3. Postupak se ponavlja tako dugo dokle god ima kandidata u listi čiji je koeficijent pouzdanosti veći od graničnog željenog koeficijenta pouzdanosti.

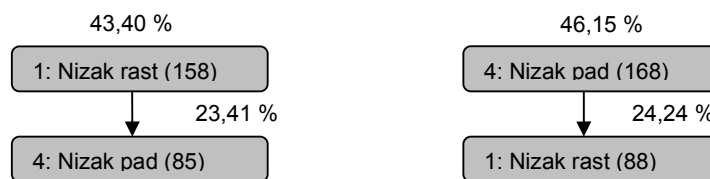
U ovom algoritamskom postupku se krenulo od Hanovog postulata [Han, 2000], koji govori da ako glava uzorka ne zadovoljava kriterij koeficijenta pouzdanosti, tada ni cjelokupan uzorak ne zadovoljava kriterij koeficijenta pouzdanosti. Ušteda vremena redukcijom kombinatorne eksplozije postiže se korištenjem vremenskih rešetki po uzoru na samoorganizirajuće mape, koje u svakom narednom koraku vrše proračune samo na temelju frekventnih uzoraka pohranjenih u stablu frekventnih uzoraka.

6.1.5. Rezultati istraživanja

Nakon aktiviranja opisanih algoritama za koeficijent pouzdanosti 0,2 dobiveni su slijedeći rezultati na populaciji vremenske serije automobilske nesreće 2001 godine na Pentiumu 1.6 GHz, 512 MB RAM-a. Proces obrade trajao je 3,4 sekunde za transformaciju u REFII model, dok je traženje uzoraka trajao 3 sekunde.

cvor [['0', '1', 158, ['1']], ['0', '4', 168, ['4']]]
 prvi sloj [['1', '4', 85, ['1', '4']], ['4', '1', 88, ['4', '1']]]

Stablo uzoraka na temelju dobivenih rezultata izgledaju kao što je to prikazano na slici 6.3.



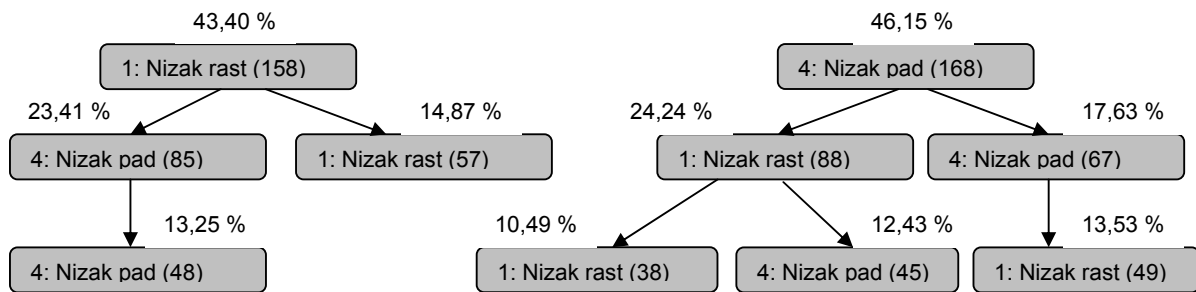
SLIKA 6.3. OTKRIVENI UZORCI SA KOEFICIJENOM POUZDANOSTI 0.2

Pouzdanost uzorka računa se prema formuli (klasa *brisi_nefrekventne*) kao :
 broj frekventnih uzoraka/ broj kandidata uzoraka za testiranje, pri čemu se broj uzoraka za testiranje računa kao :
 broj kandidata u listi-(broj elemenata u uzorku-1). Frekvencija na slici 6.3. izražena kao vrijednost u zagradi odnosi se na pojavnost cjelokupnog uzorka, pri čemu broj instanci prikazuje dužinu uzorka. Tako primjerice uzorak "Nizak rast" dužine uzorka 1 ima frekvenciju pojavnosti 158 i 43,40 % pojavnosti ili koeficijent pouzdanosti 0,43, dok uzorak "Nizak rast"- "Nizak pad " ima frekvenciju pojavnosti 85 i 23,41% pojavnosti odnosno koeficijent pouzdanosti 0,2341.

Ako koeficijent pouzdanosti smanjimo na 0.1 dobivamo slijedeće rezultate:

korijen [['0', '1', 158, ['1']], ['0', '4', 168, ['4']]]
 prvi sloj [['1', '1', 57, ['1', '1']], ['1', '4', 85, ['1', '4']], ['4', '1', 88, ['4', '1']], ['4', '4', 67, ['4', '4']]]
 drugi sloj [['1', '1', 38, ['4', '1', '1']], ['4', '4', 48, ['1', '4', '4']], ['4', '1', 49, ['4', '4', '1']], ['1', '4', 45, ['4', '1', '4']]]

Koje možemo prikazati slikom 6.4.



SLIKA 6.4. OTKRIVENI UZORCI SA KOEFICIJENOM POUZDANOSTI 0.1

Smanjivanjem koeficijenta pouzdanosti može se dobiti razgranatije stablo. Program prikazan u prilogu : "Modul za otkrivanje uzoraka u vremenskim serijama." Pouzdanost i frekvencije za stablo frekventnih uzoraka računa prema preporuci Hanovog rada [Han, 2000] o formiranju stabla frekventnih uzoraka. Veličina populacije se smanjuje svakim novim korakom u kome se povećava dužina uzorka.

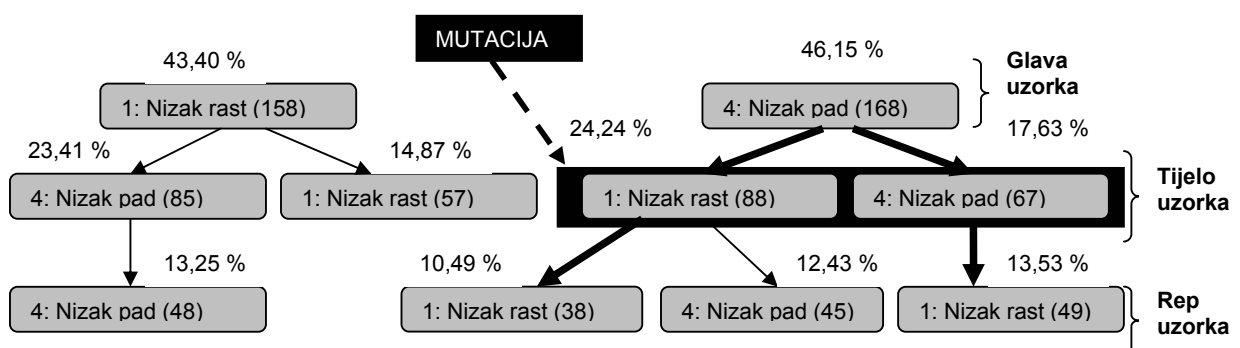
6.2. Otkrivanje mutacija unutar uzoraka

Ovako riješeno otkrivanje uzoraka, prikazano putem lista, pogodno je za rješavanje skupine problema opisanim u radovima [Han, 2002], [Brazma, 1998], [Wang, 2001], koja se odnosi na trženje "mutacija" unutar uzoraka vremenskih serija. Ovaj problem izražen je u primjerice u genetici gdje se traže modaliteti mutacije gena, procjenu djelovanja smetnji na vremensku seriju signala kod telekomunikacijskih sustava, te moguće modalitete diversifikacije uzoraka.

Svaki od navedenih radova iz ovog područja svodi se na vlastiti necjeloviti modalitet rješenja koji rješava samo taj problem.

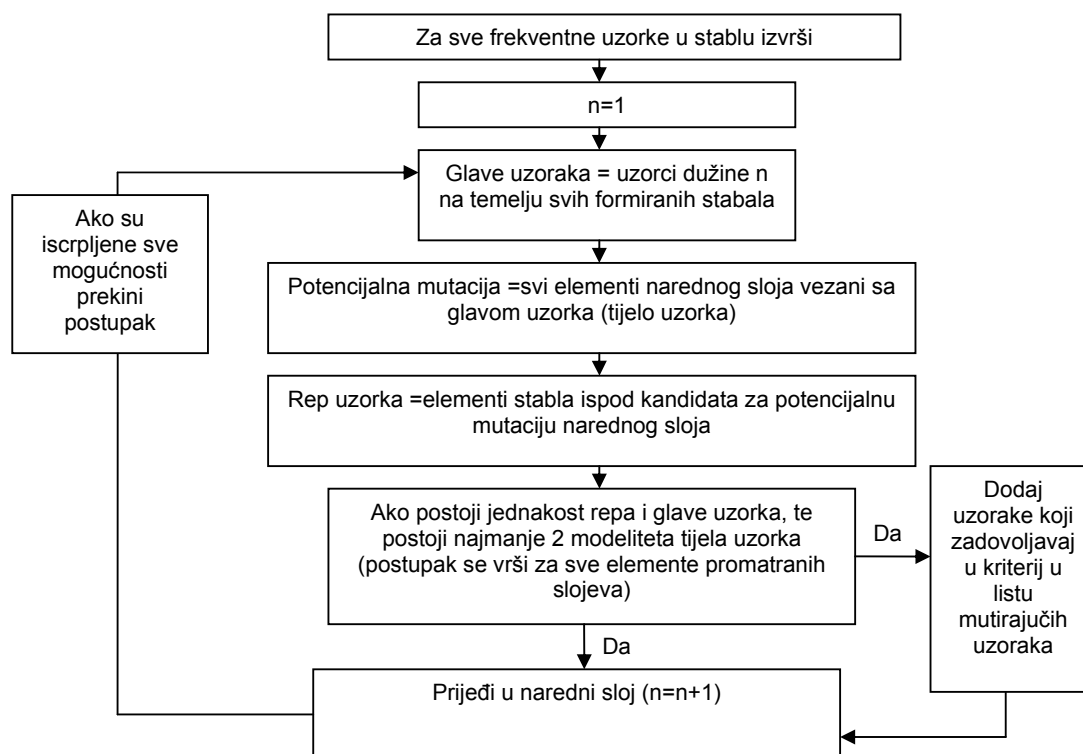
U svjetlu REFIL modela, analizom stabala frekventnih uzoraka možemo dobiti odgovor na pitanje o postojanju mutacije unutar uzorka vremenskih serija.

Osnovni postupak svodi se na usporedbu intervalnih jednakosti glave uzorka, i repa uzorka. Ako su glava uzorka i rep uzorka intervalno jednaki, a tijelo može biti sastavljeno iz različitih uzoraka, tada govorimo o mutacijama, kao što je to prikazano na slici 6.5., proizašloj iz empirijskih podataka na temelju kojih su otkriveni uzorci u prethodnoj analizi..



SLIKA 6.5. OTKRIVANJE MUTACIJE UNUTAR UZORAKA PRI KOEFICIJENU POUZDANOSTI 0.1

Algoritamski postupak ekstrakcije prikazan je na slici 6.6.



SLIKA 6.6. ALGORITAM ZA OTKRIVANJA MUTACIJE UNUTAR UZORAKA

Iz 6.5. vidi se da se pojmovi glave uzorka, repa uzorka i tijela odnose na hijerarhiju u stablu.

Traženje mutacija svodi se na uspoređivanje intervalnih jednakosti glave uzorka i repa uzorka, pri čemu može postojati n glava uzoraka i n repova uzoraka. Ako se otkrije intervalna jednakost među elementima uzorka, tada međuelemente ako ima više od jednog modaliteta međuelementa zovemo mutacijama.

Na primjeru auto nezgoda, između modaliteta "Nizak pad" kao glavu uzorka, i "Niskog rasta" kao repa uzorka postoje međuelementi "Nizak rast" i "Nizak pad" koje nazivamo mutacijama uzorka. Ovako postavljeno rješenje mutaciju promatra kao varijaciju jednog elementa između glave i tijela uzorka. Algoritam je, s obzirom na problemski prostor moguće prilagoditi na način koji mutaciju može promatrati kao varijaciju više od jednog elementa između glave i repa uzorka.

Mutacije, prikazane na slici 6.4. dobivene su na način da je algoritam u prvom koraku glavama mutirajućih uzoraka proglasio elemente prvog sloja (*Nizak rast*; *Nizak pad*).

U istom koraku vezao je uz glave uzoraka pripadajuće elemente drugog sloja, te ih je proglasio tijelima uzoraka (tijela vezano uz *Nizak rast*: *Nizak pad*, *Nizak rast*; tijela vezano uz *Nizak pad*: *Nizak rast*, *Nizak pad*). Nadalje su formirani repovi uz tijela (tijelu *Nizak pad* formiran je rep *Nizak pad*, tijelu *Nizak rast* formiran je rep *Nizak rast*, tijelu *Nizak pad* formiran je rep *Nizak rast*). Ovdje je bitno naglasiti da su tijela i repovi zadržavaju vezu stabla prema glavama. Prema danom algoritmu postoje dva

modaliteta u tijelima (*Nizak rast* i *Nizak pad*) za glavu uzorka *Nizak pad*, sa repom uzorka *Nizak rast*, pa je taj uzorak proglašen mutirajućim uzorkom. U slučaju razgranatijeg stabla ovaj bi se postupak ponavljao sve dok se ne bi iscrpile sve mogućnosti.

Kao što je već spomenuto, s obzirom na analitički zahtjev, tijelo uzorka može sadržavati i više od jednog elementa. U znanstvenim radovima [Han, 2002], [Brazma, 1998], [Wang, 2001] prisutni su različiti pristupi u poimanju mutacija. U ovom primjeru krenulo se je od procjene mutacija gdje se repovi formiraju na temelju elementa iz sloja ispod tijela uzorka na temelju kojeg se procjenjuje mutacija. U spomenutim radovima obrazlažu se i mogućnosti, koje bi u svjetlu prikazanog primjera u formiranju repova obuhvatile sve vezane elemente ispod tijela uzorka. Otkrivanje mutacija unutar uzoraka vrlo je živo znanstveno područje, te tako postoji čitav niz pristupa u rješavanju ove problematike. Cilj prikazanog rješenja, bio je pokazati kako se primjenom REFII modela mogu rješavati ovakve vrste problema, te da je moguće prilagodbom algoritama pretrage na temelju REFII modela obuhvatiti različite pristupe ovoj problematici.

6.3. Otkrivanje sezonskih oscilacija

Na postojećoj populaciji podataka moguće je otkriti sezonske oscilacije po uzoru na algoritam opisan u poglavlju 2.2.1, a odnosio se je na otkrivanje sezonskih oscilacija u REF modelu.

Cilj analize je otkrivanje sezonskih oscilacija na dnevnoj razini. Listing objekta za otkrivanje sezonskih oscilacija, dan je u prilogu sa podnaslovom :Objekt za otkrivanje sezonskih oscilacija. Iako je već prikazana primjena ovog pristupa, ona će biti prikazana i na populaciji vremenske serije auto nezgoda s ciljem dokazivanja hipoteze o efikasnosti jedinstvenog modela transformacije u povezivanju različitih modaliteta analize vremenskih serija na istom uzorku podataka.

Pomoću SPSS naredbe za kreiranje krostabulacijskih tabela :

CTABLES

```
/VARIABLES=ref indeks DISPLAY=DEFAULT
/TABLE indeks [C][ROWPCT.COUNT PCT40.1] BY ref
/CATEGORIES VARIABLES=ref indeks ORDER=A KEY=VALUE EMPTY=EXCLUDE.
```

Dobivena je tablica :

Tablica 6.5. Sezonske oscilacije na temelju trendova

		REF		
		E	F	R
		Red %	Red%	Red %
Indeksi trendova	1	1,9%	86,5%	11,5%
	2	,0%	1,9%	98,1%
	3	3,8%	51,9%	44,2%
	4	1,9%	57,7%	40,4%
	5	3,8%	42,3%	53,8%
	6	,0%	25,0%	75,0%
	7	,0%	86,5%	13,5%

Indeksi trendova imaju značenje :

- 1- Trend pojave sa subote na nedjelju
- 2- Trend pojave s nedjelje na ponedjeljak
- 3- Trend pojave s ponedjeljka na utorak
- 4- Trend pojave s utorka na srijedu
- 5- Trend pojave sa srijede na četvrtak
- 6- Trend pojave sa četvrtka na petak
- 7- Trend pojave sa petka na subotu

Iz izvršene analize vidljivo je da postoje sezonski trendovi rasta broja nezgoda sa nedjelje na ponedjeljak, i to sa 98,1 % pouzdanosti, te sa četvrtka na petak sa 75 % pouzdanosti. Sa petka na subotu bilježi se trend pada broja nezgoda sa stupnjem pouzdanosti 86,5%, kao i sa subote na nedjelju.

Ako želimo veći stupanj preciznosti procjena vezan uz jačinu trendova otklona, u analizu uvrštavamo preciznije razrede, nakon čega dobivamo narednu tablicu sezonskih oscilacija.

Tablica 6.6. Sezonske oscilacije na temelju razreda otklona kutova

		Razred kutnog otklona				
Indeksi trendova		Bez promjene	Nizak pad	Nizak rast	Srednji pad	Srednji rast
		Red %	Red%	Red %	Red %	Red %
	1	1,9%	84,6%	11,5%	1,9%	,0%
	2	,0%	1,9%	86,5%	,0%	11,5%
	3	3,8%	48,1%	38,5%	3,8%	5,8%
	4	1,9%	48,1%	34,6%	9,6%	5,8%
	5	3,8%	38,5%	51,9%	3,8%	1,9%
	6	,0%	25,0%	67,3%	,0%	7,7%
	7	,0%	76,9%	13,5%	9,6%	,0%

Na osnovu ove tablice vidljiva je preciznija razrada sezonskih oscilacija gdje se vidi jačina trendova rasta i pada u određenoj promatranoj sezoni.

Iz izvršene analize vidljivo je da postoje sezonski trendovi niskog rasta broja nezgoda sa nedjelje na ponedjeljak, kao i sa četvrtka na petak.

Sa subote na nedjelju bilježi se trend niskog pada broja nezgoda kao i sa petka na subotu.

S obzirom na otkrivene uzorke "Nizak pad-Nizak rast" i dobivene rezultate po pitanju sezonskih oscilacija, lako je za pretpostaviti da je ovaj uzorak ima jaki sezonski utjecaj i to u periodu od subote do ponedjeljka.

Isto tako možemo pretpostaviti da na uzorak "Nizak pad-Nizak pad " utječe sezonska komponenta u periodu od petka do nedjelje. Ako u algoritam uključimo kriterij pretrage uzoraka u kojem se traži međuzavisnost u sezonskim oscilacijama, kao što je to prikazano u poglavlju 4.4.2., tada je frekventnost pojavnosti uzorka "Nizak pad-Nizak rast", niža nego što je to slučaj kada vremenske odsječke promatramo parcijalno, i ona iznosi 81,7%. Frekventnost pojavnosti uzorka "Nizak pad-Nizak pad " iznosi 71,21%.

Ovi rezultati su dobiveni na temelju primjene algoritma (programa) za otkrivanje frekventnih uzoraka opisanog i već primijenjenog u poglavlju 6.1.4..

Program je izvršen na Pentiumu 1.6 GHz, 512 MB RAM-a., proces obrade traženja uzorka sezonskih oscilacija trajao je 2.4 sekunde.

Uzevši u obzir podatke o strukturi korisnika polica osiguranja, na temelju koje je rađena ova analiza, vidljivo je da su to uglavnom osobe srednjih godina, sa prosječnom starošću automobila 7,8 godina, sa jednim ili više djece. Na temelju rezultata analize sezonskih oscilacija, vidljivo je da postoji trend pad broja automobilskih nesreća u periodu vikenda, a početkom i krajem radnog tjedna, raste broj nezgoda. Uz pomoć ovih pokazatelja može se zaključiti da je riječ o obiteljskim ljudima koji automobile uglavnom koriste za odlazak na posao, te je veća vjerojatnost doživljavanja nesreće početkom i krajem radnog tjedna. Isto tako vikendom je manja vjerojatnost doživljavanja nezgode, što se može pripisati, ili smanjenom korištenju automobila vikendom, ili pojačanom oprezu vozača, jer ako kreću na put, vjerojatno voze obitelj, što im je potencijalno dodatni motivator za oprez na cesti.

6.4. Otkrivanje događajnosti

Događajnost unutar vremenskih serija, također je čest predmet razmatranja znanstvenih radova [Guralnik, 1999], [Manilla, 1997], [Williams, 2003a].

Događajnost se može promatrati i sa perspektive očekivanih uzoraka (motiva) koji se temelje na ekspertnom znanju [Lin, 2001].

Neki znanstveni radovi pojam događajnosti unutar vremenskih serija promatraju u kontekstu događaja koji ne moraju nužno proizlaziti iz numeričkih vrijednosti vremenske serije nego se svode na niz događaja vezanih uz provale i aktivaciju alarma [Manilla, 1997] .

Na temelju koncepcije koja se ovdje obrazlaže moguće je analizirati događajnost nezavisno o rezultatima prethodnih analiza, dakle kao samostalan analitički projekt.

Kako bi se dokazala efikasnost povezivanja različitih analitičkih koncepcija vremenskih serija temeljem jedinstvenog modela transformacije, ovo poglavlje donosi rješenje kako otkrivene uzorke iz vremenske serije o automobilskim nezgodama promatrati kao događaje, te analizirati pojavnost tih događaja unutar vremenske serije.

Zadatak je frekventne uzorke dužine od 3 elementa proglašiti događajima, te promatrati međuzavisnost tih događaja unutar vremenske serije.



SLIKA 6.7. ALGORITAM ZA OTKRIVANJE DOGAĐAJNOSTI

Prvi korak sastoji se u deklariranju događaja na temelju otkrivenih frekventnih uzoraka. Tako možemo deklarirati događaje na temelju rezultata istraživanja iz poglavlja 6.1. :

$A = ['4', '1', '1']$
 $B = ['1', '4', '4']$
 $C = ['4', '4', '1']$
 $D = ['4', '1', '4']$

gdje oznaka 4 predstavlja nizak pad, a oznaka 1 nizak rast u listama.

Događaje možemo proizvoljno deklarirati, kao primjerice "Period kontinuiranog rasta", "Period visokog rasta", "Period volatilnosti" i slično.

Osnovna ideja kreiranja događaja svodi se na prepoznavanje događaja na temelju kreiranih kandidata za potrebe otkrivanja uzoraka i formiranja nove liste događaja sa elementima događaja.

Nakon što je kreirana nova lista događaja, nad njom se provodi postupak otkrivanja uzoraka sukladno algoritmu opisanom u poglavlju 6.1.

Listing ovog postupka pisanog u programskom jeziku Python dan je u prilogu: Kod za otkrivanje događajnosti

Program je izvršen na procesoru Pentiumu 1.6 GHz, 512 MB RAM-a., proces formiranja događaja na temelju otkrivenih uzoraka trajao je 1,7 sekundi, a proces otkrivanja događaja 2,1 sekunde.

Nakon aktivacije programa sa željenim faktorom sigurnosti 0,1 dobiven je naredni rezultat:

korijen [['0', 'A', 38, ['A']], ['0', 'B', 48, ['B']], ['0', 'C', 49, ['C']], ['0', 'D', 45, ['D']]]
 prvi sloj [['A', 'B', 22, ['A', 'B']], ['B', 'C', 45, ['B', 'C']], ['C', 'D', 27, ['C', 'D']], ['D', 'B', 24, ['D', 'B']], ['C', 'A', 20, ['C', 'A']]]
 drugi sloj [['B', 'C', 20, ['A', 'B', 'C']], ['C', 'D', 24, ['B', 'C', 'D']], ['B', 'C', 23, ['D', 'B', 'C']], ['C', 'A', 19, ['B', 'C', 'A']]]

Za potrebe kreiranja razreda površina, preporuča se izvršiti normiranje u intervalu od 0-1.

Također je moguće vidjeti osnovne karakteristike površine ispod vremenskih odsječaka pa izvršiti kreiranje razreda površine za određivanje intervalnih jednakosti na temelju originalnih vrijednosti, gdje se za potrebe preliminarne analize može koristiti naredba u SPSS skriptnom jeziku za određivanje osnovnih pokazatelja kao što su minimum, maksimum, standardna devijacija, i prosječna vrijednost.

To možemo postići narednom naredbom u SPSS –u na uzorku podataka auto odgovornosti transformiranih u REFII sintaksu.

DESCRIPTIVES

VARIABLES=povrsina

/STATISTICS=MEAN STDDEV MIN MAX .

Za potrebe analize, izvršeno je normiranje površine ispod krivulje, te je izvršeno grupiranje u razrede na slijedeći način:

Tablica 6.7. Klasifikacija normirane površine

Raspon vrijednosti	Oznaka intervala	Sekundarna transformacija
0-0.4	Niska površina	1
0.4-0.7	Srednja površina	2
0.7-1	Visoka površina	3

Zadatak je otkriti motive unutar vremenske serije štetnika, gdje poslije perioda niskog rasta sa niskom površinom nastupa period niskog rasta sa srednjom površinom.

Traženi uzorak možemo izraziti narednom strukturom:

Trazeni_motiv=[[1,1], [1,2]]

Za potrebe rješavanja ovog problema, u listu za pretraživanje ubačene su vrijednosti otklona kutova indeksa i površina.

Lista za pretraživanje nakon ove intervencije ima strukturu:

Lista=[[C₁,I₁,O₁,P₁], [C₂,I₂,O₂,P₂] ... [C_n,I_n,O_n,P_n]]

Gdje C - označava indeks ciklusa, u konkretnom slučaju, tjedan,

I – Označava indeks dana

O – Označava intervalu oznaku otklona kuta

P- Označava intervalnu oznaku pripadnosti površine ispod odsječka

Postupak se svodi na prolazak kroz listu za pretraživanje, i registriranje indeksa gdje se pojavljuje traženi motiv.

Kao rezultat obrade (vidi u prilogu program za traženje motiva)dobivena je naredna lista:

Lista_izlaznih_indeksa= [[1,3,1,4], [2,1,2,2], [3,2,3,3], [6,2,6,3], [14,2,14,3], [16,2,16,3], [52,3,52,4]]

Elementi izlazne liste su indeksi, gdje se nalazi traženi motiv, tako primjerice prvi element iz liste [1,3,1,4] ukazuje na to da se traženi motiv nalazi u prvom tjednu od utorka do četvrtka. (Indeks 1 odnosi se na prvi tjedan 2001. godine, indeks 3 označava trend pojave s utorka na srijedu, a indeks 4 označava trend pojave sa srijede na četvrtak).

Za potrebe analize, moguće je (sukladno metodologiji obrađenoj u poglavlju 5.5.) mijenjati širine razreda, kako otklona kutova, tako i razreda površine ispod vremenskih odsječaka.

Motivi se mogu koristiti sa nevremenskim atributima u procesima otkrivanja međuzavisnosti motiva sa nekim od događaja. To je posebno interesantno kod predikcija događaja na primjerice burzi, gdje je pojava nekog vremenskog uzorka u uskoj korelaciji sa događajem značajnim za poslovanje, što možemo izraziti kao :

$$P(M_1|D_n) = p_n$$

Gdje M_1 označava motiv, a D_n događaj, pri čemu p_n označava vjerojatnost događaja D_n uz uvjet pojave motiva M_1 .

Moguće je izvršiti dodatne analize koje bi trebale dati odgovor na pitanje o sezonskim faktorima koji utječu na pojavnost motiva, što ovdje nije slučaj

6.7. Integracija metoda data mininga sa vremenskim serijama

Jedan od vrlo izraženih problema u dome temporalnog *data mininga* je povezivanje vremenskih serija sa tradicionalnim *data mining* algoritmima [Williams, 2002]. Spomenuti autor kao rješenje ovog problema nudi prilagođavanje postojećih algoritama *data mininga* kako bi se oni mogli primijeniti na vremenske serije.

REFII model samom svojom koncepcijom omogućava direktnu primjenu *data mining* algoritama na vremenskoj seriji, pri čemu je upravo REFII model most koji povezuje vremenske serije i tradicionalne *data mining* algoritme.

U poglavlju 4.4. (Izravno otkrivanje pravila iz vremenske serije) dana je jedna od mogućnosti integracije asocijativnih algoritama sa vremenskim serijama.

Potreba povezivanja tradicionalnih *data mining* metoda sa vremenskim serijama proizlazi iz potencijalnih mogućnosti vezanih uz segmentaciju, procjene utjecaja vremenskih i nevremenskih komponenti na trendove vremenske serije iz čega se mogu kreirati modeli ponašanja tržišnih subjekata i segmenata.

REFII model omogućava uvođenje u analizu vremenskih serija i vremenski nezavisnih atributa. Ova metodologija teorijski je objašnjena u poglavlju 5.2., a oslanja se na ekspanziju vremenskih odsječaka netemporalnim atributima. Ova je tehnika detaljnije teorijski obrazložena u poglavlju 5.2.

Osnovni princip proizlazi iz činjenice da na trend neke pojave (trend vremenskog odsječaka) utječu subjekti uz koje se vežu klasični atributi (npr. spol, dob, regija, visina primanja, stupanj stručne sprema). Da bi i ovi atributu ušli u proces analize potrebno je izvršiti postupak ekspanzije vremenskih odsječaka sa netemporalnim atributima. Tako pripremljene podatke možemo analizirati kako statističkim metodama, tako i tradicionalnim *data mining* algoritmima. Ekspanzija vremenskih odsječaka nije preduvjet za primjenu *data mining* algoritama nad vremenskom serijom, već analizu možemo provesti i nad elementima matrice transformacije.

6.7.1. Postupak ekspanzije vremenskih odsječaka

Za potrebe empirijskog istraživanja proveden je postupak ekspanzije vremenskih odsječaka na bazi štetnika iz prethodnih primjera.

Za potrebe empirijskog istraživanja proveden je postupak vremenske ekspanzije na bazi štetnika iz prethodnih primjera. Kako je od postojećih atributa na temelju JMBG-a i registarskih oznaka ekstrahirati dob, spol, i mjesto registracije, (mjesto štete nije u relacijskoj zavisnosti te je s toga neupotrebljivo za analizu) nakon osnovnog pretprocesiranja podataka gdje su otklonjeni nedostajući atributi JMBG-a (strani državljani počinitelji šteta i neupisani JMBG-ovi, te mjesta registracije) preostalo je 30,756 slogova, od 39,139 koji su korišteni kao elementi prethodnih analiza.

Ova populacija ponovo je transformirana u REFII notaciju postupkom koji je opisan u poglavlju 6.1.

Nakon toga provedeno je spajanje matrice transformacije i originalne tablice štetnika preko datumskog polja s ciljem pridruživanja nevremenskih atributa vremenskim odsječaka putem sljedećeg SQL upita:

```
SELECT * from stetnici,refii WHERE datum_stet=vezni AND ((LEN(stetn)=13 OR LEFT(stetn,2)<>"00")) AND (YEAR(datum_stet)=2001 ) INTO TABLE expa
```

Također je izdvojeno polje registarske oznake, gdje prva dva mjesta u stringu označavaju mjesto registracije pomoću naredbe :

```
REPLACE ALL reg_ozn WITH LEFT(regozn1,2)
```

U narednom koraku ekstrahirani su dob i spol na temelju JMBG-a pomoću klasa pisanih u Visual Fox Pro –u transformiraj.dob() i transformiraj.spol()

transformiraj.dob()

PARAMETERS jmbg

local godina

godina = substr(jmbg,5,3)

godina = val(godina)

godina=YEAR(DATE())-(godina+1000)

RETURN godina

transformiraj.spol()

PARAMETERS jmbg

local vrijednost

vrijednost =val(substr(jmbg,10,3))

if vrijednost <500

RETURN "M"

ENDIF

if vrijednost >=500

RETURN "Z"

ENDIF

Nakon ekstrakcije podatka o dobi i spolu provedeno je kreiranje razreda na temelju dobi i to u rasponima

- do 25 godina
- od 26-36 godina
- od 36 do 45 godina
- od 46 do 55 godina
- od 56 do 65 godina
- od 66 godina na više

Kreiranje razreda provedeno je u SPSS skriptnom jeziku pomoću narednog programa :

```
RECODE dob
  ( MISSING = COPY )
  ( LO THRU 25 =1 )
  ( LO THRU 35 =2 )
  ( LO THRU 45 =3 )
  ( LO THRU 55 =4 )
  ( LO THRU 65 =5 )
  ( LO THRU HI = 6 )
  ( ELSE = SYSMIS ) INTO GRUPEDOBI.
VARIABLE LABELS GRUPEDOBI 'GRUPEDOBI'.
FORMAT GRUPEDOBI (F5.0).
VALUE LABELS GRUPEDOBI
  1 '<= 25'
  2 '26 - 35'
  3 '36 - 45'
  4 '46 - 55'
  5 '56 - 65'
  6 '66+'.
MISSING VALUES GRUPEDOBI ( ).
VARIABLE LEVEL GRUPEDOBI ( ORDINAL ).
EXECUTE.
```

Ovim postupkom je okončan postupak ekspanzije vremenskih odsječaka nevremenskim atributima.

Kao što je već spomenuto, ekspanzija vremenskih odsječaka nije preduvjet za provođenje analize vremenske serije primjenom tradicionalnih *data mining* algoritama. Ako primjerice primijenimo stablo odlučivanja direktno na matricu transformacije, tada u analizu možemo uključiti otklone kutova vremenske serije, površinu ispod krivulja i vremenske indekse. Na ovaj način možemo otkriti korelacijske zavisnosti između otklona kutova, površine i indeksa. Za potrebe segmentacije i procjene ponašanja subjekata potrebno je uključiti i što više popratnih atributa, te se s toga provodi ekspanzija vremenskih odsječaka.

6.7.2. Tablična i OLAP analiza

Jedan od izazova koji se često puta spominje u okviru radova vezanih uz upitni jezik za vremenske serije je i mogućnost analize vremenskih serija pomoću OLAP procesa, te ih integrirati u skladišta podataka s ciljem multidimenzionalne analize [Perng, 2002]. REFIL model rješava ovaj problem, integrirajući i mogućnost ekspanzije vremenskih odsječaka nevremenskim atributima što daje niz dodatnih mogućnosti po pitanju provođenja analiza posebice za potrebe segmentacije tržišta, procjene ponašanja tržišnih subjekata i segmenata.

Tako možemo provesti analizu utjecaja varijable spol na trendove odsječaka vremenske serije pomoću jednostavnog upita na ekspanzirani model čiji je postupak formiranja prikazan u poglavlju 6.7.1.

Naredba koja će nam prikazati tablicu utjecaja varijable spol na trendove odsječaka vremenske serije u SPSS skriptnom jeziku glasi :

CTABLES

```
/VLABELS VARIABLES=spol razred DISPLAY=DEFAULT
```

```
/TABLE razred [ROWPCT.COUNT PCT40.1] BY spol
```

```
/CATEGORIES VARIABLES=spol razred ORDER=A KEY=VALUE  
EMPTY=EXCLUDE.
```

Kao rezultat upita dobivena je tablica :

Tablica 6.8. Otkloni kutova prema spolu

		Spol	
		M	Z
		%	%
Razred	Bez promjene	81.5%	18.5%
	Nizak pad	79.5%	20.5%
	Nizak rast	79.3%	20.7%
	Srednji pad	80.2%	19.8%
	Srednji rast	77.7%	22.3%

Kako je populacija štetnika dominantno muškog spola, ipak su vidljiva određena odstupanja po spolu s obzirom na trend vremenskih odsječaka. Tako primjerice prilikom trenda srednjeg rasta prometnih nesreća udio pripadnika ženskog spola viši je nego kod ostalih vrsta trendova.

Daljnja analiza može se provesti s ciljem procjene utjecaja mjesta registracije na trend vremenskog odsječka.

Za tu svrhu u koristimo naredbu u SPSS skriptnom jeziku :

CTABLES

```
/VLABELS VARIABLES=razred reg_ozn DISPLAY=DEFAULT
```

```
/TABLE reg_ozn [ROWPCT.COUNT PCT40.1] BY razred
```

```
/CATEGORIES VARIABLES=razred ORDER=A KEY=VALUE EMPTY=EXCLUDE
```

```
/CATEGORIES VARIABLES=reg_ozn ORDER=D KEY=COUNT EMPTY=EXCLUDE.
```

Kao rezultat obrade dobivena je tablica :

Tablica 6.9. Otkloni kutova prema mjestu registracije

		Razred				
		Bez promjene	Nizak pad	Nizak rast	Srednji pad	Srednji rast
		Red %	Red %	Red %	Red %	Red %
	ZG	1.4%	38.3%	48.1%	3.5%	8.7%
	ST	1.5%	40.9%	46.1%	3.9%	7.6%
	PU	1.3%	42.9%	44.9%	3.6%	7.2%
	RI	1.9%	39.9%	45.9%	4.1%	8.2%
	OS	1.1%	36.7%	51.5%	2.4%	8.3%
	ZD	.9%	40.7%	46.3%	5.3%	6.9%
	ČK	1.9%	41.3%	46.5%	4.1%	6.1%
	KA	1.4%	36.8%	49.7%	4.4%	7.7%
	KR	2.3%	39.1%	45.4%	4.0%	9.3%
	SK	1.3%	38.6%	47.3%	4.2%	8.6%
	VŽ	1.6%	40.3%	45.9%	2.7%	9.4%
	SB	1.5%	36.7%	46.1%	6.5%	9.2%
	KC	1.3%	38.2%	47.9%	5.3%	7.4%
Registarska oznaka	DU	1.7%	43.9%	43.2%	4.9%	6.3%
	ŠI	2.3%	39.1%	48.8%	3.6%	6.1%
	BJ	.8%	45.5%	46.8%	2.2%	4.7%
	VK	1.2%	43.6%	46.5%	2.3%	6.4%
	NA	1.5%	39.8%	47.5%	5.6%	5.6%
	PŽ	.6%	41.7%	40.7%	7.1%	9.9%
	VU	1.5%	43.5%	44.6%	1.9%	8.5%
	DA	1.8%	40.3%	48.2%	2.7%	7.1%
	BM	2.3%	34.5%	50.0%	5.5%	7.7%
	DJ	1.0%	37.9%	47.1%	3.4%	10.7%
	KT	2.5%	40.7%	42.6%	3.4%	10.8%
	IM	.5%	43.1%	47.0%	1.5%	7.9%
	VT	1.5%	46.4%	38.1%	6.2%	7.7%
	KŽ	3.1%	44.3%	41.7%	2.6%	8.3%
	NG	.5%	37.4%	50.5%	4.4%	7.1%
	MA	1.7%	37.1%	50.3%	5.1%	5.7%
	GS	1.7%	39.3%	48.0%	3.5%	7.5%
	PS	2.5%	38.5%	48.4%	.8%	9.8%
	ŽU	.0%	33.7%	57.1%	1.0%	8.2%
	OG	.0%	43.3%	51.7%	.0%	5.0%
	DE	.0%	34.6%	50.0%	.0%	15.4%

Iz tablice je vidljivo da je primjerice veća vjerojatnost sudjelovanja vozila Zagrebačke registracije u prometnim nesrećama u razdobljima niskog rasta trendova vremenskih odsječaka, dok je primjerice veća vjerojatnost sudjelovanja vozila Križevačkih registracija u prometnim nesrećama u razdobljima niskog pada trendova vremenskih odsječaka. Ovakve analize nam mogu pomoći prilikom kreiranja uvjetnih vjerojatnosti sudjelovanja vozila sa određenim registarskim oznakama s obzirom na očekivani trend pojave. Ako ovu analizu promatramo u kontekstu provedene analize otkrivanja vremenskih uzoraka i sezonskih oscilacija (poglavlja 6.1. i 6.3.), tada možemo procijeniti primjerice da je s nedjelje na ponedjeljak veća vjerojatnost sudjelovanja

nesreće Križevačkih registarskih oznaka, nego Zagrebačkih, jer je u tom periodu zabilježena sezonska oscilacija niskog pada. Ovim principom moguće je izgraditi i Bayesovu mrežu utjecaja kao model ponašanja tržišnih segmenata i subjekata.

Nadalje možemo promatrati ukupnu strukturu utjecaja na trend s obzirom na registarsku oznaku automobila, gdje više do izražaja dolazi frekvencija pojavnosti nezgoda s obzirom na mjesto registracije.

Za tu svrhu u koristimo naredbu u SPSS skriptnom jeziku :

CTABLES

/VLABELS VARIABLES=razred reg_ozn DISPLAY=DEFAULT

/TABLE reg_ozn [COLPCT.COUNT PCT40.1] BY razred

/CATEGORIES VARIABLES=razred ORDER=A KEY=VALUE EMPTY=EXCLUDE

/CATEGORIES VARIABLES=reg_ozn ORDER=D KEY=COUNT EMPTY=EXCLUDE

Kao rezultat obrade dobivena je naredna tablica :

Tablica 6.10. Otkloni kutova prema mjestu registracije(struktura)

		Razred				
		Bez promjene	Nizak pad	Nizak rast	Srednji pad	Srednji rast
		Kolona %	Kolona %	Kolona %	Kolona %	Kolona %
	ZG	34.1%	33.8%	35.5%	32.4%	37.7%
	ST	10.4%	10.5%	9.9%	10.6%	9.5%
	PU	6.8%	8.2%	7.2%	7.4%	6.8%
	RI	9.0%	6.9%	6.7%	7.5%	7.0%
	OS	4.0%	5.1%	6.0%	3.6%	5.6%
	ZD	1.8%	3.0%	2.9%	4.1%	2.5%
	ČK	3.5%	2.9%	2.7%	3.0%	2.1%
	KA	2.6%	2.5%	2.9%	3.2%	2.6%
	KR	4.0%	2.6%	2.5%	2.8%	3.0%
	SK	2.0%	2.3%	2.3%	2.6%	2.5%
	VŽ	2.4%	2.3%	2.2%	1.6%	2.7%
	SB	2.0%	1.8%	1.9%	3.4%	2.2%
	KC	1.3%	1.5%	1.6%	2.2%	1.4%
Registarska oznaka	DU	1.8%	1.7%	1.4%	2.0%	1.2%
	ŠI	2.4%	1.5%	1.6%	1.5%	1.2%
	BJ	.7%	1.4%	1.2%	.7%	.7%
	VK	.9%	1.2%	1.1%	.7%	.9%
	NA	1.1%	1.1%	1.1%	1.6%	.8%
	PŽ	.4%	1.1%	.9%	2.0%	1.3%
	VU	.9%	.9%	.8%	.4%	.9%
	DA	.9%	.7%	.8%	.5%	.6%
	BM	1.1%	.6%	.8%	1.0%	.7%
	DJ	.4%	.6%	.7%	.6%	.9%
	KT	1.1%	.7%	.6%	.6%	.9%
	IM	.2%	.7%	.7%	.3%	.6%
	VT	.7%	.7%	.5%	1.0%	.6%
	KŽ	1.3%	.7%	.6%	.4%	.6%
	NG	.2%	.6%	.6%	.7%	.5%
	MA	.7%	.5%	.6%	.8%	.4%
	GS	.7%	.6%	.6%	.5%	.5%
	PS	.7%	.4%	.4%	.1%	.5%
	ŽU	.0%	.3%	.4%	.1%	.3%
	OG	.0%	.2%	.2%	.0%	.1%
	DE	.0%	.1%	.1%	.0%	.2%

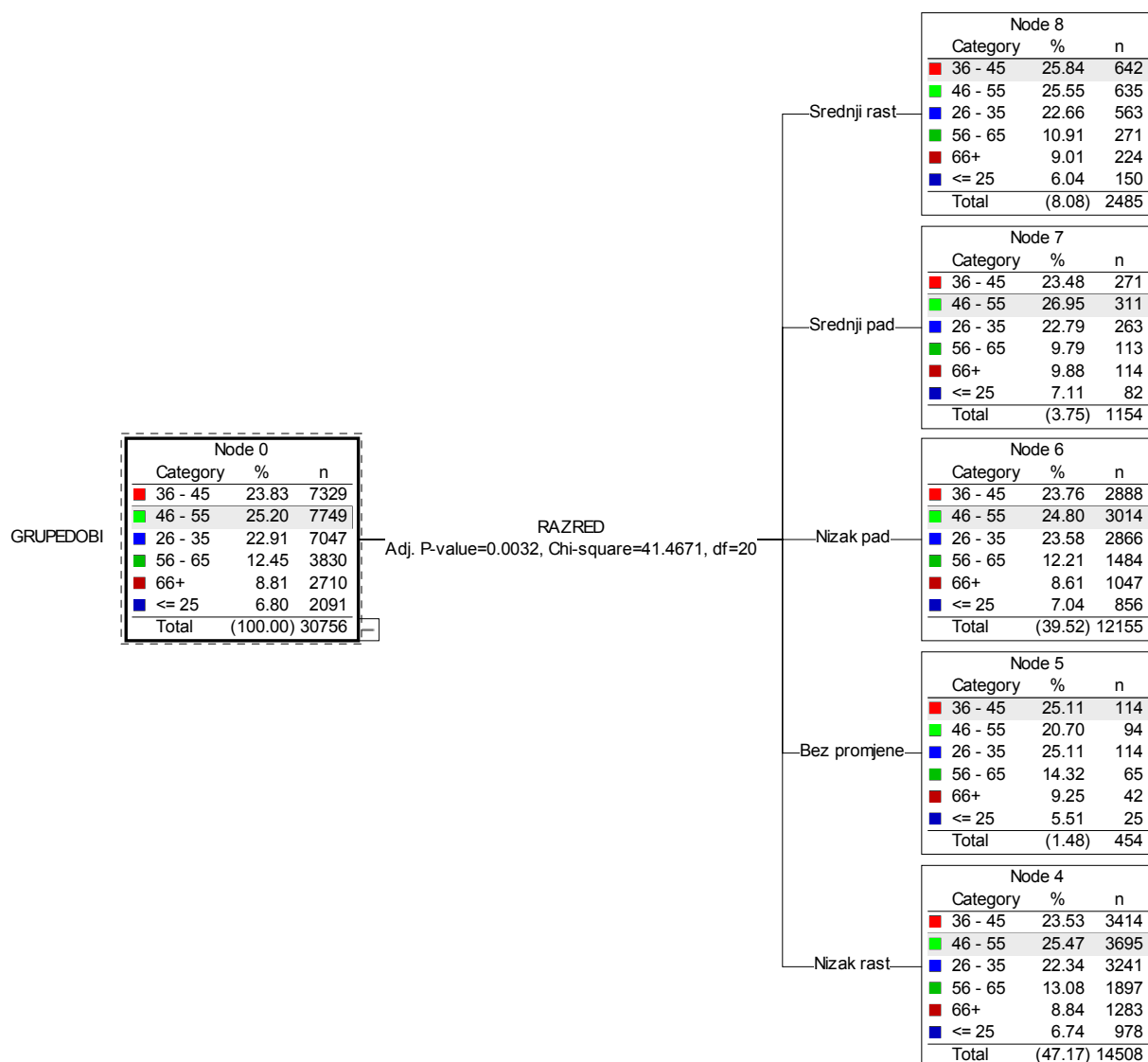
Iz dobivene tablice vidljivi su utjecaji mjesta registracija na trend s posebnim osvrtom na frekvenciju pojavnosti nezgoda. Ovdje je vidljivo da Zagreb prednjači u broju nezgoda u periodima srednjeg rasta trendova vremenskih odsječaka. Postoji čitava paleta mogućnosti koja prvenstveno ovisi o analitičkom cilju.

6.7.3. Primjena stabla odlučivanja na vremensku seriju

Nad vremenskom serijom u REFII notaciji, bilo u obliku izvorne matrice transformacije, bilo u obliku ekspanzirane matrice transformacije, moguće je direktno primijeniti stablo

odlučivanja. Pomoću stabla odlučivanja između ostalog možemo segmentirati populaciju te generirati pravila sa pripadajućom vjerojatnošću.

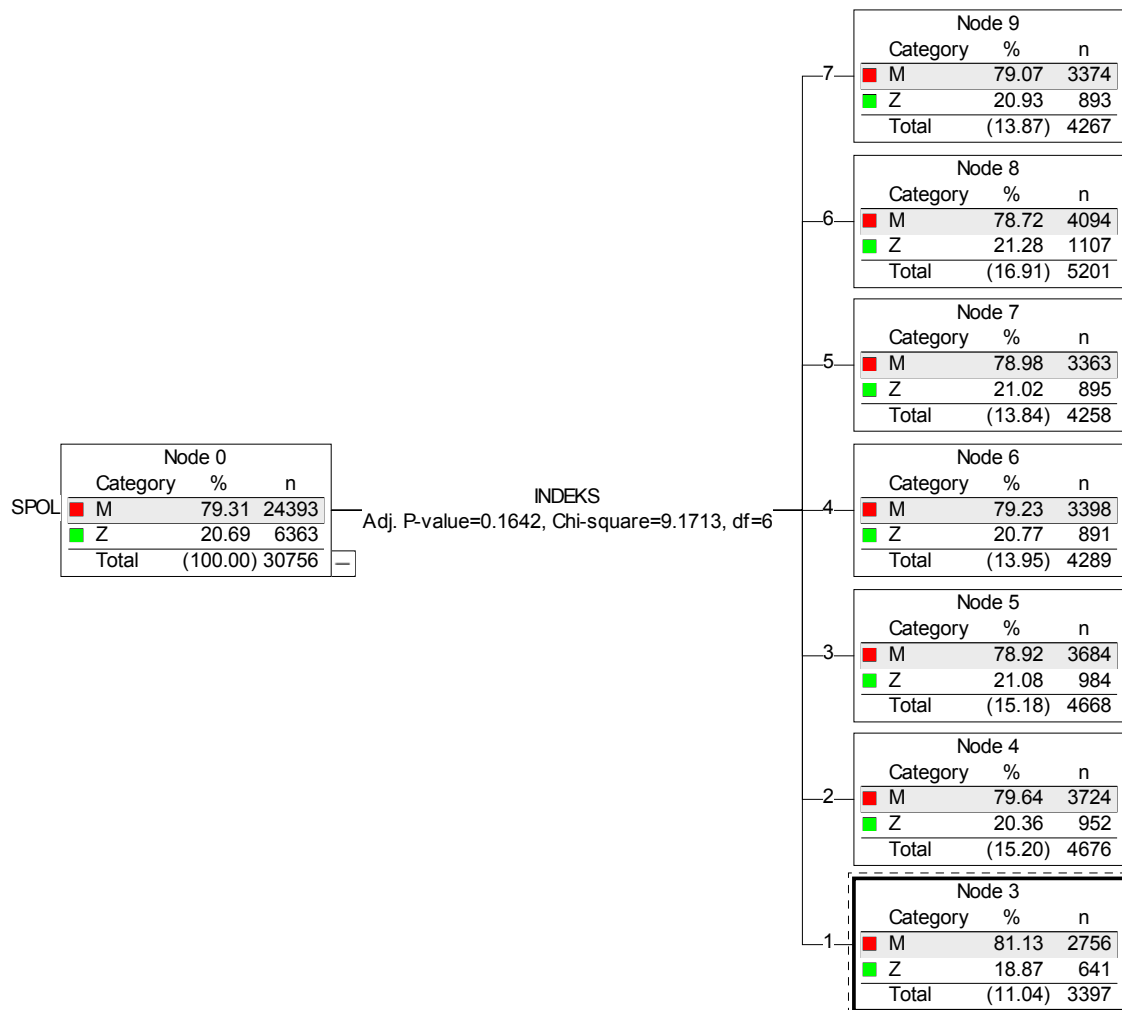
Nakon što je na ekspaniranu matricu transformacije primijenjen CHAID algoritam stabla odlučivanja (SPSS Answer tree 3.0) sa ciljnom varijablom razreda dobi dobiveno je slijedeće stablo odlučivanja :



SLIKA 6.9. STABLO ODLUČIVANJA GRUPE DOBI- TRENDovi VREMENSKIH OTKLONA

Iz stabla odlučivanja vidljivo je najveća vjerojatnost pojavnosti dobi od 46-55 godina kod trendova niskog rasta, niskog pada, i srednjeg pada, a najveća vjerojatnost pojavnosti dobi od 36-45 kod trendova srednjeg rasta i bez promjene.

Ako se promatra spol kao ciljna varijabla a periodi trendova kao zavisne varijable, tada se dobiva naredno stablo odlučivanja:



SLIKA 6.10. STABLO ODLUČIVANJA SPOL-TJEDNI TRENDovi

Prema ovom stablu odlučivanja vidljiva je vjerojatnost pojavnosti modaliteta varijable spol s obzirom na trendove. Tako je evidentno da je sa subote na nedjelju (indeks 1) manji utjecaj pripadnika ženskog spola na formiranje trenda. Odnosno dominantniji je utjecaj pripadnika muškog spola na formiranje trenda. To možemo izraziti u formi pravila kao :

```

IF (INDEKS = 1)
THEN
    Node = 3
    Prediction = M
    Probability = 0.811304
  
```

6.7.4. Primjena klasteriranja na vremenskim serijama

Metoda *data mininga* koja se gotovo tradicionalno primjenjuje na vremenskim serijama je klasteriranje. Razlog popularnosti primjene ove metode u analizi vremenskih serija proizlazi iz činjenice što je možemo primijeniti bez nužnog preprocesiranja.

U svjetlu REFII modela, ona je primijenjena na matricu transformacije, a moguće ju je primijeniti i na ekspanzirani model matrice transformacije, te može biti vrlo snažan alat za segmentaciju.

Kako bi se ilustrirala primjena, prilikom klasteriranja uzete su u obzir dvije varijable, koeficijent kutnog odklona, i oznaka trenda REF zbog jasnoće i moguće lakše vizualizacije rezultata.

Nad ekspanziranom matricom transformacije izvršeno je klasteriranje "K-mean" klasteriranjem primjenom naredbe iz SPSS skriptnog jezika:

QUICK CLUSTER

```
koef REF
/MISSING=LISTWISE
/CRITERIA= CLUSTER(4)
/METHOD=CLASSIFY
/SAVE CLUSTER
/PRINT INITIAL.
```

Kao rezultat obrade dobiveni su klasteri :

Tablica 6.11. Broj slučajeva u klasterima

Broj slučajeva u svakom od klastera		
Cluster	1	15320.000
	2	1318.000
	3	2127.000
	4	11991.000
Valid		30756.000
Missing		.000

Sa vrijednostima centroida:

Tablica 6.12. Vrijednosti centroida

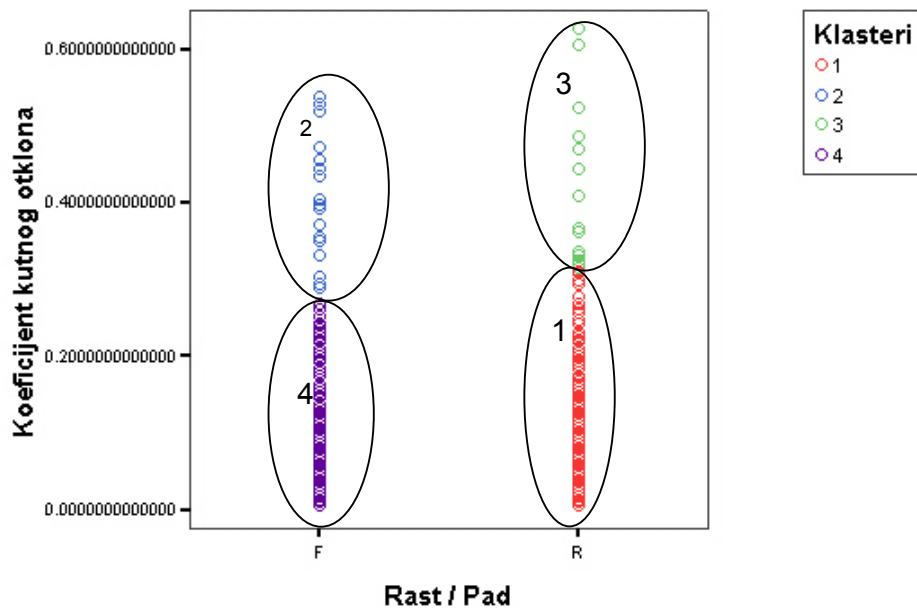
Centroidi klastera

	Cluster			
	1	2	3	4
Koeficijent kutnog odklona	.1221468251764	.4019500855584	.4319338995088	.1076498863509
REF (R=1, F=-1)	1	-1	1	-1

naredbe :

```
IGRAPH /VIEWNAME='Scatterplot' /X1 = VAR(ref) TYPE = CATEGORICAL /Y =
VAR(koef) TYPE = SCALE /COLOR = VAR(PRAVICLUST) TYPE =
CATEGORICAL /COORDINATE = VERTICAL /X1LENGTH=3.0 /YLENGTH=3.0
/X2LENGTH=3.0 /CHARTLOOK='NONE' /CATORDER VAR(PRAVICLUST)
(ASCENDING VALUES OMITEMPTY) /CATORDER VAR(ref) (ASCENDING
VALUES OMITEMPTY) /SCATTER COINCIDENT = NONE.
```

EXE.



SLIKA 6.11. KLASITERI KONCENTRACIJE KUTNOG OTKLONA

Na temelju provedene analize vidljiva je koncentracija klastera, što nam može pomoći prilikom definiranja razreda kutnog otklona. Na temelju klasteriranja također možemo provesti analizu sličnosti tržišnih subjekata na način uključivanja kako vremenskih tako i nevremenskih atributa.

6.8. Konsolidacija rezultata kroz prizmu tržišnih analiza

Kroz poglavlje 6. prikazane su mogućnosti provođenja analize vremenskih serija nad istom vremenskom serijom temeljem jedinstvenog modela transformacije (REFII modela), gdje je na temelju jedinstvene matrice transformacije provedena čitava lepeza analiza : otkrivanje uzoraka, otkrivanje sezonskih oscilacija, otkrivanje mutacija unutar vremenskih serija, pronalaženje motiva unutar vremenskih serija, otkrivanje događajnosti, analiza vremenskih serija posredstvom tradicionalnih metoda *data mininga*. Na ovaj je način dokazana uspješnost i efikasnost REFII modela u domeni povezivanja različitih koncepcija temporalne *data mining* analize, na način da je polazišna osnova za sve vrste analiza jedinstveni model transformacije -REFII model. Nadalje pokazano je kako je moguće vršiti "dublju analizu", pri čemu "Dublja analiza" pojmovno znači mogućnost provedbe različitih tipova analiza na reduciranom setu podataka iz vremenske serije, koji je proizašao kao rezultat obrade prije primijenjene metode ili analitičkoga postupka.

Primjer za dublje analize u poglavlju 6. vidljiv je iz analize događajnosti koja se nastavlja kao lančani proces na rezultate analize proizašle iz rezultata otkrivanja uzoraka unutar vremenske serije. Otkrivanja mutacija, je analiza koja se također

nadovezuje na rezultate analize proizašle iz rezultata otkrivanja uzoraka unutar vremenske serije.

U 6. poglavlju također je pokazano kako analizirati vremensku seriju primjenom *data mining* metoda.

Veći dio ovih problema naveden je kao budući izazov istraživanja u domeni temporalnog *data mininga* [Williams, 2002].

Ako sve ove analize promatramo sa perspektive analize tržišta, tada se mogu povezivati rezultati proizašli iz provedenih analiza s ciljem otkrivanja zakonitosti tržišta, segmentacije tržišta, procjene ponašanja tržišnih segmenata, i tržišnih subjekata. Ako primjerice uzmemo u obzir otkrivene uzorke "Nizak pad-Nizak rast" i dobivene rezultate po pitanju sezonskih oscilacija, lako je za pretpostaviti da je ovaj uzorak ima jaki sezonski utjecaj i to u periodu od nedjelje do utorka s obzirom rezultate proizašle iz analize sezonskih oscilacija.

Također se u obzir mogu uzeti otkriveni događaji "Nizak pad-Nizak rast- Nizak rast " nakon kojeg se može očekivati događaj "Nizak rast-Nizak pad- Nizak pad ", a nakon njega događaj "Nizak pad-Nizak pad- Nizak rast".

Temeljem rezultata analize gdje su primijenjene metoda OLAP-a , te stabla odlučivanja, može se pretpostaviti (ako se promatra Zagrebačka regija) da će u periodima trenda niskog rasta biti povećana vjerojatnost nesreća od strane vozila koje imaju Zagrebačku registraciju. U periodu trenda niskog rasta ugroženija je dobna skupina od 46-55 godina, muškog spola.

Na ovaj se način mogu raditi segmentacije, te promatrati ili cijeli uzorak, ili primjerice Varaždinska regije, tražiti uzorci i sezonske oscilacije samo za tu regiju, ili samo za određenu skupinu i slično. Isto tako prikazanom metodologijom moguće je provoditi procjene rizičnosti kako tržišnih segmenata u određenom vremenskom periodu, tako i segmentaciju s obzirom na unaprijed zadane kriterije.

Prilikom promatranja rezultata empirijskih istraživanja u obzir je potrebno uzeti činjenicu zadanih koeficijenata pouzdanosti, te relativno signifikantno male razlike u vjerojatnostima kod procjena primjenom metoda stabla odlučivanja. Ovi pokazatelji mogu varirati s obzirom na izvorne empirijske podatke nad kojima se provodi analiza. Cilj je bio pokazati efikasnost REFI modela, njegovu sposobnost povezivanja različitih koncepcija analize, od tradicionalnih koje spadaju u područje *data miniga* vremenskih serija, pa sve do povezivanja sa tradicionalnim metodama *data mininga*.

Ovdje je također dokazana teza spomenuta u prvom poglavlju rada da nije potrebno prilagođavati algoritme *data mininga* za primjenu u vremenskim serijama [Williams, 2002], nego da je prvenstveno potrebno izgraditi kvalitetan model transformacije koji će služiti poput mosta između algoritama *data mininga* i vremenskih serija.

Koncepcija REFI modela i predložena metodologija može biti korisna u provođenju tržišnih analiza iz različitih područja.

Tako primjerice u bankarstvu možemo otkrivati uzorke modela ponašanja s obzirom na dizanje gotovine sa bankomata. Uzorke možemo tražiti u segmentiranoj populaciji (prema dobi, spolu, primanjima...), te na temelju otkrivenih uzoraka analizirati događajnost, i tražiti skriveno znanje o karakteristikama klijenata pomoću neke od metoda *data mininga* kao što je stablo odlučivanja, uzimajući u obzir attribute poput razreda prosječno iskorištenog limita po kreditnoj kartici, razredu iznosa depozita, sklonosti određenom tipu valute, vrste korištenog kredita i slično.

U trgovini se primjerice mogu promatrati vremenske serije potražnje za određenim skupinama artikala, pratiti sezonske oscilacije, te analizirati karakteristike klijenata koji najviše utječu na sezonske trendove pomoću primjerice stabla odlučivanja.

Isto je tako moguće segmentirati klijente iz turističke, trgovačke, bankarske ili neke druge branše, formiranjem klastera čiji su elementi osim tradicionalnih atributa i atributi koji se odnose na koeficijente kutnog otklona te površinu ispod krivulje.

Postoji čitav niz mogućnosti, te tijek analize ponajviše ovisi o cilju koji se želi postići.

U radu nisu prikazane sve potencijalne mogućnosti primjene REFII modela, poput primjene metoda poput *fuzzy* logike, *survival* analize, memorijski temeljenog razlučivanja i slično u kombinaciji sa vremenskim serijama.

Dana je generalna metodologija na osnovu koje je intuitivno lako provesti i integraciju ovih metoda, a kao što je već spomenuto o cilju analize ovisi izbor alata za analizu.

6.9. Analiza tržišta tekstila pomoću REFII modela

REFII model primijenjen je na podacima poduzeća Tekstilpromet d.d. i Lantea d.d. s ciljem otkrivanja zakonitosti ponašanja kupaca tih poduzeća. Ovo istraživanje je provedeno u okviru izrade magistarskog rada Lea Mršića (mentor prof. dr.sc. Željko Panian), zaposlenika Tekstilprometa d.d., te se je model primijenio i za provođenje kompleksnijih analiza ponašanja kupaca za interne potrebe poduzeća. U daljnjem tekstu navodim dio opisa projekata otkrivanja tržišnih zakonitosti tržišta tekstila pomoću REFII modela, sa najznačajnijim rezultatima istraživanja³.

" U skladu s nastojanjima za spoznajom zakonitosti u prodaji te pružanjem što kvalitetnije usluge vlastitim kupcima u tvrtkama Tekstilpromet d.d. i Lantea d.d. primijenjen je REFII model, autora Gorana Klepca (Raiffeisen Consulting) razvijen u okviru istraživanja za potrebe doktorske disertacije (mentor prof. dr. sc. Božidar Kliček – FOI). Na podacima tvrtke model je primijenjen u izradi magistarskog rada autora ovog članka uz dodatne analize pojedinih područja za interne potrebe.

Primjeni modela prethodila je analiza i prikupljanje podataka a za korišteni su podaci o vremenskim prilikama i statistički podaci o naseljenosti i strukturi stanovništva gradskih četvrti Grada Zagreba. Za potrebe istraživanja od Gradskog zavoda za planiranje razvoja Grada i zaštitu okoliša kupljeni su statistički podaci o stanovništvu podijeljeni po gradskim četvrtima. Državni hidrometeorološki Zavod ustupio je podatke o klimatološkim uvjetima proteklih godina. Važno je napomenuti da su ovi podaci javno dostupni na navedenim mjestima.

Cilj istraživanja bio je ukazati na potencijale primjene modela te postavljanje temeljnih pretpostavki za daljnja, interna, istraživanja. Interni podaci na kojima je provedeno istraživanje odabrani su između pojedinih proizvoda za koje je procijenjeno da bi mogli relevantno prikazivati okolnosti te nazvani po grupama roba kojima pripadaju, te agregirani na razini dana za ukupno promatranu prodaju. Svrha ovakve pripreme je ukazivanje na potencijal primjene uz istovremeno poštovanje odgovarajuće tajnosti poslovnih podataka koja je ovim pristupom osigurana.

³ Autor Mršić Leo, članak pod naslovom: "*Utjecaj vremenskih promjena na kupnju tekstilnih proizvoda*", objavljen u "*Business Intelligence*" prilogu u dnevnom poslovnom listu "*Dnevnik*" 30.12.2004., 16. stranica, odobren od strane Tekstilpromet d.d. i Lantea d.d. za objavu.

Podaci su uzeti za period od godine dana te kategorizirani prema godišnjem dobu, danu u tjednu, prazniku, temperaturi, vlažnosti zraka, količini oborina i snijega i količinama prodaje. Transformacija podataka u REFII model vršena je programom Time Explorer autora Gorana Klepca (Raiffeisen Consulting) također razvijenog u okviru istraživanja za potrebe doktorske disertacije. Nakon transformacije provedena je analiza vremenskih serija te na dobivenim podacima izgrađena Bayesova mreža.

Provedenim istraživanjem otkrivene su zanimljive, na prvi pogled iznenađujuće, zakonitosti odnosa vremenskih prilika i razine prodaje (npr. visok dnevni temperaturni prosjek utječe pozitivno na razinu prodaje). Vremenske prilike imaju sve veći utjecaj na trgovinu tekstilom uslijed neuobičajenih promjena uvjetovanih globalnim zatopljenjem kojih smo svjedoci proteklih godina, stoga je ovaj pristup vrlo zanimljiv. Uz to uočene su preferencije istovremene kupnje pojedinih grupa promatranih proizvoda u značajnom postotku. Odnosi razina prodaje omogućuju inteligentniju specijalizaciju ponude uz olakšano praćenje rezultata.

Ekspertno znanje osoba koje prate prodaju omogućilo je dodatnu kategorizaciju razina prodaje koja promatrana u vremenskim odsječcima u određenim situacijama snažno implicira na određenu zakonitost (npr. povećanje razine prodaje) odnosno pokazuje prediktivnu stranu modela. "

Iz navedenog teksta vidljiva je primjenjivost REFII modela u istraživanju tržišta tekstila. Ovaj poslovni slučaj samo potkrepljuje hipoteze o primjenjivosti REFII modela u tržišnim analizama, te o mogućnostima koje ovaj model pruža u povezivanju vremenskih serija sa metodama *data mininga*.

7. Ostala područja primjene REFII modela (idejni modeli i konceptualna rješenja)

S obzirom na to da je REFII model konstruiran sa ciljem šire primjenjivosti na strukture podataka koje pokazuju karakteristike vremenskih serija, ili pak djelomične karakteristike vremenskih serija, postoje i alternativni načini korištenja REFII modela gdje postoji sličnost struktura podataka sa strukturom podataka vremenske serije.

Tako se REFII model može primijeniti u tekst *miningu*, *web miningu*, medicinskim sustavima (EEG, EKG, mjerenje tlaka), prepoznavanje oblika u 3D prostoru.

Ključ za primjenu leži u činjenici da je svako od spomenutih područja potrebno promatrati u svjetlu vremenskih serija, odnosno REFII modela bilo u cjelini, bilo u kontekstu elementa korištenih kroz empirijska istraživanja.

Pri tome treba razlikovati primjenu REFII modela kao jedinstvenog modela transformacije tradicionalnih vremenskih serija i primjenu REFII modela u kontekstu primjene za potrebe tekst *mininga*, *web mininga*, EEG-a, EKG-a, prepoznavanja oblika u 3D prostoru jer u navedenim slučajevima nije riječ o analizi vremenskih serija, već uglavnom o simulacijama vremenskih serija.

Ove se simulacije provode s ciljem iskorištavanja potencijala koji su sastavni dio bilo REFII modela, bilo njegovih analitičkih potencijala prikazanima kroz empirijska istraživanja.

Za potrebe tekst *mininga*, tekst je potrebno promatrati kao vremensku seriju, gdje se svaka oznaka unutar teksta može pretvoriti u ASCII kod. Tekst prikazan kao niz ASCII kodova možemo promatrati kao vremensku seriju. Tako formiranu vremensku seriju moguće je dalje transformirati na način da se svaka riječ ili korijen riječi predstavi jedinstvenom numeričkom vrijednošću. Na tako formirani niz moguće je primijeniti algoritme poput algoritama za pretraživanje uzoraka, otkrivanja mutacija i slično, u svrhu otkrivanja sličnosti između tekstova. Alternativni pristup mogao bi se svesti na formiranje matrice transformacije (transformirati tako kodiranu vremensku seriju u REFII model) te nad matricom transformacije provoditi analize.

Za potrebe *Web mininga*, izvršene sesije posjetitelja možemo također predstaviti poput vremenskih serija. Ovdje se pojavljuje problem reprezentacije vrijednosti unutar vremenske serije, odnosno kako pristup pojedinoj stranici na site-u reprezentirati brojčanom oznakom koja će biti dovoljno reprezentativan za potrebe analize pomoću REFII modela.

Od do sada spomenutih područja alternativne primjene REFII modela, primjena u okviru medicine i medicinskih istraživanja (EEG, EKG, mjerenje tlaka) se čini na prvi pogled najbliža osnovnoj koncepciji REFII modela, jer spomenuti elementi su u osnovni vremenske serije.

Da bi se provela analiza prepoznavanja objekata u 3D prostoru, potrebno je REFII model prilagoditi za 3D prostor kao što je to konceptualno prikazano u poglavlju 2.8. Na taj način moguće je prepoznavati 3D oblike (površina zemlje - GIS sustavi, prepoznavanje lica, prepoznavanje objekata...).

Ovo poglavlje daje prikaz potencijalnih mogućnosti po pitanju provođenja analiza posredstvom REFII modela, ili njegovih metodoloških dijelova. Metodologija i područja koja se preliminarno obrazlažu u ovom poglavlju spadaju u kategoriju budućih istraživanja. S obzirom na kompleksnost svakog od obrađivanih područja, dane su osnovne smjernice kako svako od područja koje se obrađuje u okviru ovog poglavlja promatrati u svjetlu REFII modela ili njegovih metodoloških dijelova. Isto tako dane su smjernice za kreiranje potencijalnih modela rješenja za svako od ovih područja, što također spada u kategoriju budućih istraživanja.

Ovdje se žali ukazati na potencijale REFII modela i njegovih metodoloških dijelova, kao i na to da navedena područja nisu ujedeno i jedina područja gdje bi se navedena koncepcija analiza vremenskih serija mogla pokazati efikasnom.

S obzirom na specifičnosti spomenutih područja, posebice u kontekstu budućih istraživanja, predloženi modaliteti rješenja mogu se ravnopravno promatrati bilo kroz korištenje originalne koncepcije REFII modela, bilo kroz modalitete rješenja koji koriste pojedinačne elemente proizašle na temelju empirijskih istraživanja primjene REFII modela.

7.1. Tekst mining

Kao što je već spomenuto, temeljni preduvjet za promatranja tekstova u kontekstu vremenskih serija i tekst *mininga* je izražavanje znakova (brojčano -slovčanih oznaka) pomoću ASCII kodova. Ako primjerice imamo tekst u datoteci koji glasi: 'Goran test.',

poželjno ga je transformirati u niz sa malim ili sa velikim slovima, maknuti znakove interpunkcije i formirati listu na temelju tekstualne datoteke kao što je to prikazano na narednom primjeru u Python sintaksi:

```
lista= ['g','o','r','a','n',' ','t','e','s','t']
```

U sljedećem koraku potrebno je transformirati znakove u njihove ASCII vrijednosti na slijedeći način :

```
nova=[]  
for c in range(len(lista)):  
    print ord(lista[c])  
    nova.append(ord(lista[c]))
```

Kao rezultat obrade dobije se lista :

```
nova= [103, 111, 114, 97, 110, 32, 116, 101, 115, 116]
```

Daljnji tijek provođenja analize umnogome ovisi o vrsti jezika koji se koristi. Ideja je svaku riječ predstaviti sa odgovarajućom numeričkom vrijednošću. Kod jezika koji u sebi sadrže padeže preporučljivo je tu transformaciju provoditi na temelju korijena riječi i to prema principu sa slike 4.7.3. pri čemu se baza brojevnog sustava određuje se kao suma broja znakova u korištenoj abecedi i suma broja korištenih numeričkih oznaka (ako brojevi ulaze u postupak analize) sa specijalnim oznakama koje su relevantne s obzirom na vrstu analize.

Kao separator za kreiranje inicijalnih brojevnih uzoraka služi oznaka razmaka (' ') odnosno ASCII= 32.

Nakon provođenja ovog koraka imamo listu brojčanih oznaka koje predstavljaju riječi. Jedan od mogućih pravaca analize koji bi mogao biti predmet budućih istraživanja, je korištenje metodologije otkrivanja uzoraka, otkrivanja mutacija, otkrivanje događajnosti, otkrivanje cikličnosti direktno na tako formiranoj listi brojčanih vrijednosti koje predstavljaju riječi, bez posredstva jedinstvenog modela transformacije vremenske serije.

Drugi mogući pravac analize koji bi mogao biti predmet budućih istraživanja, je korištenje metodologije otkrivanja uzoraka, otkrivanja mutacija, otkrivanje događajnosti, otkrivanje cikličnosti na bazi REFI modela.

S obzirom na prirodu tekstualnih datoteka, na ovaj način bi se mogle istražiti mogućnosti otkrivanja rečeničnih uzoraka unutar tekstova, što može biti korisno prilikom procjenjivanja sličnosti među tekstovima na temelju hipoteze da su tekstovi slični ako u sebi sadrže visoku frekvenciju pojavnosti istih tekstualnih uzoraka.

Pri tome se može napraviti klasifikacija sličnosti s obzirom na dužine rečeničnih uzoraka jer što je veća dužina frekventnog uzorka to je i potencijalno veća sličnost između tekstova.

To znači da je dužina frekventnih uzoraka unutar tekstova proporcionalna visini stupnja pouzdanosti sličnosti među tekstovima, kao i broju frekventnih uzoraka. Hipotetska tablica za procjenu sličnosti tekstova može sadržavati sljedeće podatke:

Tablica 7.1. Hipotetska tablica za procjenu sličnosti tekstova

Dužina uzorka	Frekvencija pojavnosti uzorka	Očekivana sličnost
Mala (1-3)	Mala (do 0.3)	Mala
Srednja (3-7)	Srednja (0.7)	Srednja
Velika (>7)	Velika (>0.7)	Velika
...

U tablicu bi trebalo obuhvatiti sve moguće kombinacije pojavnosti između dužine uzoraka i frekvencije pojavnosti uzoraka sa procjenama očekivane sličnosti.

Daljnji postupci koji bi trebali razlučiti sličnost između tekstova mogli bi se temeljiti na proračunima cikličkih pojavnosti uzoraka teksta. Pretpostavka je ako se unutar N tekstova ciklički frekventno pojavljuju uzorci teksta, tada se dva teksta mogu smatrati sličnima.

Isto tako na temelju analize mutacija unutar teksta, odnosno procjene koji se frekventni rečenični uzorci međusobno variraju i na koji način (vidi sliku 6.4.) moguće je postaviti hipotezu o sličnosti tekstova.

Daljnji izazovi mogli bi se svesti na postavljanje modela uvjetnih vjerojatnosti korištenja rečeničnih uzoraka s obzirom na prethodno korištene rečenične uzorke, te procjena sličnosti na temelju konteksta.

Iako se tekstualne datoteke na prikazani način mogu izraziti kao vremenske serije, one ipak u sebi zadržavaju određene specifičnosti koje ih razlikuju od tradicionalnih vremenskih serija.

Te specifičnosti se očituju u činjenicama

- Svaka riječ ili korijen riječi predstavljen jedinstvenom brojčanom vrijednošću je relevantan uzorak na kojima se direktno mogu iskoristiti algoritmi primijenjeni u empirijskom dijelu istraživanja poput algoritama za otkrivanje uzoraka, mutacija, događajnosti..
- “Tekstualne vremenske serije” nisu u osnovi trendovski orijentirane strukture kao što su to klasične vremenske serije poput vremenskih serija podizanja gotovine sa bankomata
- Način na koji je provedena transformacija “tekstualnih vremenskih serija” u okviru ovog poglavlja može poslužiti kao jedinstveni model transformacije koji je kompatibilan za provođenje algoritama prikazanih u empirijskom dijelu istraživanja
- Na tako provedenom transformiranom strukturu tekstualnih datoteka, uspoređivanje uzoraka se može vršiti na temelju eksplicitne, a ne intervalne jednakosti

Ako se u obzir uzmu navedene činjenice, postavlja se logično pitanje o potrebitosti i nužnosti transformacije tekstualnih datoteka u REFII model prije provođenja analiza, što se može svrstati u kategoriju budućih istraživanja. Ova dilema proizlazi

prvenstveno iz prirode tekstualnih datoteka, koje se promatraju u kontekstu vremenskih serija, pri čemu nemaju sve karakteristike tradicionalnih vremenskih serija. Primjer iz ovog područja izabran je iz razloga:

- ilustracije kako neočite podatkovne strukture možemo reprezentirati vremenskim serijama
- prikaza kako priroda samih podataka (u ovom slučaju tekstualnih) ima utjecaja na shvaćanje problematike vremenskih serija
- prikaza kako priroda samih podataka (ako nisu tradicionalne vremenske serije) utječe na selekciju i metodologiju analize

7.2. Web mining

Područje u kojemu REFII model sa pripadajućom metodologijom može biti iskoristiv je i *Web mining*, i to područje analize "klikova" (eng. clickstreams).

Metodologiju REFII modela moguće je u domeni *Web mininga* iskoristiti u nekoliko osnovnih tipova analize koje se temelje na analizi "klikova" :

- Segmentacija posjetitelja Web stranica s obzirom na sličan model ponašanja prilikom posjete stranicama
- Prepoznavanje uzoraka ponašanja posjetitelja Web stranica
- Procjena ergonomske efikasnosti stranica

Da bi se postigli spomenuti analitički ciljevi, potrebno je efikasno pretprocesirati podatke sadržane u *Web log* datoteci. *Web log* datoteka sadrži informacije o svakoj pojedinoj *sesiji* koja predstavlja reprezentaciju svakog pojedinačnog posjeta stranicama.

Osim podataka o aktivnostima posjetitelja ("klikovima") ove datoteke mogu sadržavati i informacije da li je posjetitelj došao na promatranoj stranicu direktno, upisom Web adrese u preglednik, ili putem *bannera*, linka sa neke druge stranice, ili nekim drugim putem. Pojedini *Web log-ovi* bilježe i vrijeme između aktivacije dvaju linkova.

S obzirom na to da linkovi unutar *Web log-ova* nisu reprezentirani numeričkim vrijednostima, potrebno je izvršiti transformaciju vrijednosti za svaki aktivirani link u numeričke vrijednosti, kako bi se dobila simulacija vremenske serije za potrebe daljnjih analitičkih aktivnosti za potrebe daljnjih analiza.

S obzirom na prirodu REFII modela i strukturu *Web* stranica , predlaže se numerička transformacija gdje će tematska područja unutar stranica biti grupirana temeljem numeričkih vrijednosti. Primjerice, ako želimo analizirati *Web* stranicu distributera boja i lakova, te pribora i popratnog materijala za bojenje, tada numeričku transformaciju stranica možemo izvršiti prema hipotetskom obrascu:

Tablica 7.2. Sustav numeričke transformacije stranica

Najviša hijerarhijska instanica	Numerički klasifikator	Niža hijerarhijska instanica	Numerički klasifikator	Niža hijerarhijska instanica
Boje	1000000	Auto-lakovi	1010000	...
Boje	1000000	Boje za čamce i brodove	1020000	...
Boje	1000000	Zidne boje	1030000	...
Trake	2000000	Platnene	2010000	...
Trake	2000000	Plastične	2020000	...
Zaštitni pribor	3000000	Respiratornog sustava	3010000	...
Zaštitni pribor	3000000	Mehanička oštećenja	3020000	...
Zaštitni pribor	3000000	Kemijska oštećenja	3030000	...
...

Transformacija se u vremensku seriju vrši prema narednom obrascu:

Ako u *Web log* datoteci postoji informacija da je posjetitelj došao na stranicu koja je generalno vezana uz sve vrste boja, tada se u vremensku seriju upisuje vrijednost 1000000. Ako je nakon toga odlučio ući na stranicu koja se bavi zaštitnim priborom, tada se u vremensku seriju upisuje vrijednost 3000000, te ako je nakon toga birao stranicu koja prikazuje proizvode koje spadaju u zaštitna sredstva od mehaničkih oštećenja, tada se u vrijednost vremenske serije upisuje vrijednost 3020000, i tako redom.

Na ovaj način vidljivo je da su prema kriteriju sličnosti, a u duhu REFII modela dodijeljene približno sličnije numeričke vrijednosti prema tematici *Web* stranica. S obzirom na prirodu REFII modela, poželjno je i pridodijeljivati inicijalne vrijednosti najviših hijerarhijskih razina prema kriterijima sličnosti. Tako primjerice u kontekstu spomenutog primjera, boje i lakovi bi u obrascu imati "sličnije" (prema funkcijama udaljenosti) vrijednosti od primjerice boja i potrošnog materijala.

Ovakva transformacija se provodi iz razloga što na taj način prilikom analize vremenske serije pomoću REFII modela, postoje veći otkloni prilikom prelaska sa stranica sa različitom tematikom, što može imati značajan utjecaj na procjenu intervalnih jednakosti. Teoretski, se u istu intervalnu skupinu jednakosti mogu smjestiti sve kategorije koje spadaju u slična područja. Do koje će hijerarhijske instance biti modelirani kriteriji intervalnih jednakosti, ovisi o cilju analize. To se može ilustrirati primjerom, gdje se želi analizirati model ponašanja posjetitelja na način da je analitičar zainteresiran samo za model ponašanja na temelju glavnih kategorija, odnosno da li postoje određeni uzorci ponašanja posjetitelja, na razini zainteresiranosti za primjerice boje, trake, zaštitni pribor i slično. Na temelju ove analize koja je više generalnog karaktera na temelju otkrivanja uzoraka posredstvom REFII modela, metodologijom koja je opisana u poglavlju 6.1. mogu se dobiti uzorci ponašanja posjetitelja. Na temelju uzoraka moguće je pretpostaviti nizove kategorija iz asortimana predstavljenih na stranicama koje kategorije proizvoda su interesantne posjetiteljima i to u nizovima. Da li primjerice nakon posjeta stranica sa bojama, posjećuju stranice za lakove ili zaštitna sredstva, da li su skloniji ciljano posjećivati stranice upisom URL-a u preglednik, ili u većem broju slučajeva dolaze do određenih stranica preko *bannera*? Isto je tako moguće analizirati modele ponašanja frekventnijih posjetitelja, ili posjetitelja koji dolaze prvi puta na temelju ostavljenih *kolačića* (eng. cookies).

Daljnji postupci analize mogu se provoditi s ciljem segmentacije posjetitelja, s obzirom na slične modele ponašanja prema metodologiji prikazanoj u poglavlju 4.7.2..

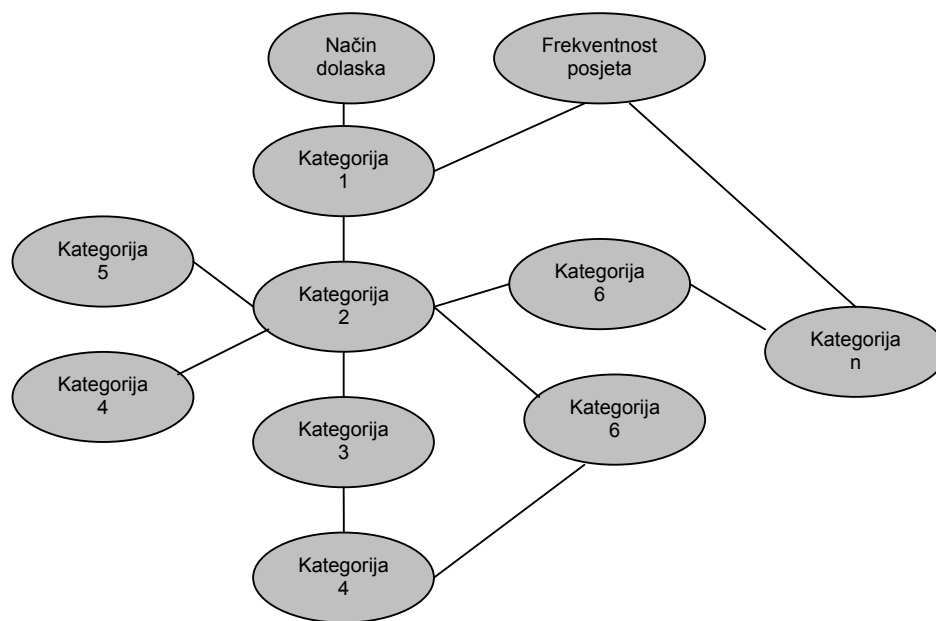
Principom izravnog otkrivanja pravila prikazanog u poglavlju 4.4.1 moguće je otkrivati pravilnosti na temelju niza posjeta kako bi se otkrile pravilnosti prilikom posjeta stranicama.

Problematiku procjene efikasnosti ergonomije stranica moguće je analizirati uz uvjet da su u *Web log* datoteci zabilježena vremena zadržavanja na stranici.

Na taj način primjerice možemo odrediti očekivana trajanja zadržavanja na određenim stranicama s obzirom na sadržaj. Tako bi primjerice navigacijske stranice trebale imati manje očekivano vrijeme zadržavanja od sadržajnih stranica. Ako se primijeti da se posjetitelj prosječno mnogo više vremena zadržava na navigacijskom stranicama od očekivanog vremena, to znači da navigacijske stranice nemaju zadovoljavajući stupanj ergonomske karakteristika, te ih je potrebno modificirati. Ista je stvar i primjerice sa stranicama za narudžbe. Uz pomoć ovih pokazatelja i analize, moguće je kreirati numeričke pokazatelje ergonomije koji bi izražavali prosječna odstupanja od očekivanih vrijednosti, te na temelju tih pokazatelja procjenjivati stupanj ergonomije stranica.

Daljnji modeli analize mogu se oslanjati na probabilističke koncepte poput Bayesovih mreža.

Primjer ovakvog modela dan je na slici integracije REFII modela i probabilističkog koncepta u *Web miningu*.



SLIKA 7.1. INTEGRACIJA REFII MODELA I PROBABLISTIČKOG KONCEPTA U *WEB MININGU*

Prikazani probabilistički model može primjerice davati odgovore na pitanja, kolika je vjerojatnost da će posjetitelj koji je na stranicu došao putem određenog *bannera* posjetiti stranicu koja se odnosi na boje (u kontekstu hipotetskog primjera). Svaki čvor prikazan na slici u sebi sadrži tablicu uvjetnih vjerojatnosti sukladno definiciji kategorija i potkategorija. Tako se na primjer kategorija 1 može odnositi na boju, a pripadajuća tablica uvjetnih vjerojatnosti za ovaj čvor kao elemente može sadržavati : Auto-lakove, boje za čamce i brodove, zidne boje i slično.

Temeljem ovog modela moguće je procjenjivati efikasnost *bannera*, jer se mogu procjenjivati vjerojatnosti posjećenosti onog dijela site-a na koji se *banner* odnosi, te se u odnos mogu staviti posjete putem *bannera* i posjete koje nisu inicirane sa *bannerom*. Isto tako u probabilistički model na temelju *kolačića* možemo uključiti i kategorije posjetitelja koji su barem jednom ili više puta (moguća je podijeljenost i u dobne razrede) posjetili određene stranice, te procjenjivati frekvenciju dinamičnosti sadržaja sa frekventnošću posjeta. Ova mjera također može biti korisna za siteove koji su portalski orijentirani.

Napredniji način kreiranja Bayesovih mreža za ovakve potrebe mogao bi se referencirati na rezultate analize proizašle kao rezultat pretraživanja uzoraka. Ovakav pristup daje informacije o frekventnim uzorcima, te već u samom startu sugerira oblike modela Bayesovih mreža.

Naravno, ovakav pristup ne mora biti pravilo, već se model može formirati na temelju ekspertize i područja interesa, te zadanim ciljevima analize koji mogu varirati s obzirom na vrijeme provođenja analize i okolnosti zbog kojih se provodi analiza.

Web mining u svjetlu REFII modela baziranog na vremenskim serijama obuhvaća nekoliko područja o kojima je bilo riječi kroz rad:

- Integraciju (simuliranih) vremenskih komponenti sa nevremenskim komponentama. (*clickstreams* i elementi poput *bannera*, ili upisa URL-a u preglednik)
- Integracija *data mining* metoda sa simuliranim vremenskim komponentama (Bayesove mreže)
- Moguće promatranje aktivnosti posjetitelja s obzirom na nove sadržaje i nove ponude na *Webu* (akcijska prodaja, ponuda novih artikala i slično)

7.3. Medicina

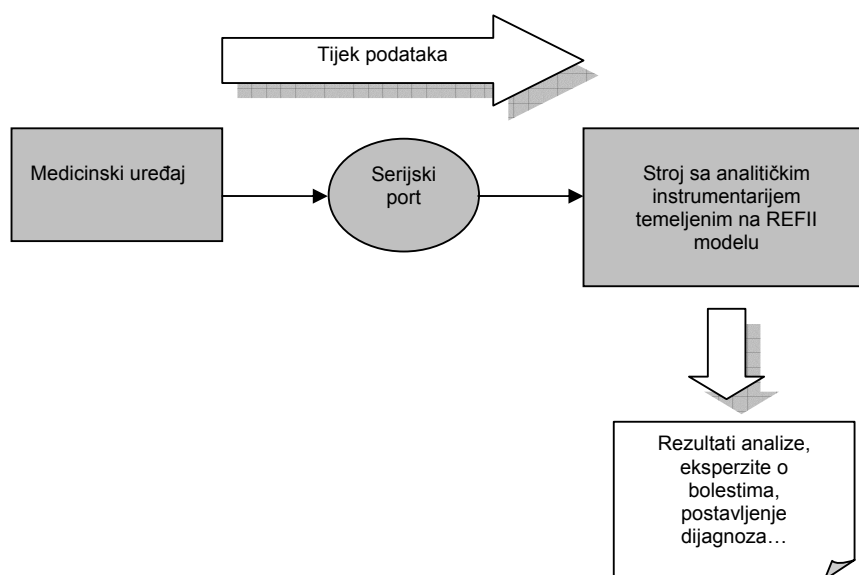
REFII model ima široki spektar primjene u okviru medicinskih sustava koji se baziraju na vremenskim serijama, poput uređaja za EEG, EKG, mjerenja krvnog tlaka i slično. Vrijednosti proizašle iz ovakvih vrsta mjerenja imaju sva obilježja vremenskih serija, što znači da nije potrebno provoditi dodatne transformacije vrijednosti da bi se simulirale vremenske serije, kao što je to slučaj sa tekst *miningom* i *web miningom*. Sama analiza tako proizašlih vremenskih serija može se provoditi na dva osnovna načina:

- Direktna analiza podataka na temelju izlaznih rezultata za svako pojedinačno mjerenje iz uređaja (EKG, EEG ...) gdje se odmah nakon provedenih mjerenja putem algoritama za otkrivanje uzoraka i motiva postavlja ekspertiza na temelju provedenih mjerenja
- Eksplorativna, koja vrši analizu na temelju n izlaznih rezultata, gdje svaki izlazni rezultat predstavlja mjerenje sa ciljem otkrivanja novih činjenica poput klasificiranja vremenskih serija s obzirom na sličnost prema određenim markerima (motivima u vremenskoj seriji) koji upućuju na određene devijacije ili upućuju na određene tipove poremećaja
- Analiza procjene uspješnosti određenih vrsta terapija, gdje se mogu ispitivati utjecaji određenih medikamenata na populacijama oboljelima od određenih tipova

bolesti čiji je trend i razvoj moguće pratiti putem EEG-a, EKG-a, i mjerenjima krvnog tlaka

Mogući način prikupljanja podataka iz uređaja za medicinska mjerenja može se postići putem serijskog porta za komunikaciju (COM porta), pri čemu svaki uređaj u tehničkoj specifikaciji sadrži uputstva za parametrizaciju za potrebe prikupljanja podataka.

Stroj na kojem se provode analize mora imati module za prikupljanje podataka putem serijskog porta, spremanje podataka u datoteku, te modul za analizu putem REFII modela. Shema opisanog sustava dana je na slici:



SLIKA 7.2. SHEMA SUSTAVA ZA PRIKUPLJANJE PODATAKA SA MEDICINSKIH UREĐAJA ZA POTREBE ANALIZA

Osim puke analize podataka, potrebno je skladištiti svako mjerenje za potrebe eksplorativne analize i otkrivanje skrivenih zakonitosti proizašlih na temelju vremenskih serija. Isto tako moguće je u takav sustav uključiti i rezultate mjerenja za potrebe analiza uspješnosti medikamenata, te se tako objedinjeni podaci i rezultati analize mogu objediniti u medicinski ekspertni sustav.

Takav ekspertni sustav može se kreirati prema konceptu prikazanom u poglavlju 5.7., te može bit baziran na *fuzzy* logici, i može obuhvaćati sustave pravila koja osim što obuhvaćaju znanje iz domene vremenskih serija kombiniraju i ekspertna znanja.

Na primjer ako se empirijski dokaže da testni lijek A utječe na sniženje krvnog tlaka, i povećanje određenih hormona u tijelu, potrebno je vezati utjecaj povišenja te vrste hormona na neku drugu bolest. Na taj način kada se unesu varijable o pacijentu u sustav, on će preporučiti određenu terapiju, uzimajući u obzir ne samo jednodimenzionalni pregled proizašao iz analize vremenskih serija i utjecaja testnog lijeka na razvoj bolesti, već će u obzir uzeti i ostale faktore rizika proizašlih na temelju podataka pacijentove dijagnoze.

Metodologija REFII modela koja se efikasno može iskoristiti u za analize u okviru ovih sustava je :

- Otkrivanje motiva
- Otkrivanje uzoraka
- Upitni jezik (TSSQL) na vremensku seriju

Za potrebe dijagnostike zasigurno bi bilo najiskoristivije otkrivanje motiva unutar vremenske serije, gdje se na temelju ekspertnog znanja unaprijed mogu definirati anomalije unutar provedenih mjerenja na temelju ekspertnog znanja. Pri tome, pojavnost određenog uzorka, niza uzoraka, ili pravilnosti unutar vremenske serije može upućivati na prisutnost određenog tipa bolesti.

Otkrivanje uzorka može biti korisno za potrebe otkrivanja zakonitosti unutar izvršenih mjerenja, te ovaj pristup može uvelike pomoći prilikom definicije ekspertnog znanja o pojavnosti uzoraka unutar vremenskih serija za određene tipove bolesti. Nadalje, ovaj pristup može biti koristan u području istraživanja rezultata proizašlih na temelju novih medicinskih uređaja koji kao izlazni rezultat bilježe vremenske serije, gdje još, zbog nedovoljne istraženosti nije definirano ekspertno znanje o pojavnosti uzoraka i određeni tip bolesti. Metodologijom otkrivanja uzoraka i uspoređivanja postavljene dijagnoze, moguće je nekom od metoda *data mininga* poput stabla odlučivanja dati smjernice za postavljenja dijagnoza po obrascu :

AKO uzorak A u seriji TADA bolest A (koeficijent pouzdanosti k)

AKO uzorak B u seriji TADA bolest B (koeficijent pouzdanosti k)

AKO uzorak A i uzorak B u seriji tada bolest C (koeficijent pouzdanosti k)

...

Cjelokupna analiza na višoj razini može biti provođena pomoću strukturiranog upitnog jezika za vremenske serije TSSQL [Perng, 2002], koji može objedinjavati niz metodoloških postupaka, i može se svesti na upite tipa: "Pronađi sve pacijente koji su imali uzorak A, a nisu imali uzorak B", ili "Pronađi sve pacijente koji su imali uzorak C i nisu imali dobru reakciju na medikament M".

Ovakav sustav morao bi obuhvatiti kako dio sa vremenskim serijama za potrebe generiranja rezultata upita tako i dio klasične baze podataka. Konceptom koji je predstavljen u okviru REFII modela moguće je ostvariti ovaj spoj. Tradicionalni pristup upitnih jezika na vremensku seriju oslanjaju se uglavnom na pretragu motiva i uzoraka unutar vremenskih serija. Postavljanjem REFII modela kao temelja sustava moguće je proširiti spektar mogućnosti upita, kao i proširiti analitiku na nevremenske attribute. Daljnje mogućnosti odnose se na korištenje *data mining* tehnika s ciljem otkrivanja pravilnosti vezanih uz oboljenja.

S obzirom na to da REFII model omogućuje direktnu primjenu *data mining* metoda nad vremenskim serijama, a s obzirom na prirodu medicinskog područja gdje se u fokus interesa stavljaju uzročno-posljedične veze, moguće je vršiti analize uvodeći jednu novu dimenziju u otkrivanju skrivenih pravilnosti.

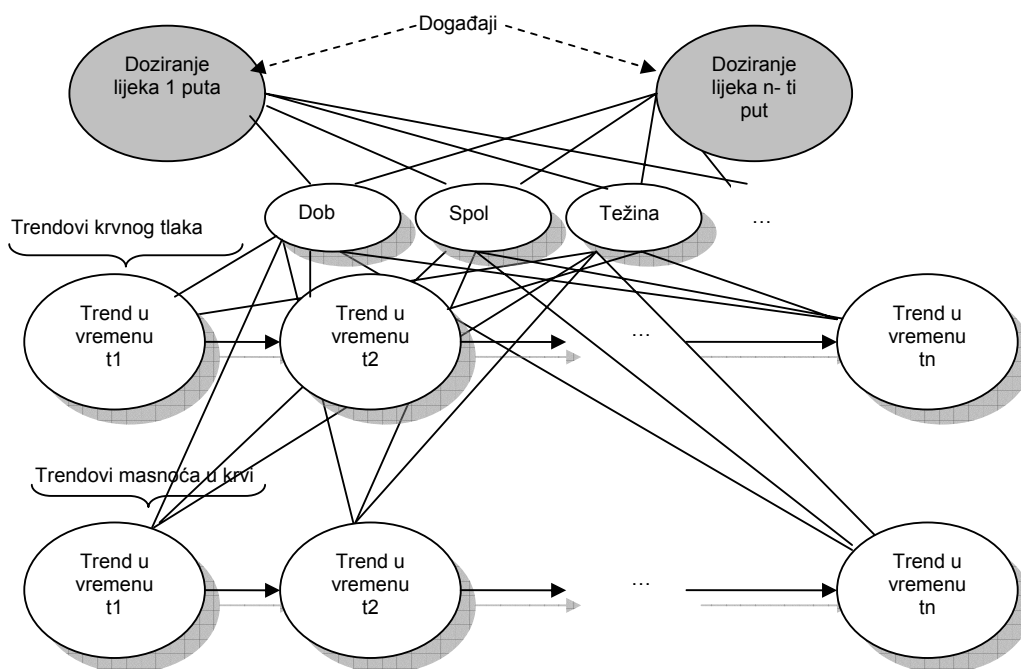
REFII model može biti vrlo koristan u okviru medicinskih istraživanja gdje se procjenjuje efikasnost djelovanja nekog medikamenta ili novog lijeka na razvoj/sprečavanje određene bolesti. Ako na primjer želimo testirati koji je od dva lijeka uspješniji za smanjivanje krvnog tlaka kod osoba sa povišenim krvnim tlakom, te primjerice istovremeno djelovati na smanjenje masnoća u krvi, te na niz drugih faktora

moguće je uz pomoć REFII modela i *data mining* tehnika postaviti model rješenja vezan uz ovu problematiku.

Takav model mogao bi se formirati na temelju promatranja niza pacijenata sa spomenutim bolestima. Vrijednosti krvnog tlaka i masnoća mogle bi se mjeriti kroz jednake vremenske razmake (mjerilo vremenske kompleksnosti) nakon što su bolesnici tretirani sa lijekovima koji se ispituju, te se u svakom tom vremenskom razmaku mogu bilježiti trendovi krvnog tlaka i masnoća u krvi. Model se može nadograditi na način ako se bolesnici lijekovima tretiraju više puta u tijeku perioda promatranja, te se tretiranje lijekovima može u kontekstu REFII modela izraziti kao događaj.

Cjelokupni izloženi model može se riješiti u okviru Bayesovih mreža, gdje bi svaki čvor predstavljao jedan promatrani vremenski razmak sa trendovima, a trendovi bi se mogli promatrati s obzirom na dob, spol, godinu života, težinu, visinu i ostale karakteristike pacijenta. Osim po spomenutim karakteristikama, glavni fokus interesa bio bi usredotočen na vrijeme uzimanja lijeka. Sve ove varijable mogu se izraziti kao elementi tablice uvjetnih vjerojatnosti, gdje bi ciljna varijabla za promatranje bila trend vrijednosti krvnog tlaka i masnoća, te otkrivanje koji je od dva lijeka efikasniji, i da li postoje karakteristike pacijenata s obzirom na određene karakteristike koji preferiraju jedan ili drugi lijek.

Generalni model rješenja ovog problema koji se bazira na Bayesovim mrežama i REFII modelu prikazan je slikom :



SLIKA 7.3. INTEGRACIJA REFII MODELA I BAYESOVIH MREŽA U SVRHU ISTRAŽIVANJA EFIKASNOSTI LJEKOVA

Cilj je u modelu pronaći trendove pada masnoća u krvi i krvnog tlaka, analizirati što je utjecalo na taj trend ? Da li primjerice uzimanje lijeka A ili B i nakon koliko puta, što se može definirati kao događaj? Da li su sve promatrane skupine s obzirom na promatrane karakteristike jednako reagirale na lijek A ili B, te ako su reagirale različito koje su karakteristike skupina koje su reagirale različito ? Da li je za to presudan dob, spol, godine života, težinu, visinu ili neko drugo obilježje?

Na temelju postavljenog modela možemo isfiltrirati sve skupine prema zadanim kriterijima koje su nakon primanja lijeka A i B jedan ili više puta pokazale :

- Padajuće trendove razine masnoća u krvi
- Padajuće trendove krvnog tlaka
- Padajuće trendove razine masnoća u krvi i krvnog tlaka
- Rastuće trendove razine masnoća u krvi
- Rastuće trendove krvnog tlaka
- Rastuće trendove razine masnoća u krvi i krvnog tlaka
- ...

Na temelju dobivenih rezultata moguće je procijeniti efikasnost utjecaja lijeka A i B na smanjivanje krvnog tlaka i/ili masnoće u krvi, te da li ta efikasnost varira s obzirom na karakteristike promatrane populacije po dobi, spolu, godinama života, težini, visini i ostalim karakteristikama pacijenata.

Ovo su smjernice kako je moguće koristiti REFII model u području medicinskih istraživanja, odnosno šire gledajući kako pomoću vremenskih serija potpomognutima *data mining* tehnikama možemo efikasnije vršiti istraživanja u domeni medicine.

Ova metodologija također može biti predmet budućih istraživanja na REFII modelu, te se može proširiti i na druga područja medicine koja se mogu poistovjetiti sa vremenskim serijama kao promatranim strukturama podataka.

7.4. "Profile" analize

"Profile" analize u marketinškim istraživanjima provode se kao uvodni postupak s ciljem temeljnog upoznavanja karakteristika klijenata, radi lakšeg definiranja ciljeva i narednih postupaka analize. Tradicionalne profile analize svode se na temeljnu analizu demografskih karakteristika klijenata, te osnovnih pokazatelja proizašlih iz prirode analize.

Rezultati ovakve vrste analiza često puta služe kao ulazni parametri, odnosno smjernice za provođenje složenijih oblika segmentacije klijenata.

U slučajevima kada s obzirom na cilj analize postoji podatak, vezan uz svakog klijenta, koji se odnosi na pojave sadržane unutar vremenskih serija, tada u profile analize možemo uključiti i elemente analize proizašle iz vremenskih serija.

Za razliku od tradicionalnih pristupa, gdje se primjerice promatraju dobni razredi korisnika proizvoda X , primjenom REFII modela moguće je napraviti precizniju analizu, koja se primjerice odnosi na dijagnostiku dobnih skupina , korisnike proizvoda X koji pokazuju veću sklonost kupnji u periodu završetka radnog tjedna.

Na ovaj način mogu se dobiti precizniji pokazatelji za provođenje segmentacije posebice ako su za segmentaciju i planiranje kampanja važni trendovski pokazatelji. Na taj način na mikro – razini u trgovačkim centrima moguće je organizirati promocije

novih proizvoda u periodima kada se pojavljuju kupci obilježja koji odgovaraju ciljnoj skupini potencijalnih potrošača te vrste proizvoda.

Na višoj razini moguće je planirati medijske objave neposredno prije očekivanja posjećenosti trgovačkih centara od strane kupaca čija obilježja odgovaraju ciljnoj skupini potencijalnih potrošača.

Napredniji oblik *profile* analiza odnosio bi se na analizu osjetljivosti primjerice na određenu kampanju prema osnovnim obilježjima populacije klijenata. Da li su primjerice nakon provedene kampanje trend potražnje od prodaje ostvarili kupci sa točno određenim karakteristikama i kojim. Koja segment klijenata je stagnirao u potražnji nakon kampanje i slično.

Ovakav pristup moguće je provesti i u tijeku procjene uspješnosti kampanja prije samog nastupa na tržištu, gdje se mogu formirati fokus grupe i analiziraju učinci nekoliko različitih pristupa kampanjama.

Na ovaj način mogu se dobiti informacije, koja strategija oglašavanja će polučiti najveći uspjeh s obzirom na ciljnu skupinu kojoj je proizvod namijenjen, te se ovim putem povećava vjerojatnost uspješnosti kampanje.

Također je vrlo važno pratiti učinke kampanji, i upravo na način gdje se ispituju kritični faktori i obilježja segmenata koji su pokazali najveći stupanj osjetljivosti na kampanju (trend rasta) potražnje nakon provedene kampanje, moguće je alocirati resurse iz potencijalno neprofitabilnih segmenata u potencijalno profitabilne segmenta s ciljem maksimizacije profita.

U slučaju da se pojedini segmenti izdvoje kao neosjetljivi, *campaign manager* može donijeti odluku o novoj segmentaciji i traženu uzroka toj pojavi. Problem nastaje ako je analizom ustanovljeno da je cjelokupni ciljni segment neosjetljiv na samu kampanju. Da bi se izbjegao ovakav problem, koji generira velike troškove, provode se spomenute analize osjetljivosti radi reduciranja ove vrste rizika.

REFII model i vremenske serije daju novu dimenziju u domeni marketinških istraživanja, te uvode i trendovsku komponentu u procese segmentacije tržišta.

Ako cjelokupnu problematiku promatramo i sa perspektive primjene *data mining* algoritama kako na vremenskim serijama, tako i na tradicionalnim nevremenskim atributima, kao što je to prikazano u poglavlju 6.7., tada je segmentaciju primjerice moguće izvršiti primjenom metoda klasteriranja, stabla odlučivanja, gdje ulazni parametri za klasteriranje mogu biti nevremenski atributi (kategorije dobi, spol, regije...), te vremenski atributi (trendovi pojava, vremenski indeksi). Na temelju ovog modela moguće je istražiti trendove primjerice potražnje za nekim artiklom s obzirom na vremenske indekse (dani u tjednu, mjeseci, godine) s obzirom na obilježja dobi, spola regija, ili nekih drugih obilježja kao što je to ilustrativno prikazano na empirijskim podacima u poglavlju 6.7.3.

Metode *data mininga* su same po sebi donijele novi način provođenja marketinških analiza. Integracija REFII modela u analitičke instrumentarij marketinških analiza otvara jednu novu dimenziju u analize gdje pojave vezane vremenskom komponentom sudjeluju ravnopravno u procesima analize poput tradicionalnih nevremenskih atributa. S tim u skladu moguće je osim tradicionalnim *data mining* metodama ovako kreirane podatkovne strukture analizirati i metodama procjene relevantnosti atributa [Han, 2001] u postupcima pretprocesiranja podataka.

Na ovaj način u samim začetima analize moguće je procijeniti koliko su značajne trendovske komponente u provođenju daljnjih aktivnosti analize. Kako je zadatak pretprocesiranja podataka redukcija osnovnog seta atributa stavljanjem u fokus naglasaka na relevantne attribute, u ovom koraku je moguće testirati važnost

trendovskih komponenti u modelima, te ovisno o rezultatima analize relevantnosti trendovskih atributa ulaziti u daljnje analitičke aktivnosti sa tim atributima.

Sa perspektive nomenklature, vrlo je teško povući oštru granicu gdje prestaje *profile analiza*. Generalno gledajući, sve aktivnosti vezane uz temeljno upoznavanje populacije (klijenata, proizvoda, tržišta) spada u domenu *profile analiza*.

Potencijali REFII modela mogu doći do izražaja kako u periodu provođenja *profile analiza*, tako i tijekom provođenja narednih koraka analiza, kao što je to prikazano na empirijskim podacima u 6. poglavlju.

7.5. Trodimenzionalni prostori i REFII model

Kao što je već prikazano u poglavlju 2.8., temeljna koncepcija REFII modela može se primijeniti i na 3D prostor i to za prepoznavanje objekata u prostoru koji leže na ravnini. Ovu metodologiju moguće bi bilo primijeniti za potrebe prepoznavanja lica, površine na zemlji (GIS sustavi). Uvjet za provođenje ovih vrsta analiza je reprezentacija trodimenzionalnih oblika (lica, površine zemlje) u žičanom modelu.

Na tako formiranom žičanom modelu, potrebno je prilagoditi REFII model kako bi zadovoljio zahtjeve vezane uz treću dimenziju, kako je to konceptualno prikazano u poglavlju 2.8. Na osnovu tog modela, potrebno je definirati pojmove intervalnih jednakosti, definirati pojmove sličnosti, i što se podrazumijeva pod pojmom sličnosti vezano uz 3d model kao što je to učinjeno za REFII model u 3. poglavlju koji opisuju pojamovne strukture.

S obzirom na 3D prostor elementi koje promatramo sastavljeni su od “krnjih” kvadrova, a ne od “krnjih” pravokutnika kao u 2D prostoru. Za razliku od 2D prostora gdje površinu ispod krivulje računamo metodom pravokutnika, u 3D prostoru trebali bismo koristiti monte carlo metodu integracije za određivanje površine ispod 3D krivulje.

Sukladno tome potrebno bi bilo prilagoditi algoritme za prepoznavanje uzoraka u 3D prostoru, mutacije, te definirati pojam događajnosti. Pojam događajnosti proizlazi iz potreba vezanih uz trodimenzionalne prostore, gdje uz komponentu vremena i varijable koja se pojavljuje u dvije dimenzije možemo uvesti i drugu varijablu, te vršiti analizu s obzirom na vrijeme i dvije varijable.

Ako želimo uvesti još jednu varijablu u model te izaći iz 3D prostora u n dimenzionalni prostor, putem ove metodologije nailazimo na probleme, zbog vezanosti REFII modela na geometrijsku interpretaciju.

U svakom slučaju kao i većina područja navedenih u okviru ovog poglavlja ovo područje spada u kategoriju predmeta budućih istraživanja, a navedene ideje mogu poslužiti kao smjernice za daljnje korake istraživanja.

8. Zaključak

REFII model konstruiran je s ciljem objedinjavanja različitih koncepcija analize vremenskih serija, tradicionalnih metoda rudarenja podataka te s ciljem konstrukcije novih algoritamskih postupaka iz oblasti analize tržišta, te pronalaženju rješenja problema automatskog pretprocesiranja vremenskih serija.

Važnost problema razjedinjenosti metodoloških postupaka analiza vremenskih serija uočio je Graham Williams u svom radu [Williams, 2002] koji daje pregled *data mining* metoda u domeni vremenskih serija. U istom radu naveo je izazove budućih istraživanja (eng. challenge questions) gdje bi trebalo riješiti problem generalne teorije analize vremenskih serija koja bi objedinila sadašnja i buduća istraživanja na području *data mining* analize vremenskih serija.

Osnovni nedostatak spomenutog Williamsovog modela proizlazi iz činjenice što je premala pažnja usmjerena na model transformacije, te je on stavljen u drugi plan prilikom definicije rješenja.

Kao implikacija ovakvog zapostavljanja modela transformacije pojavljuju se problemi:

- Svi problemi proizlaze zbog nedovoljno dorađenog i nefleksibilnog modela transformacije vremenske serije koji se u hodu dorađuje s obzirom na analitičke potrebe [Williams, 2000], [Williams, 2001], [Williams, 2001], [Williams, 2001a], [Williams, 2002a], [Williams, 2003] , [Williams, 2003a]
- Zanemarivanje kontingencijskog pristupa analizi
- Nije predviđena metodologija povezivanje temporalnih atributa sa netemporalnim
- Kao aktivnosti budućih istraživanja napominje se prilagođavanje i dorada postojećih *data mining* algoritama, umjesto da se putem modela transformacije izgrade mostovi za spajanje, što naravno implicira i generalni problem sa primjenom *fuzzy* logike unutar modela

REFII model ima odgovore na probleme koji su navedeni kao rezultati budućih istraživanja u Williamsovom radu, te nudi puno fleksibilniji pristup analizi vremenskih serija od one koju je zamislio spomenuti autor.

Cilj istraživanja je bio istražiti mogućnosti nadogradnje REFII modela, kao i njegovih mogućnosti sintetiziranja postojećih algoritama rudarenja podataka i analize vremenskih serija u jedinstven sustav analize tržišnih zakonitosti i segmentacije tržišta na temelju vremenskih serija.

Kroz rad su dani modaliteti nadogradnje REFII modela, te njegova sposobnost sintetiziranja postojećih metodoloških postupaka analize vremenskih serija u jedinstveni sustav *data mining* analize vremenskih serija, koji unutar objedinjava, *data mining* tehnike, otkrivanje uzoraka, traženje motiva, traženje mutacija unutar algoritama, otkrivanje sezonskih oscilacija i traženje događajnosti.

Svi ovi postupci izvršeni su na istom uzorku podataka, temeljem jedinstvenog modela transformacije vremenske serije.

S tim u skladu ponuđena su rješenja za različite modalitete segmentacije tržišta i otkrivanja tržišnih zakonitosti (modeli ponašanja tržišnih segmenata i subjekata) na

temelju vremenskih serija, posredstvom REFII modela na uzorku podataka baze štetnika na temelju automobilske police osiguranja.

Dokazane su hipoteze :

1: Jedinstveni model transformacije vremenske serije (REFII) omogućuje povezivanje različitih konceptualnih modela analize vremenske serije, što je unapređenje u odnosu na tradicionalan način korištenja niza nepovezanih metoda prilikom analize.

Hipoteza je dokazana na temelju provedbe niza analitičkih postupaka na uzorku podataka baze štetnika, automobilske police osiguranja, gdje su rađene analize otkrivanje uzoraka, pronalaženje motiva, otkrivanja mutacija, analize sezonskih oscilacija, stabla odlučivanja i klasteriranja posredstvom jedinstvenog modela transformacije vremenske serije, omogućivši provođenje detaljnije analize u svjetlu ulančavanja ovih metodoloških postupaka.

2: REFII model također omogućuje analize vremenskih serija primjenom tradicionalnih metoda rudarenja podataka (stabla odlučivanja, klasteriranja, metode potrošačke košarice,...)

Hipoteza je dokazana na primjerima primjene stabla odlučivanja, klasteriranja i primjene asocijativnih algoritama na empirijskim podacima, te je pokazan način kako ovakve postupke analize možemo ulančavati sa tradicionalnim vrstama analiza, poput otkrivanja uzoraka i slično.

3: Povezivanjem različitih konceptualnih modela analize u okviru REFII modela možemo riješiti složene analitičke zadatke na temelju vremenskih serija, poput segmentacije tržišta te otkrivanja tržišnih zakonitosti (modeli ponašanja tržišnih segmenata i subjekata) .

U radu je prikazana sinergija primjene informacija proizašla iz rezultata analiza nakon primjene niza ulančanih analitičkih postupaka gdje su se na uzorku podataka štetnika, korisnika polica auto – odgovornosti otkrile zakonitosti :

Ako se u obzir uzmu otkriveni frekventni uzorci : "Nizak pad-Nizak rast" i dobivene rezultate po pitanju sezonskih oscilacija, lako je za pretpostaviti da je ovaj uzorak ima jaki sezonski utjecaj i to u periodu od nedjelje do utorka s obzirom rezultate proizašle iz analize sezonskih oscilacija.

Također se u obzir mogu uzeti otkriveni događaji "Nizak pad-Nizak rast- Nizak rast " nakon kojeg se može očekivati događaj "Nizak rast-Nizak pad- Nizak pad ", a nakon njega događaj "Nizak pad-Nizak pad- Nizak rast".

Temeljem rezultata analize gdje su primijenjene metoda OLAP-a , te stabla odlučivanja, može se pretpostaviti (ako se promatra Zagrebačka regija) da će u periodima trenda niskog rasta biti povećana vjerojatnost nesreća od strane vozila koje imaju Zagrebačku registraciju. U periodu trenda niskog rasta ugroženija je dobna skupina od 46-55 godina, muškog spola.

Na ovaj se način mogu raditi segmentacije, te promatrati ili cijeli uzorak, ili primjerice Varaždinska regije, tražiti uzorci i sezonske oscilacije samo za tu regiju, ili samo za

određenu skupinu i slično. Isto tako prikazanom metodologijom moguće je provoditi procjene rizičnosti kako tržišnih segmenata u određenom vremenskom periodu, tako i segmentaciju s obzirom na unaprijed zadane kriterije.

4: Primjena REFII modela nudi unapređenje u otkrivanju tržišnih zakonitosti iz vremenskih serija prilikom "ad hoc" analize tržišnih problema za razliku od tradicionalnoga pristupa, prilikom kojega koristimo niz nepovezanih i nekompatibilnih metoda, koje nam ponekad ne dopuštaju mogućnost dublje analize i modeliranje rješenja za nestandardne probleme povezivanjem niza različitih metodoloških postupaka analize.

Unapređenje pojmovno znači povezivanje različitih metodoloških koncepcija analize vremenskih serija, primjenu tradicionalnih metoda rudarenja podataka na vremensku seriju, ulančavanje metoda za analizu vremenskih serija te nadogradnja temeljnog REFII modela s izvornim algoritmima s ciljem izgradnje modela za segmentaciju tržišta na osnovi vremenskih serija i procjene ponašanja tržišnih segmenata i subjekata.

Dublja analiza pojmovno znači mogućnost provedbe različitih tipova analiza na reduciranom setu podataka iz vremenske serije, koji je proizašao kao rezultat obrade prije primijenjene metode ili analitičkoga postupka.

Povezivanje se ostvaruje posredstvom jedinstvenoga modela transformacije vremenske serije, koji je temelj analitičkim postupcima, metodama te razvoju novih metoda s obzirom na problemski prostor.

Hipoteza je dokazana kroz empirijska istraživanja, primjerice, otkriveni frekventni uzorci poslužili su kao input za provedbu analize događajnosti i provedbu analize otkrivanje mutacije unutar uzoraka, te su pronađeni uzorci promatrani u kontekstu sezonskih oscilacija s ciljem pronalaženje utjecaja sezonskih oscilacija na pojavnost uzoraka.

Uz povezivanje tradicionalnih do sada nepovezanih metoda za analizu vremenskih serija, postignuto je i povezivanje vremenskih serija sa metodama *data mininga*. Temelj svih aktivnosti povezivanja metodoloških postupaka, unapređenja i dublje analize je upravo jedinstveni model transformacije vremenske serije.

Kako je efikasnost modela definirana kao sposobnost povezivanja različitih analitičkih koncepcija, bilo izvornih bilo tradicionalnih, s ciljem provođenja složenih analiza vremenskih serija u području istraživanja tržišnih zakonitosti i segmentacije tržišta, na temelju provedenih empirijskih istraživanja dokazana je efikasnost modela.

Znanstveni doprinos očituje se u predstavljanju jedne nove koncepcije u analizi vremenskih serija u području istraživanja i segmentacije tržišta, koja, za razliku od tradicionalnoga pristupa ovoj problematici, u žarište stavlja model jednoznačne transformacije vremenske serije.

Prednost ovakvoga pristupa očituje se u sintezi niza različitih pristupa i metoda u analizi vremenskih serija, u koje spadaju i tradicionalne metode rudarenja podataka, kao i nadogradnja novim algoritamskim postupcima analize.

Ovakav pristup omogućuje provedbu različitih tipova analiza na reduciranom setu podataka iz vremenske serije, koji je proizašao kao rezultat obrade već primijenjene metode ili analitičkoga postupka.

Ovakav pristup daje odgovore na to kako provoditi kompleksne tržišne analize temeljene na vremenskim serijama koje je tradicionalnim pristupom vrlo teško ili gotovo nemoguće izvesti.

Ove mogućnosti su do sada tradicionalnim pristupima analizi vremenskih serija zbog primjene niza nepovezanih metoda (nepostojanja generalne strategije povezivanja) bile nemoguće ili vrlo teško izvesti, a pojavljuju se kao nužnost u analizi tržišnih zakonitosti putem vremenskih serija.

Sažetak

Rad izlaže novi pristup u analizi vremenskih serija, koji se oslanja na jedinstveni model transformacije vremenske serije. Prednost ovakvog pristupa očituje se u mogućnosti integracije čitavog niza analitičkih postupaka nad vremenskom serijom, mogućnost ulančavanja postupaka analize nad vremenskom serijom, te integraciju metoda *data mininga* unutar vremenskih serija, što dosadašnjom poznatom metodologijom nije bio slučaj, upravo iz razloga što se zanemaruje važnost modela transformacije. Dosadašnji znanstveni radovi upućuju na taj problem i uglavnom rješavanje ovog problema uvjetuju razvojem moćnije hardverske podrške.

Prof. Williams [Williams, 2002] u svom znanstvenom radu također upućuje na problem nepovezanosti i metodologije analiza vremenskih serija, te definira koncept nazvan *general hidden distribution based analysis theory for temporal data mining*.

Osnovni nedostatak spomenutog Williamsovog modela proizlazi iz činjenice što je premala pažnja usmjerena na model transformacije, te je on stavljen u drugi plan prilikom definicije rješenja.

Premala je pažnja usmjerena na fleksibilnost prilikom analize (fleksibilno kreiranje razreda otklona od strane analitičara), te se ovaj problem predstavlja kao predmet budućih istraživanja. Predloženo rješenje također zanemaruje kvantitativne aspekte vremenskih serija (površinu ispod krivulje).

Zbog toga su najavljena buduća istraživanja usmjerena na povećanje fleksibilnosti, iznalaženje rješenja kako vremensku seriju povezati direktno sa tradicionalnim *data mining* metodama, te je izražen problem povezivanja vremenske serije sa *fuzzy* logikom unutar Williamsovog predloženog rješenja, i ovaj se problem spominje kao predmet budućih istraživanja. Paradoksalno je to što za primjerice povezivanje vremenske serije sa tradicionalnim *data mining* metodama, Williams namjerava prilagođavati tradicionalne *data mining* algoritme bez razmišljanja o izgradnji jedinstvenog modela transformacije vremenske serije.

Glavna kritika Williamsove koncepcije odnosi se na :

- Svi problemi proizlaze zbog nedovoljno dorađenog i nefleksibilnog modela transformacije vremenske serije koji se u hodu dorađuje s obzirom na analitičke potrebe [Williams, 2000], [Williams, 2001], [Williams, 2001], [Williams, 2001a], [Williams, 2002a], [Williams, 2003] , [Williams, 2003a]
- Zanemarivanje kontingencijskog pristupa analizi
- Nije predviđena metodologija povezivanje temporalnih atributa sa netemporalnim
- Kao aktivnosti budućih istraživanja napominje se prilagođavanje i dorada postojećih *data mining* algoritama, umjesto da se putem modela transformacije izgrade mostovi za spajanje, što naravno implicira i generalni problem sa primjenom *fuzzy* logike unutar modela

REFII model ima odgovore na probleme koji su navedeni kao rezultati budućih istraživanja u Williamsovom radu, te nudi puno fleksibilniji pristup analizi vremenskih serija od one koju je zamislio spomenuti autor. Williams upada u klasičnu zamku prilikom pokušaja objedinjavanja metoda, a to je fokusiranje na rješavanje parcijalnih

analitičkih zadataka, pri čemu ne postoji univerzalno razrađeni model transformacije vremenske serije, već sakuplja dijelova niza parcijalnih rješenja koje se pokušava objediniti u konačni mozaik, posredstvom nedovoljno razrađenog modela transformacije koji bi trebao imati centralno mjesto u povezivanju različitih koncepcija. REFII model na način na koji je prikazan daje cjeloviti rješenje za analizu vremenskih serija, te je posebno pogodan za rješavanja problema iz domene tržišnih analiza koje zahtijevaju integraciju i ulančavanje čitavog niza postupaka i metoda za analizu vremenskih serija.

Rad se sastoji iz osam poglavlja.

Prvo poglavlje rada pod naslovom Uvod definira cilj i svrhu rada, prikazuje do sada najznačajnija istraživanja iz područja vremenskih serija, trenutnu metodologiju analiza vremenskih serija te ukratko daje osnovni koncept REFII modela kao i njegove prednosti u odnosu na dosadašnju praksu vezanu uz analize vremenskih serija u otkrivanju tržišnih zakonitosti.

Drugo poglavlje rada donosi detaljan opis REFII modela s temeljitom razradom njegovih komponenti REF podmodela., proračuna površine ispod krivulje te proračuna nagiba pravca. U ovom poglavlju prikazan je REFII model u cjelini, kao i metodologija transformacije vrijednosti proizašlih iz REFII modela u strukture čitljive algoritmima. Kao ilustracija u okviru ovoga poglavlja, prikazana je trodimenzionalna interpretacija REFII modela.

Treće poglavlje opisuje osnovne pojmovne strukture vezane uz REFII model. Ovo poglavlje donosi pojmovnu definiciju jednakosti vremenskih odsječaka, inverzne oscilacije vremenskih odsječaka, njihove korelacije i sličnosti. Ove pojmovne definicije ključ su za konstrukciju algoritama koji se oslanjaju na REFII model.

Četvrto poglavlje donosi metodologiju konstrukcije algoritama za različita područja analize vremenskih serija. U sklopu ilustracije ove metodologije ovo poglavlje prikazuje kako konstruirati algoritme za otkrivanje znanja na temelju vremenske serije, odnosno kako iskoristiti postojeće metode za otkrivanje znanja u REFII modelu.

Peto poglavlje donosi metodologiju primjene REFII modela od pretprocesiranja izvornih podataka, procesiranja u REFII sintaksu, principe integracije različitih koncepcija i modela analize do interpretacije i primjene rezultata analize.

Šesto poglavlje je empirijski orijentirano te daje rješenje za problematiku otkrivanja uzoraka, traženje događajnosti, traženje mutacija, traženje sezonskih oscilacija, integraciju metoda *data mininga* sa vremenskim serijama.

Sedmo poglavlje prikazuje REFII model kao model koji se može uz analizu vremenskih serija koristiti i u područjima kao što je tekst *mining*, grafičko modeliranje, profile analiza te za medicinska istraživanja. Ovo poglavlje je ujedno i prikaz mogućih smjerova budućih istraživanja.

Osmo poglavlje - zaključno poglavlje

Životopis

Goran Klepac rođen je u Zagrebu 10. lipnja 1972. godine.
Završio je srednju školu OC Nikola Tesla, smjer: prirodoslovno-matematički tehničar.

Dodiplomski studij upisuje 1991. godine na Ekonomskom fakultetu u Zagrebu te završava smjer poslovne informatike gdje diplomira na temi *"Primjena umjetne inteligencije kao potpora poslovnom odlučivanju"*.

Poslijediplomski studij "Informatički menadžment" upisuje 1998. godine te magistrira na temi *"Prepoznavanje tržišnih zakonitosti iz perspektive poduzeća primjenom metoda umjetne inteligencije"*.

Prva iskustva u informatici stekao je već u periodu osnovne škole kada uči prve programske jezike BASIC, Assembler Z80, Assembler 6502.

Nakon upisa u srednju školu OC "Nikola Tesla", u tijeku školovanja počinje raditi kao programer te taj angažman kombinira s redovitim školskim aktivnostima.

Prvi projekt u koji je bio uključen je izrada kompjutorskoga programa "Percepcija I" , "Percepcija II" za potrebe centra za rehabilitaciju djece s poteškoćama u razvoju "Goljak", koji je služio kao pomoćno sredstvo u procesu rehabilitacije. Osnovna namjena toga projekta bila je pomoć u rehabilitaciji kod problema vizualne percepcije djece s poteškoćama u razvoju. Projekt je rađen u suradnji s prof. Stanec, tada djelatnicom toga centra.

Daljnji njegov profesionalan angažman u tijeku srednjoškolskoga obrazovanja odnosi se na razvoj aplikativnih rješenja za administrativne poslove u programskom jeziku CLIPPER za poduzeće KONČAR u Sessvetskom Kraljevcu.

Tijekom srednje škole uči programske jezike C++, ASSEMBLER 20x86 te honorarno razvija systemske aplikacije, kao što su antivirusni programi te utilitarni systemski programi pod DOS operativnim sustavom.

Nakon upisa na Ekonomski fakultet u Zagrebu, počinje honorarno raditi za poduzeće ZEER, gdje isprva radi kao programer te postepeno napreduje do projektanta informacijskih sustava. U sklopu ovoga poduzeća sudjeluje kao projektant informacijskih sustava u projektima razvoja informacijskih sustava u poduzećima, od kojih valja spomenuti CHROMOS KUTRILIN, AMERICAR, AUTOSERVIS SAVSKA.

Zadnji period njegova studiranja obilježen je honorarnim radom u poduzeću INFOBIT, gdje radi u svojstvu projektanta informacijskih sustava te sudjeluje na projektima razvoja istih, od kojih treba izdvojiti : JANKOMIR ROBNi TERMINALI, ŽITNJAK ROBNi TERMINALI, INDUSTROGRADNJA, TEP.

U periodu školovanja na Ekonomskom fakultetu uči niz novih programskih jezika: Visual Fox pro, Cobol, Prolog, GPSS.

Istodobno u praksi te u sklopu redovitoga školskog programa usvaja znanja iz oblasti računovodstva menadžmenta i informatike, što će utjecati kasnije na područje primjene stečenih informatičkih znanja.

Od 1996. godine radi u poduzeću PROFIT PP, koje je za razliku od prijašnjih poduzeća, uz djelatnost razvoja i projektiranja informacijskih sustava orijentirano i na razvoj i implementaciju sustava za podršku odlučivanju.

U sklopu ovoga poduzeća, uz poslove razvoja informacijskih sustava i vođenja programerskih timova, vodi te aktivno sudjeluje u projektima vezanima uz razvoj sustava za podršku odlučivanju i analizu podataka.

Od projekata koje je vodio i sudjelovao u sklopu ovoga poduzeća, a vezani su uz razvoj informacijskih sustava, razvoj sustava za podršku odlučivanju i analizu podataka, valja spomenuti: METAL-ELKTRO, zastupništvo SIKKENSEA za Hrvatsku, zastupnike SONAXA za Hrvatsku, PRIMAX te niz hotelijersko-ugostiteljskih i proizvodnih poduzeća na području cijele Hrvatske.

Ovo iskustvo na području sustava za podršku odlučivanju, analize te rudarenja podataka, motiviralo ga je da upiše poslijediplomski studij - Informatički menadžment 2001. godine.

2001. godine zapošljava se u Raiffeisen Banci kao voditelj odjela analize i istraživanja tržišta, gdje prvenstveno primjenjuje metode rudarenja podataka na transakcijske baze podataka s ciljem otkrivanja tržišnih trendova i zakonitosti.

Od 2004. godine radi kao direktor BI direkcije u Raiffeisen consultingu.

U sklopu svoga posla u Raiffeisen Banci a sad u Raiffeisen consultingu, koristi analitičke alate i metode (SPSS, Answer tree, Decision time, Bayesian networks, Neural networks, Fuzzy logic,...), te samostalno razvija softver u Pythonu i Visual Fox Pro za pretprocesiranje podataka, kao i za programiranje vlastitih analitičkih metoda i modela. Isto tako razvija modele za kompleksne tržišne analize, te osmišljava strategije razvoja kompleksnih analitičkih modela i vodi BI projekte.

Model predstavljen u okviru disertacije uspješno koristi za realizaciju analitičkih projekata unutar Raiffeisen banke u sklopu Raiffeisen consultinga.

Govori tečno engleski jezik te ga koristi u svakodnevnom poslu u radu s internacionalnim timovima.

Stručno usavršavanje

2001 Aris training

2001 Slovakia , "Internet and data warehousing workshop", u organizaciji Raiffeisen Zentral Bank

2004 Seminar menađerskih vještina : "Učinkovito delegiranje i interna komunikacija", RBA

2004 Seminar menađerskih vještina : "Nematerijalna motivacija", RBA

Održana predavanja na stručnim skupovima

"Segmentacija tržišta na temelju analiza vremenskih serija, Business intelligence konferencija, 10-11 veljače 2003, HLD

"Poslovna inteligencija", IDC BI Roadshow, Zagreb, 2004.

"Data mining i tržišne analize ", Prime, Zagrebački velesajam 2004.

"Business intelligence- perspectives ", SAS forum adriatic region ,Opatija, 30-31. 3 2004.

"Otkrivanje znanja iz vremenskih serija", HROUG,Umag, 2004, rad objavljen zborniku radova.

Radovi na međunarodnim konferencijama

Goran Klepac (2002.): *REFII model-Model for recognition patterns in time series*, 20th International conference METHODOLOGY AND STATISTIC, University of Ljubljana, Faculty of social sciences, center of methodology and in informatics, Ljubljana, september 15.-18., 2002; Program and abstracts, str 53.-55.

Objavljene knjige

Klepac, Goran: *Primjena inteligentnih računalnih metoda u menadžmentu*, Sinergija, Zagreb, 2001., ISBN: 953-6895-01-3

Klepac Goran; Panian Željko: *Poslovna inteligencija*, Masmedia, Zagreb, 2003, ISBN: 953-157-447-2

Sudjelovanje na znanstveno-istraživačkim projektima financiranim od strane ministarstva znanosti

Projekt broj 067003, "Modeliranje i simulacija u poslovnoj ekonomiji" - glavni istraživač prof. dr. Vlatko Čerić (1996.-2000.) (<http://bib.irb.hr/>)

Projekt broj 0067016, "Metode i modeli potpore odlučivanju" - glavni istraživač prof. dr. Vlatko Čerić (2002.-) (<http://bib.irb.hr/>)

Referentna literatura

[Aleksander, 1995] Aleksander, I.; Morton H. *An introduction to neural computing*, International Thompson Computer Press, 1995.

[Agrawal, 1994] Agrawal R., Srikant R. *Fast Algorithms for mining association rules*. Proc. 20 th VLDB .487-499, 1994

[Agrawal, 1995] Agrawal R., Srikant R. *Mining sequential patterns* . Pros IEEE ICDE, 3-14, March 1995

[Agrawal, 1996] Agrawal R., Srikant R. *Mining sequential :generalization and performance improvements*. Proc. 5th EDBT,3-17,1996

[Apostols, 1996] Apostolos-Paul Refenes, *Neural networks in capital markets*, John Wiley & Sons, Ltd., 1996.

[Bozkaya, 1997] Bozkaya, T., Yazdani, N. & Ozsoyoglu, Z. M. (1997). *Matching and indexing sequences of different lengths*. In proceedings of the 6th Int'l Conference on Information and Knowledge Management. Las Vegas, NV, Nov 10-14. pp 128-135.

[Berry, 1997] Berry, J.A. Michael ; Linoff Gordon: *Data mining techniques for marketing sales and customer support*, John Wiley & Sons Inc, 1997.

[Brazma, 1998] Brazma A., Jonassen I., Vilo J. , Ukkonen E. *Pattern discovery in biosequences*. Lecture notes in Artificial Intelligence, 1433 ,256-270, 1998

[Caraça-Valente, 2000] Caraça-Valente, J. P. & Lopez-Chavarrias, I. (2000). *Discovering similar patterns in time series*. In proceedings of the 6th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data mining. Boston, MA, Aug 20-23. pp 497-505.

[Chabris,1987] Chabris Christopher F. *Artificial intelligence & Turbo Pascal*, Dow Jones- Irvin , 1987

[Chen, 2003] Chen, L., Ozsu, T. & Oria, V. (2003). *Symbolic Representation and Retrieval of Moving Object Trajectories*. Technical Report CS-2003-30. University of Waterloo.

[Cheng, 1997] Cheng-Jian Lin, *SISO nonlinear system identification using a fuzzy-neural hybrid system*, International journal of neural systems, Vol. 8, No 3 (June,1997) 325-337

[Chiu, 2003] Chiu, B., Keogh, E. & Lonardi, S. (2003). *Probabilistic Discovery of Time Series Motifs*. In the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 24 - 27. Washington, DC, USA. pp 493-498.

[Chu, 1999] Chu, K. & Wong, M. (1999). *Fast time-series searching with scaling and shifting*. In proceedings of the 18th ACM Symposium on Principles of Database Systems. Philadelphia, PA, May 31-Jun 2. pp 237-248.

[Craven, 1997] Craven, Mark W. *Understanding time series networks: a case study in rule extraction*, International journal of neural systems, Vol. 8, No 4 (August,1997) 373-384

[Das, 1997] Das, G., Gunopulos, D. & Mannila, H. (1997). *Finding similar time series*. In proceedings of Principles of Data Mining and Knowledge Discovery, 1st European Symposium. Trondheim, Norway, Jun 24-27. pp 88-100.

[Debregeas, 1998] Debregeas, A. & Hebrail, G. (1998). *Interactive interpretation of kohonen maps applied to curves*. In proceedings of the 4th Int'l Conference of Knowledge Discovery and Data Mining. New York, NY, Aug 27-31. pp 179-183.

[Dougherty, 1988] Dougherty, R. Edward; Giardina R. Charles *Mathematical methods for artificial intelligence and autonomous systems*, Prentice-Hall, 1988

[Faloutsos, 1997] Faloutsos, C., Jagadish, H., Mendelzon, A. & Milo, T. (1997). *A signature technique for similarity-based queries*. In proceedings of the Int'l Conference on Compression and Complexity of Sequences. Positano-Salerno, Italy, Jun 11-13.

[Fanchi, 2000] Fanchi, John ; *Math Refresher for Scientists and Engineers*, 2nd Edition, Wiley-IEEE Press; 2 edition (May 11, 2000)

[Freeman, 1991] Freeman, A. James; Skapura M. David : *Neural networks , algorithms, applications and programming techniques*. Addison-Wesley Publishing Company, 1991.

[Ge, 2000] Ge, X. & Smyth, P. (2000). *Deformable markov model templates for time-series pattern matching*. In proceedings of the 6th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining. Boston, MA, Aug 20-23. pp 81-90.

[Geurts, 2001] Geurts, P. (2001). *Pattern extraction for time series classification*. In proceedings of Principles of Data Mining and Knowledge Discovery, 5th European Conference. Freiburg, Germany, Sept 3-5. pp 115-127.

[Guralnik, 1999] Guralnik, V. & Srivastava, J. (1999). *Event detection from time series data*. In proceedings of the 5th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining. San Diego, CA, Aug 15-18. pp 33-42.

[Han, 2000] Han Jiawei, Pei J., Yin j. *Mining frequent patterns without candidate generation*, Proc. ACM SIGMOID, 1-12, 2000

[Han, 2001] Han Jiawei, Kamber Micheline, *Data mining-concepts and techniques*, Morgan Kaufmann publishers, 2001

[Han, 2002] Han Jiawei, Wang Wei, Yu S. Philips, Yang Jiong (2002). *Mining long patterns in a noisy environment*, ACM SIGMOID June 2002, Madison USA

[Han, 2003] Han Jiawei, Xifeng Yan, Ashfar Jamin (2003). *CloSpan: Mining closed sequential patterns in large datasets*. NSF IIS-02-09199, University of Illinois

[Han, 2004] Han, J. Pei, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach", IEEE Transactions on Knowledge and Data Engineering, 16(10), 2004.

[Huang, 1999] Huang, Y. & Yu, P. S. (1999). *Adaptive query processing for time-series data*. In proceedings of the 5th Int'l Conference on Knowledge Discovery and Data Mining. San Diego, CA, Aug 15-18. pp 282-286.

[Indyk, 2000] Indyk, P., Koudas, N. & Muthukrishnan, S. (2000). *Identifying representative trends in massive time series data sets using sketches*. In proceedings of the 26th Int'l Conference on Very Large Data Bases. Cairo, Egypt, Sept 10-14. pp 363-372.

[Jain, 1999] A. K. Jain, M. N. Murty, and P. J. Flynn, Data clustering: A review, ACM Computing Surveys, vol.31, no. 3, pp. 264-323, 1999.

[Javor, 1988] Javor Petar, *Uvod u matematičku analizu*, Školska knjiga- Zagreb, 1988

[Jensen, 2001] Jensen V. Finn, *Bayesian networks and decision graph*, Springer, NY, 2001

[Kennedy, 2001] Kennedy L. Rub, Van Roy Benjamin, Reed D. Christopher, Lippman P.

Richard. *Solving data mining problems through pattern recognition*, Prentice Hall, New Jersey, 1997.

[Keogh, 1997] Keogh, E. & Smyth, P. (1997). *A probabilistic approach to fast pattern matching in time series databases*. In proceedings of the 3rd Int'l Conference on Knowledge Discovery and Data Mining. Newport Beach, CA, Aug 14-17. pp 24-20.

[Keogh, 1998] Keogh, E. & Pazzani, M. (1998). *An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback*.

In proceedings of the 4th Int'l Conference on Knowledge Discovery and Data Mining. New York, NY, Aug 27-31. pp 239-241.

[Keogh, 2001] Keogh Eamon, Chu Selina, Hart David, Pazzani Michael, *An online algorithm for segmenting time series*, IEEE, 2001

[Kim, 2000] Kim, E., Lam, J. M. & Han, J. (2000). *AIM: approximate intelligent matching for time series data*. In proceedings of Data Warehousing and Knowledge Discovery, 2nd Int'l Conference. London, UK, Sep 4-6. pp 347-357.

[Klepac, 2000] Klepac Goran, *Otkrivanje zakonitosti iz perspektive poduzeća primjenom metoda umjetne inteligencije*, Ekonomski fakultet Zagreb, magistarski rad, 2000

[Klepac, 2001] Klepac Goran, *Primjena inteligentnih računalnih metoda u menedžmentu*, Sinergija, 2001

[Klepac, 2003] Klepac Goran, Panian Željko, *Poslovna inteligencija*, Masmedia, 2003

[Kliček, 2002] Kliček B., Zekić-Sušac M., *A Nonlinear Strategy of Selecting NN Architectures for Stock Return Predictions, Finance*, Proceedings from the 50th Anniversary Financial Conference Svishtov, Bulgaria, 11-12 April, 2002., ABAGAR, Veliko Tarnovo, 2002, pp. 325-355.

[Kohonen, 2001] Kohonen T. *Self-organizing maps*, Springer 2001, 3th edition

[Korn, 1997] Korn, F., Jagadish, H. & Faloutsos, C. (1997). *Efficiently supporting ad hoc queries in large datasets of time sequences*. In proceedings of the ACM SIGMOD Int'l Conference on Management of Data. Tucson, AZ, May 13-15. pp 289-300.

[Lam, 1998] Lam, S. K. & Wong, M. H. (1998). *A fast projection algorithm for sequence data searching*. *Data & Knowledge Engineering*, Vol. 28(3). pp 321-339.

[Lavrenko, 2000] Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D. & Allan, J. (2000). *Mining of concurrent text and time series*. In proceedings of the 6th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining Workshop on Text Mining. Boston, MA, Aug 20-23. pp 37-44.

[Lee, 2000] Lee, S., Chun, S., Kim, D., Lee, J. & Chung, C. (2000). *Similarity search for multidimensional data sequences*. In proceedings of the 16th Int'l Conference on Data Engineering. San Diego, CA, Feb 28-Mar 3. pp 599-608.

[Lin, 2001] Lin, J. Keogh, E. Patel, Lonardi S. (2001) *Finding motifs in time series*, 1st workshop on temporal data mining at 7th ACM SIGKDD International conference of knowledge discovery and data mining July 27-30 ,2001 Edmont Alberta Canada

[Lin, 2002] Lin, J. Keogh, E. Patel, Lonardi S. (2002) *Clustering of time series subsequences in meanings: Implications for previous and future research*, 2nd workshop on temporal data mining at 8th ACM SIGKDD International conference of knowledge discovery and data mining July 23-26 ,2002 Edmont Alberta Canada

[Lin, 2003] Lin, J., Keogh, E., Lonardi, S. & Chiu, B. (2003). *A Symbolic Representation of Time Series, with Implications for Streaming Algorithms*. In proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA. June 13.

[Manilla, 1997] Heikki Mannila, Hanu Toivonen, Verkamo Inkeri , *Discovery of frequent episodes in event sequences*, University of Helsinki Finland, Report C-1997-15

[Manilla, 2001] Heikki Mannila, Gunopulos Dimitrios, Das Gautam, *Finding similar time series*, Technical Report D-2001-4, University of Helsinki Finland

[Mardešić, 1977] Mardešić Sibe, *Matematička analiza I*, Školska Knjiga, 1977

[Ohsaki, 2003] Ohsaki, M., Sato, Y., Yokoi, H. & Yamaguchi, T. (2003). *A Rule Discovery Support System for Sequential Medical Data* In the Case Study of a Chronic Hepatitis Dataset. ECML 2003.

[Ozsu, 2003] Ozsu T.& Chen, L. (2003). *Multi-Scale Histograms for Answering Queries over Time Series Data*. In proceedings of the 20th International Conference on Data Engineering. Mar 30 - Apr 2. Boston, MA.

[Park, 1999] Park, S., Lee, D. & Chu, W. W. (1999). *Fast retrieval of similar subsequences in long sequence databases*. In proceedings of the 3rd IEEE Knowledge and Data Engineering Exchange Workshop. Chicago, IL, Nov 7.

[Park, 2001] Park, S., Kim, S. & Chu, W. W. (2001). *Segment-based approach for subsequence searches in sequence databases*. In proceedings of the 16th ACM Symposium on Applied Computing. Las Vegas, NV, Mar 11-14. pp 248-252.

[Patel, 2002] Patel, P., Keogh, E., Lin, J., & Lonardi, S. (2002). *Mining Motifs in Massive Time Series Databases*. In proceedings of the 2002 IEEE International Conference on Data Mining. Maebashi City, Japan. Dec 9-12.

[Perng, 2002] Perng C. S. , Parker D.S., *SQL/LPP:a time series extension of SQL based on limited patience patterns*, Technical report 980034 UCLA , Computer Science

[Polly, 2001] Polly, W. P. M. & Wong, M. H. (2001). *Efficient and robust feature extraction and pattern matching of time series by a lattice structure*. In proceedings of the 10th ACM CIKM Int'l Conference on Information and Knowledge Management. Atlanta, GA, Nov 5-10. pp 271-278.

[Popivanov, 2002] Popivanov, I. & Miller, R. J. (2002). *Similarity search over time series data using wavelets*. In proceedings of the 18th Int'l Conference on Data Engineering. San Jose, CA, Feb 26-Mar 1. to appear.

[Pratt, 2001] Pratt Kevin, *Locating patterns in discrete time series*, University of south Florida, M.sc these, 2001

[Pyle, 2001] Pyle Dorian , *Data preparation for data mining*, Morgan Kaufmann publishers, Inc, 1999.

[Radić, 1979] Radić Mirko, *Algebra II dio linearna algebra*, Školska knjiga, Zagreb 1979

[Rafiei, 1998] Rafiei, D. & Mendelzon, A. O. (1998). *Efficient retrieval of similar time sequences using dft*. In proceedings of the 5th Int'l Conference on Foundations of Data Organization and Algorithms. Kobe, Japan, Nov 12-13.

[Sardi, 2001] Sardi R , *Optimiation of sequence queries in database systems*, Ph.D Thesis , UCLA 2001

[Shapiro, 1986] Shapiro, E; Sterling, L. : *The art of prolog*, MIT Press, 1986

[Shatkay, 1998] Shatkay, H. & Zdonik, S. (1996). *Approximate queries and representations for large data sequences*. In proceedings of the 12th IEEE Int'l Conference on Data Engineering. New Orleans, LA, Feb 26-Mar 1. pp 536-545.

[Somayajulu , 2003] Somayajulu G. Sripada, Ehud Reiter, Jim Hunter and Jin Yu (2003). *Generating English Summaries of Time Series Data using the Gricean Maxims*. In the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 24 - 27. Washington, DC, USA.

[Šošić, 1990] Šošić Ivan, Serdar Vladimir, *Uvod u statistiku*. Školska knjiga-Zagreb 1990

[Tanaka, 2003] Tanaka, Y. & Uehara, K. (2003). *Discover Motifs in Multi Dimensional Time-Series Using the Principal Component Analysis and the MDL Principle*. In proceedings of the 3rd International Conference on Machine Learning and Data Mining in Pattern Recognition. pp.252-265.

[Taylor, 1996] Taylor, J.G. *Neural networks and their applications*, John Wiley & Sons Inc, 1996.

[Wang, 2001] Wang W., Yang J. Yu P. *Mining long sequential patterns in a noisy environment*. IBM research report 2001

[Westphal, 1998] Westphal Christopher, Blaxton Teresa *Data mining solutions – methods and tools for solving real world problems*, John Wiley & sons inc. 1998

[Williams, 2000] Williams, J. Graham, Weiqiang Lin, Mehmet Orgun, *Temporal Data Mining using Multi-Level Local Polynomial Models*, Proceedings of the 2nd International Conference on Intelligent Data Engineering and Automated Learning (IDEAL00)Lecture Notes in Computer Science, Volume 1983, Springer Hong Kong, December 2000

[Williams, 2001] Williams, J. Graham, Weiqian Lin, Mehmet Orgun, *Temporal Data Mining Using Hidden Markov-Local Polynomial Models* Edited by David Cheung, Graham Williams, Qing Li, Lecture Notes in Artificial Intelligence, Volume 2035, Springer, April 2001 Proceedings of the 5th Pacific Asia Conference on Knowledge Discovery and Data Mining(PAKDD01)Hong Kong, China, April 2001

[Williams, 2001a] Williams, J. Graham, Rohan Baxter, Graham Williams, Hongxing He, *Feature Selection for Temporal Health Records*, Advances in Knowledge Discovery and Data Mining Edited by David Cheung, Graham Williams, Qing Li Lecture Notes in Artificial Intelligence, Volume 2035, Springer, April 2001 Proceedings of the 5th Pacific Asia conference on Knowledge Discovery and Data Mining (PAKDD01) Hong Kong, China, April 2001

[Williams, 2002] Williams, J. Graham , Weiqiang Lin, Mehmet A Orgun, *An Overview of Temporal Data Mining*, Proceedings of the 1st Australian Data Mining Workshop (ADM02) Canberra, Australia, December 2002 Edited by Simeon J. Simoff, Graham J. Williams, Markus Hegland Published by University of Technology, Sydney, Pages 83-90, ISBN 0-9750075-0-5

[Williams, 2002a] Williams, J. Graham, Weiqiang Lin, Mehmet Orgun, *Mining Temporal Patterns from Health Care Data* , Proceedings of the 4th International Conference on Data Warehousing and Knowledge Discovery (DaWaK02) Lecture Notes in Computer Science, Vol 2454, Springer, 2002 Pages 221-231, ISBN 3-540-44123-9

[Williams, 2003] Williams, J. Graham, *Mining the Data Stream*, Invited Plenary, International Conference on Hybrid Intelligent Systems Melbourne, Australia, December 2003.

[Williams, 2003a] Williams, J. Graham, Chris Kelman, Rohan Baxter, Lifang.Gu, Simon Hawkins, Hongxing He, Chris Rainsford, Deanne Vickers, *Temporal Event Mining of Linked Medical Claims Data* , Proceedings of the PAKDD03 Workshop on Data Mining for Actionable Knowledge DMAK-2003 Seoul, Korea, April 2003

[Wang, 2000] Wang, C. & Wang, X. S. (2000). *Supporting content-based searches on time series via approximation*. In proceedings of the 12th Int'l Conference on Scientific and Statistical Database Management. Berlin, Germany, Jul 26-28. pp 69-81.

[Yi, 1998] Yi, B., Jagadish, H. & Faloutsos, C. (1998). *Efficient retrieval of similar time sequences under time warping*. In proceedings of the 14th Int'l Conference on Data Engineering. Orlando, FL, Feb 23-27. pp 201-208.

**Otkrivanje zakonitosti temeljem
jedinstvenoga modela transformacije vremenske serije**

G. Klepac

Fakultet organizacije i informatike, Varaždin, Hrvatska

Doktorska disertacija izlaže novi pristup u analizi vremenskih serija, koji se oslanja na jedinstveni model transformacije vremenske serije (REFII model). Model se sastoji od tri osnovna elementa (smjer nagiba pravca, koeficijent kutnog odklona odsječka, površine ispod odsječka), pomoću kojih se jednoznačno opisuje krivulja vremenske serije, u obliku matrice transformacije.

Na empirijskim je podacima pokazano, kako se pomoću REFII modela mogu otkriti uzorci iz vremenske serije, otkriti događaji, tražiti motivi, otkriti mutacije u uzorcima, otkriti sezonske oscilacije, te ulančavati analitički postupci.

Na setu empirijskih podataka, također je dokazano, da REFII model omogućava direktnu primjenu *data mining* metoda nad vremenskim serijama

Ilustracija konstrukcije novih algoritamskih rješenja pomoću REFII modela, prikazana je na primjeru otkrivanja uzoraka, sezonskih oscilacija, motiva i mutacija.

Prednost REFII modela u odnosu na tradicionalni pristup *data mininga* vremenskih serija, proizlazi iz činjenice, što on omogućava ulančavanje analitičkih postupaka nad vremenskim serijama, otvara vrata direktnoj primjeni tradicionalnih *data mining* metoda nad vremenskim serijama, te omogućava laku konstrukciju algoritamskih rješenja, za nestandardne probleme vezane uz domenu analize vremenskih serija.

Disertacija REFII modelom daje cjelovito rješenje za analizu vremenskih serija, posebno pogodnim za rješavanja problema iz domene tržišnih analiza, gdje do punog izražaja dolaze prednosti REFII modela u odnosu na tradicionalni pristup *data mininga* vremenskih serija.

Rad nije objavljen.

Voditelj rada: Prof. dr. sc. Božidar Kliček

Povjerenstvo za ocjenu : . Prof. dr. sc. Božidar Kliček

Povjerenstvo za obranu : Prof. dr. sc. Božidar Kliček

Datum obrane :

Datum promocije:

Rad je pohranjen na Fakultetu organizacije i informatike Varaždin

(211. stranica, 56 tablica, 62 slike, 86 referenci, original na hrvatskom jeziku)

G. Klepac

DD(FOI)

Tekući broj: _____

UDK 004.8(043.2)

Otkrivanje zakonitosti temeljem jedinstvenog modela
transformacije vremenske serije

I. G. Klepac

II. Fakultet organizacije i informatike, Varaždin, Hrvatska

Data mining (rudarenje
podataka)

Otkrivanje znanja

Poslovna inteligencija

Strojno učenje

Umjetna inteligencija

Vremenske serije

Doctoral dissertation

Time series rule discovering with unique model of transformation

G. Klepac

Faculty of Organisation and Informatics, Varaždin, Republic of Croatia

The doctoral dissertation introduces the REFII model, as a new approach in time series data mining.

The REFII model consists of three basic elements; curve segment trend mark, curve segment coefficient and area beneath curve segment. These three elements compose the matrix of transformation, which is a base for further analysis.

The dissertation shows the methodology of pattern recognition, event discovering, motif discovering, mutation recognition, seasonal oscillation recognition from time series with the REFII model, and analytical method chaining with REFII model.

The dissertation also shows the methodology of direct usage of data mining method on time series empirical data in REFII model and the methodology of constructing new algorithms for non-standard problems with use of REFII model.

The construction of new algorithms for non-standard problems using the REFII model is illustrated in pattern discovering, mutation recognition, motif discovering and event discovering.

The advantage of REFII model in comparison with traditional time series mining is a possibility of analytical method chaining, direct usage of traditional data mining methods on time series, and constructing of new algorithms for non-standard problems in domain of time series data mining.

The doctoral dissertation with REFII model gives a holistic solution for time series analysis, especially for problems in domain of market analysis, where advantage of REFII model gives completely new approach in times series data mining.

The thesis was not published.

Supervisor : Prof. dr. sc. Božidar Kliček

Apointed members for evaluation disertation : . Prof. dr. sc. Božidar Kliček

Apointed members for oral examination : Prof. dr. sc. Božidar Kliček

Oral examination :
Degree conferred :

This thesis is deposited at the Faculty of Organisation and Informatics, Varaždin.

(pages 211., tables 56, pictures 62, references 86, original in Croatian language)

G. Klepac

DD(FOI)
Number : _____

UDC 004.8(043.2)

Time series rule discovering with unique model of
transformation

I. G. Klepac

Artificial intelligence
Business intelligence
Data mining
Machine learning
Rule discovering
Time series

II. Faculty of Organisation and Informatics, Varaždin

Prilozi

Fuzzy definicija jednakosti odsječaka

IF				THEN	
Kutni_otklon1	Kutni_otklon2	Povrsina1	Povrsina2	DoS	Jednakost
Ostar_pad	Ostar_pad	Mala	Mala	1.00	Jednako
Ostar_pad	Ostar_pad	Mala	Srednja	1.00	Nejednako
Ostar_pad	Ostar_pad	Mala	Velika	1.00	Nejednako
Ostar_pad	Ostar_pad	Srednja	Mala	1.00	Nejednako
Ostar_pad	Ostar_pad	Srednja	Srednja	1.00	Jednako
Ostar_pad	Ostar_pad	Srednja	Velika	1.00	Nejednako
Ostar_pad	Ostar_pad	Velika	Mala	1.00	Nejednako
Ostar_pad	Ostar_pad	Velika	Srednja	1.00	Nejednako
Ostar_pad	Ostar_pad	Velika	Velika	1.00	Jednako
Ostar_pad	Srednji_pad	Mala	Mala	1.00	Nejednako
Ostar_pad	Srednji_pad	Mala	Srednja	1.00	Nejednako
Ostar_pad	Srednji_pad	Mala	Velika	1.00	Nejednako
Ostar_pad	Srednji_pad	Srednja	Mala	1.00	Nejednako
Ostar_pad	Srednji_pad	Srednja	Srednja	1.00	Nejednako
Ostar_pad	Srednji_pad	Srednja	Velika	1.00	Nejednako
Ostar_pad	Srednji_pad	Velika	Mala	1.00	Nejednako
Ostar_pad	Srednji_pad	Velika	Srednja	1.00	Nejednako
Ostar_pad	Srednji_pad	Velika	Velika	1.00	Nejednako
Ostar_pad	Blagi_pad	Mala	Mala	1.00	Nejednako
Ostar_pad	Blagi_pad	Mala	Srednja	1.00	Nejednako
Ostar_pad	Blagi_pad	Mala	Velika	1.00	Nejednako
Ostar_pad	Blagi_pad	Srednja	Mala	1.00	Nejednako
Ostar_pad	Blagi_pad	Srednja	Srednja	1.00	Nejednako
Ostar_pad	Blagi_pad	Srednja	Velika	1.00	Nejednako
Ostar_pad	Blagi_pad	Velika	Mala	1.00	Nejednako
Ostar_pad	Blagi_pad	Velika	Srednja	1.00	Nejednako
Ostar_pad	Blagi_pad	Velika	Velika	1.00	Nejednako
Ostar_pad	Jednako	Mala	Mala	1.00	Nejednako
Ostar_pad	Jednako	Mala	Srednja	1.00	Nejednako
Ostar_pad	Jednako	Mala	Velika	1.00	Nejednako
Ostar_pad	Jednako	Srednja	Mala	1.00	Nejednako
Ostar_pad	Jednako	Srednja	Srednja	1.00	Nejednako
Ostar_pad	Jednako	Srednja	Velika	1.00	Nejednako
Ostar_pad	Jednako	Velika	Mala	1.00	Nejednako
Ostar_pad	Jednako	Velika	Srednja	1.00	Nejednako
Ostar_pad	Jednako	Velika	Velika	1.00	Nejednako
Ostar_pad	Blagi_rast	Mala	Mala	1.00	Nejednako
Ostar_pad	Blagi_rast	Mala	Srednja	1.00	Nejednako
Ostar_pad	Blagi_rast	Mala	Velika	1.00	Nejednako
Ostar_pad	Blagi_rast	Srednja	Mala	1.00	Nejednako
Ostar_pad	Blagi_rast	Srednja	Srednja	1.00	Nejednako
Ostar_pad	Blagi_rast	Srednja	Velika	1.00	Nejednako
Ostar_pad	Blagi_rast	Velika	Mala	1.00	Nejednako
Ostar_pad	Blagi_rast	Velika	Srednja	1.00	Nejednako
Ostar_pad	Blagi_rast	Velika	Velika	1.00	Nejednako
Ostar_pad	Srednji_rast	Mala	Mala	1.00	Nejednako
Ostar_pad	Srednji_rast	Mala	Srednja	1.00	Nejednako
Ostar_pad	Srednji_rast	Mala	Velika	1.00	Nejednako
Ostar_pad	Srednji_rast	Srednja	Mala	1.00	Nejednako

IF				THEN	
Ostar_pad	Srednji_rast	Srednja	Srednja	1.00	Nejednako
Ostar_pad	Srednji_rast	Srednja	Velika	1.00	Nejednako
Ostar_pad	Srednji_rast	Velika	Mala	1.00	Nejednako
Ostar_pad	Srednji_rast	Velika	Srednja	1.00	Nejednako
Ostar_pad	Srednji_rast	Velika	Velika	1.00	Nejednako
Ostar_pad	Ostar_rast	Mala	Mala	1.00	Nejednako
Ostar_pad	Ostar_rast	Mala	Srednja	1.00	Nejednako
Ostar_pad	Ostar_rast	Mala	Velika	1.00	Nejednako
Ostar_pad	Ostar_rast	Srednja	Mala	1.00	Nejednako
Ostar_pad	Ostar_rast	Srednja	Srednja	1.00	Nejednako
Ostar_pad	Ostar_rast	Srednja	Velika	1.00	Nejednako
Ostar_pad	Ostar_rast	Velika	Mala	1.00	Nejednako
Ostar_pad	Ostar_rast	Velika	Srednja	1.00	Nejednako
Ostar_pad	Ostar_rast	Velika	Velika	1.00	Nejednako
Srednji_pad	Ostar_pad	Mala	Mala	1.00	Nejednako
Srednji_pad	Ostar_pad	Mala	Srednja	1.00	Nejednako
Srednji_pad	Ostar_pad	Mala	Velika	1.00	Nejednako
Srednji_pad	Ostar_pad	Srednja	Mala	1.00	Nejednako
Srednji_pad	Ostar_pad	Srednja	Srednja	1.00	Nejednako
Srednji_pad	Ostar_pad	Srednja	Velika	1.00	Nejednako
Srednji_pad	Ostar_pad	Velika	Mala	1.00	Nejednako
Srednji_pad	Ostar_pad	Velika	Srednja	1.00	Nejednako
Srednji_pad	Ostar_pad	Velika	Velika	1.00	Nejednako
Srednji_pad	Srednji_pad	Mala	Mala	1.00	Jednako
Srednji_pad	Srednji_pad	Mala	Srednja	1.00	Nejednako
Srednji_pad	Srednji_pad	Mala	Velika	1.00	Nejednako
Srednji_pad	Srednji_pad	Srednja	Mala	1.00	Nejednako
Srednji_pad	Srednji_pad	Srednja	Srednja	1.00	Nejednako
Srednji_pad	Srednji_pad	Srednja	Velika	1.00	Nejednako
Srednji_pad	Srednji_pad	Velika	Mala	1.00	Nejednako
Srednji_pad	Srednji_pad	Velika	Srednja	1.00	Nejednako
Srednji_pad	Srednji_pad	Velika	Velika	1.00	Jednako
Srednji_pad	Blagi_pad	Mala	Mala	1.00	Nejednako
Srednji_pad	Blagi_pad	Mala	Srednja	1.00	Nejednako
Srednji_pad	Blagi_pad	Mala	Velika	1.00	Nejednako
Srednji_pad	Blagi_pad	Srednja	Mala	1.00	Nejednako
Srednji_pad	Blagi_pad	Srednja	Srednja	1.00	Nejednako
Srednji_pad	Blagi_pad	Srednja	Velika	1.00	Nejednako
Srednji_pad	Blagi_pad	Velika	Mala	1.00	Nejednako
Srednji_pad	Blagi_pad	Velika	Srednja	1.00	Nejednako
Srednji_pad	Blagi_pad	Velika	Velika	1.00	Nejednako
Srednji_pad	Jednako	Mala	Mala	1.00	Nejednako
Srednji_pad	Jednako	Mala	Srednja	1.00	Nejednako
Srednji_pad	Jednako	Mala	Velika	1.00	Nejednako
Srednji_pad	Jednako	Srednja	Mala	1.00	Nejednako
Srednji_pad	Jednako	Srednja	Srednja	1.00	Nejednako
Srednji_pad	Jednako	Srednja	Velika	1.00	Nejednako
Srednji_pad	Jednako	Velika	Mala	1.00	Nejednako
Srednji_pad	Jednako	Velika	Srednja	1.00	Nejednako
Srednji_pad	Jednako	Velika	Velika	1.00	Nejednako
Srednji_pad	Blagi_rast	Mala	Mala	1.00	Nejednako
Srednji_pad	Blagi_rast	Mala	Srednja	1.00	Nejednako
Srednji_pad	Blagi_rast	Mala	Velika	1.00	Nejednako
Srednji_pad	Blagi_rast	Srednja	Mala	1.00	Nejednako
Srednji_pad	Blagi_rast	Srednja	Srednja	1.00	Nejednako
Srednji_pad	Blagi_rast	Srednja	Velika	1.00	Nejednako
Srednji_pad	Blagi_rast	Velika	Mala	1.00	Nejednako
Srednji_pad	Blagi_rast	Velika	Srednja	1.00	Nejednako
Srednji_pad	Blagi_rast	Velika	Velika	1.00	Nejednako

IF				THEN	
Srednji_pad	Blagi_rast	Velika	Srednja	1.00	Nejednako
Srednji_pad	Blagi_rast	Velika	Velika	1.00	Nejednako
Srednji_pad	Srednji_rast	Mala	Mala	1.00	Nejednako
Srednji_pad	Srednji_rast	Mala	Srednja	1.00	Nejednako
Srednji_pad	Srednji_rast	Mala	Velika	1.00	Nejednako
Srednji_pad	Srednji_rast	Srednja	Mala	1.00	Nejednako
Srednji_pad	Srednji_rast	Srednja	Srednja	1.00	Nejednako
Srednji_pad	Srednji_rast	Srednja	Velika	1.00	Nejednako
Srednji_pad	Srednji_rast	Velika	Mala	1.00	Nejednako
Srednji_pad	Srednji_rast	Velika	Srednja	1.00	Nejednako
Srednji_pad	Srednji_rast	Velika	Velika	1.00	Nejednako
Srednji_pad	Ostar_rast	Mala	Mala	1.00	Nejednako
Srednji_pad	Ostar_rast	Mala	Srednja	1.00	Nejednako
Srednji_pad	Ostar_rast	Mala	Velika	1.00	Nejednako
Srednji_pad	Ostar_rast	Srednja	Mala	1.00	Nejednako
Srednji_pad	Ostar_rast	Srednja	Srednja	1.00	Nejednako
Srednji_pad	Ostar_rast	Srednja	Velika	1.00	Nejednako
Srednji_pad	Ostar_rast	Velika	Mala	1.00	Nejednako
Srednji_pad	Ostar_rast	Velika	Srednja	1.00	Nejednako
Srednji_pad	Ostar_rast	Velika	Velika	1.00	Nejednako
Blagi_pad	Ostar_pad	Mala	Mala	1.00	Nejednako
Blagi_pad	Ostar_pad	Mala	Srednja	1.00	Nejednako
Blagi_pad	Ostar_pad	Mala	Velika	1.00	Nejednako
Blagi_pad	Ostar_pad	Srednja	Mala	1.00	Nejednako
Blagi_pad	Ostar_pad	Srednja	Srednja	1.00	Nejednako
Blagi_pad	Ostar_pad	Srednja	Velika	1.00	Nejednako
Blagi_pad	Ostar_pad	Velika	Mala	1.00	Nejednako
Blagi_pad	Ostar_pad	Velika	Srednja	1.00	Nejednako
Blagi_pad	Ostar_pad	Velika	Velika	1.00	Nejednako
Blagi_pad	Srednji_pad	Mala	Mala	1.00	Nejednako
Blagi_pad	Srednji_pad	Mala	Srednja	1.00	Nejednako
Blagi_pad	Srednji_pad	Mala	Velika	1.00	Nejednako
Blagi_pad	Srednji_pad	Srednja	Mala	1.00	Nejednako
Blagi_pad	Srednji_pad	Srednja	Srednja	1.00	Nejednako
Blagi_pad	Srednji_pad	Srednja	Velika	1.00	Nejednako
Blagi_pad	Srednji_pad	Velika	Mala	1.00	Nejednako
Blagi_pad	Srednji_pad	Velika	Srednja	1.00	Nejednako
Blagi_pad	Srednji_pad	Velika	Velika	1.00	Nejednako
Blagi_pad	Blagi_pad	Mala	Mala	1.00	Jednako
Blagi_pad	Blagi_pad	Mala	Srednja	1.00	Nejednako
Blagi_pad	Blagi_pad	Mala	Velika	1.00	Nejednako
Blagi_pad	Blagi_pad	Srednja	Mala	1.00	Nejednako
Blagi_pad	Blagi_pad	Srednja	Srednja	1.00	Jednako
Blagi_pad	Blagi_pad	Srednja	Velika	1.00	Nejednako
Blagi_pad	Blagi_pad	Velika	Mala	1.00	Nejednako
Blagi_pad	Blagi_pad	Velika	Srednja	1.00	Nejednako
Blagi_pad	Blagi_pad	Velika	Velika	1.00	Jednako
Blagi_pad	Jednako	Mala	Mala	1.00	Nejednako
Blagi_pad	Jednako	Mala	Srednja	1.00	Nejednako
Blagi_pad	Jednako	Mala	Velika	1.00	Nejednako
Blagi_pad	Jednako	Srednja	Mala	1.00	Nejednako
Blagi_pad	Jednako	Srednja	Srednja	1.00	Nejednako
Blagi_pad	Jednako	Srednja	Velika	1.00	Nejednako
Blagi_pad	Jednako	Velika	Mala	1.00	Nejednako
Blagi_pad	Jednako	Velika	Srednja	1.00	Nejednako
Blagi_pad	Jednako	Velika	Velika	1.00	Nejednako
Blagi_pad	Blagi_rast	Mala	Mala	1.00	Nejednako

IF				THEN	
Blagi_pad	Blagi_rast	Mala	Srednja	1.00	Nejednako
Blagi_pad	Blagi_rast	Mala	Velika	1.00	Nejednako
Blagi_pad	Blagi_rast	Srednja	Mala	1.00	Nejednako
Blagi_pad	Blagi_rast	Srednja	Srednja	1.00	Nejednako
Blagi_pad	Blagi_rast	Srednja	Velika	1.00	Nejednako
Blagi_pad	Blagi_rast	Velika	Mala	1.00	Nejednako
Blagi_pad	Blagi_rast	Velika	Srednja	1.00	Nejednako
Blagi_pad	Blagi_rast	Velika	Velika	1.00	Nejednako
Blagi_pad	Srednji_rast	Mala	Mala	1.00	Nejednako
Blagi_pad	Srednji_rast	Mala	Srednja	1.00	Nejednako
Blagi_pad	Srednji_rast	Mala	Velika	1.00	Nejednako
Blagi_pad	Srednji_rast	Srednja	Mala	1.00	Nejednako
Blagi_pad	Srednji_rast	Srednja	Srednja	1.00	Nejednako
Blagi_pad	Srednji_rast	Srednja	Velika	1.00	Nejednako
Blagi_pad	Srednji_rast	Velika	Mala	1.00	Nejednako
Blagi_pad	Srednji_rast	Velika	Srednja	1.00	Nejednako
Blagi_pad	Srednji_rast	Velika	Velika	1.00	Nejednako
Blagi_pad	Ostar_rast	Mala	Mala	1.00	Nejednako
Blagi_pad	Ostar_rast	Mala	Srednja	1.00	Nejednako
Blagi_pad	Ostar_rast	Mala	Velika	1.00	Nejednako
Blagi_pad	Ostar_rast	Srednja	Mala	1.00	Nejednako
Blagi_pad	Ostar_rast	Srednja	Srednja	1.00	Nejednako
Blagi_pad	Ostar_rast	Srednja	Velika	1.00	Nejednako
Blagi_pad	Ostar_rast	Velika	Mala	1.00	Nejednako
Blagi_pad	Ostar_rast	Velika	Srednja	1.00	Nejednako
Blagi_pad	Ostar_rast	Velika	Velika	1.00	Nejednako
Jednako	Ostar_pad	Mala	Mala	1.00	Nejednako
Jednako	Ostar_pad	Mala	Srednja	1.00	Nejednako
Jednako	Ostar_pad	Mala	Velika	1.00	Nejednako
Jednako	Ostar_pad	Srednja	Mala	1.00	Nejednako
Jednako	Ostar_pad	Srednja	Srednja	1.00	Nejednako
Jednako	Ostar_pad	Srednja	Velika	1.00	Nejednako
Jednako	Ostar_pad	Velika	Mala	1.00	Nejednako
Jednako	Ostar_pad	Velika	Srednja	1.00	Nejednako
Jednako	Ostar_pad	Velika	Velika	1.00	Nejednako
Jednako	Srednji_pad	Mala	Mala	1.00	Nejednako
Jednako	Srednji_pad	Mala	Srednja	1.00	Nejednako
Jednako	Srednji_pad	Mala	Velika	1.00	Nejednako
Jednako	Srednji_pad	Srednja	Mala	1.00	Nejednako
Jednako	Srednji_pad	Srednja	Srednja	1.00	Nejednako
Jednako	Srednji_pad	Srednja	Velika	1.00	Nejednako
Jednako	Srednji_pad	Velika	Mala	1.00	Nejednako
Jednako	Srednji_pad	Velika	Srednja	1.00	Nejednako
Jednako	Srednji_pad	Velika	Velika	1.00	Nejednako
Jednako	Blagi_pad	Mala	Mala	1.00	Nejednako
Jednako	Blagi_pad	Mala	Srednja	1.00	Nejednako
Jednako	Blagi_pad	Mala	Velika	1.00	Nejednako
Jednako	Blagi_pad	Srednja	Mala	1.00	Nejednako
Jednako	Blagi_pad	Srednja	Srednja	1.00	Nejednako
Jednako	Blagi_pad	Srednja	Velika	1.00	Nejednako
Jednako	Blagi_pad	Velika	Mala	1.00	Nejednako
Jednako	Blagi_pad	Velika	Srednja	1.00	Nejednako
Jednako	Blagi_pad	Velika	Velika	1.00	Nejednako
Jednako	Jednako	Mala	Mala	1.00	Jednako
Jednako	Jednako	Mala	Srednja	1.00	Nejednako
Jednako	Jednako	Mala	Velika	1.00	Nejednako
Jednako	Jednako	Srednja	Mala	1.00	Nejednako

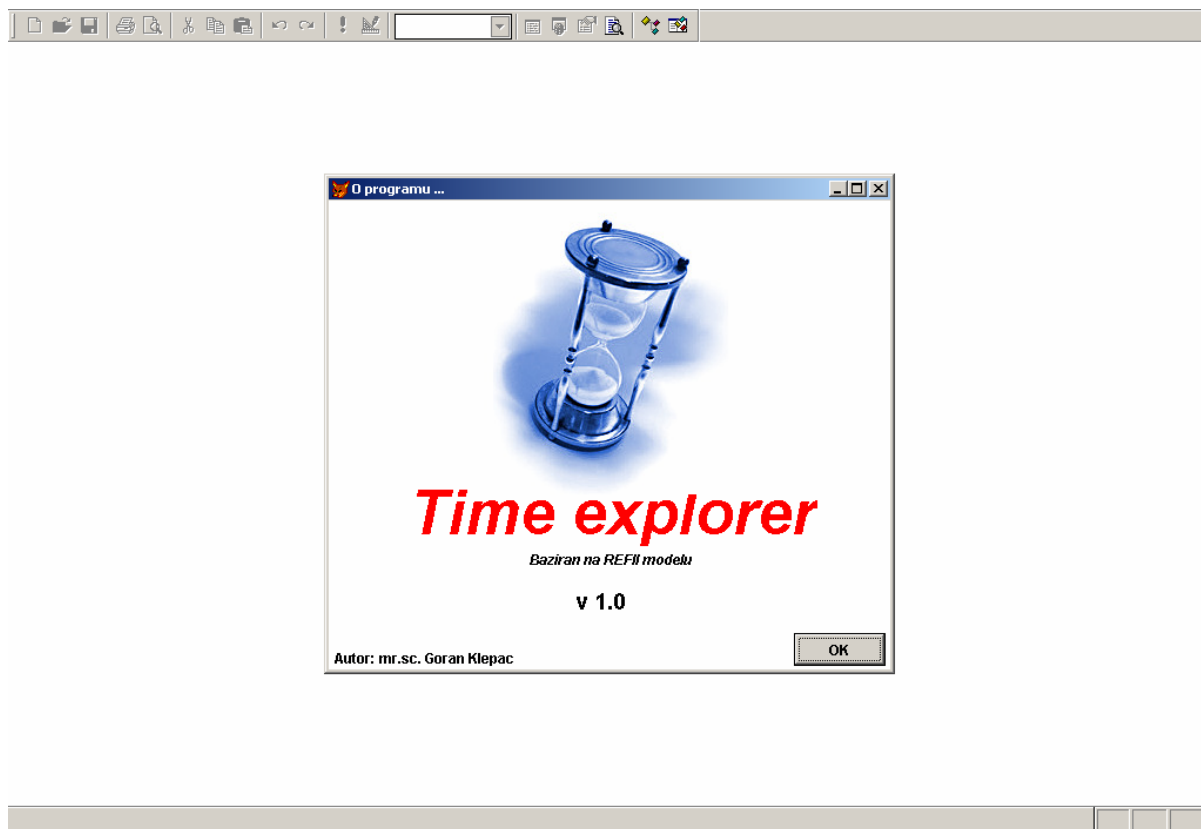
IF				THEN	
Jednako	Jednako	Srednja	Srednja	1.00	Jednako
Jednako	Jednako	Srednja	Velika	1.00	Nejednako
Jednako	Jednako	Velika	Mala	1.00	Nejednako
Jednako	Jednako	Velika	Srednja	1.00	Nejednako
Jednako	Jednako	Velika	Velika	1.00	Jednako
Jednako	Blagi_rast	Mala	Mala	1.00	Nejednako
Jednako	Blagi_rast	Mala	Srednja	1.00	Nejednako
Jednako	Blagi_rast	Mala	Velika	1.00	Nejednako
Jednako	Blagi_rast	Srednja	Mala	1.00	Nejednako
Jednako	Blagi_rast	Srednja	Srednja	1.00	Nejednako
Jednako	Blagi_rast	Srednja	Velika	1.00	Nejednako
Jednako	Blagi_rast	Velika	Mala	1.00	Nejednako
Jednako	Blagi_rast	Velika	Srednja	1.00	Nejednako
Jednako	Blagi_rast	Velika	Velika	1.00	Nejednako
Jednako	Srednji_rast	Mala	Mala	1.00	Nejednako
Jednako	Srednji_rast	Mala	Srednja	1.00	Nejednako
Jednako	Srednji_rast	Mala	Velika	1.00	Nejednako
Jednako	Srednji_rast	Srednja	Mala	1.00	Nejednako
Jednako	Srednji_rast	Srednja	Srednja	1.00	Nejednako
Jednako	Srednji_rast	Srednja	Velika	1.00	Nejednako
Jednako	Srednji_rast	Velika	Mala	1.00	Nejednako
Jednako	Srednji_rast	Velika	Srednja	1.00	Nejednako
Jednako	Srednji_rast	Velika	Velika	1.00	Nejednako
Jednako	Ostar_rast	Mala	Mala	1.00	Nejednako
Jednako	Ostar_rast	Mala	Srednja	1.00	Nejednako
Jednako	Ostar_rast	Mala	Velika	1.00	Nejednako
Jednako	Ostar_rast	Srednja	Mala	1.00	Nejednako
Jednako	Ostar_rast	Srednja	Srednja	1.00	Nejednako
Jednako	Ostar_rast	Srednja	Velika	1.00	Nejednako
Jednako	Ostar_rast	Velika	Mala	1.00	Nejednako
Jednako	Ostar_rast	Velika	Srednja	1.00	Nejednako
Jednako	Ostar_rast	Velika	Velika	1.00	Nejednako
Blagi_rast	Ostar_pad	Mala	Mala	1.00	Nejednako
Blagi_rast	Ostar_pad	Mala	Srednja	1.00	Nejednako
Blagi_rast	Ostar_pad	Mala	Velika	1.00	Nejednako
Blagi_rast	Ostar_pad	Srednja	Mala	1.00	Nejednako
Blagi_rast	Ostar_pad	Srednja	Srednja	1.00	Nejednako
Blagi_rast	Ostar_pad	Srednja	Velika	1.00	Nejednako
Blagi_rast	Ostar_pad	Velika	Mala	1.00	Nejednako
Blagi_rast	Ostar_pad	Velika	Srednja	1.00	Nejednako
Blagi_rast	Ostar_pad	Velika	Velika	1.00	Nejednako
Blagi_rast	Srednji_pad	Mala	Mala	1.00	Nejednako
Blagi_rast	Srednji_pad	Mala	Srednja	1.00	Nejednako
Blagi_rast	Srednji_pad	Mala	Velika	1.00	Nejednako
Blagi_rast	Srednji_pad	Srednja	Mala	1.00	Nejednako
Blagi_rast	Srednji_pad	Srednja	Srednja	1.00	Nejednako
Blagi_rast	Srednji_pad	Srednja	Velika	1.00	Nejednako
Blagi_rast	Srednji_pad	Velika	Mala	1.00	Nejednako
Blagi_rast	Srednji_pad	Velika	Srednja	1.00	Nejednako
Blagi_rast	Srednji_pad	Velika	Velika	1.00	Nejednako
Blagi_rast	Blagi_pad	Mala	Mala	1.00	Nejednako
Blagi_rast	Blagi_pad	Mala	Srednja	1.00	Nejednako
Blagi_rast	Blagi_pad	Mala	Velika	1.00	Nejednako
Blagi_rast	Blagi_pad	Srednja	Mala	1.00	Nejednako
Blagi_rast	Blagi_pad	Srednja	Srednja	1.00	Nejednako
Blagi_rast	Blagi_pad	Srednja	Velika	1.00	Nejednako
Blagi_rast	Blagi_pad	Velika	Mala	1.00	Nejednako

IF				THEN	
Blagi_rast	Blagi_pad	Velika	Srednja	1.00	Nejednako
Blagi_rast	Blagi_pad	Velika	Velika	1.00	Nejednako
Blagi_rast	Jednako	Mala	Mala	1.00	Nejednako
Blagi_rast	Jednako	Mala	Srednja	1.00	Nejednako
Blagi_rast	Jednako	Mala	Velika	1.00	Nejednako
Blagi_rast	Jednako	Srednja	Mala	1.00	Nejednako
Blagi_rast	Jednako	Srednja	Srednja	1.00	Nejednako
Blagi_rast	Jednako	Srednja	Velika	1.00	Nejednako
Blagi_rast	Jednako	Velika	Mala	1.00	Nejednako
Blagi_rast	Jednako	Velika	Srednja	1.00	Nejednako
Blagi_rast	Jednako	Velika	Velika	1.00	Nejednako
Blagi_rast	Blagi_rast	Mala	Mala	1.00	Jednako
Blagi_rast	Blagi_rast	Mala	Srednja	1.00	Nejednako
Blagi_rast	Blagi_rast	Mala	Velika	1.00	Nejednako
Blagi_rast	Blagi_rast	Srednja	Mala	1.00	Nejednako
Blagi_rast	Blagi_rast	Srednja	Srednja	1.00	Jednako
Blagi_rast	Blagi_rast	Srednja	Velika	1.00	Nejednako
Blagi_rast	Blagi_rast	Velika	Mala	1.00	Nejednako
Blagi_rast	Blagi_rast	Velika	Srednja	1.00	Nejednako
Blagi_rast	Blagi_rast	Velika	Velika	1.00	Jednako
Blagi_rast	Srednji_rast	Mala	Mala	1.00	Nejednako
Blagi_rast	Srednji_rast	Mala	Srednja	1.00	Nejednako
Blagi_rast	Srednji_rast	Mala	Velika	1.00	Nejednako
Blagi_rast	Srednji_rast	Srednja	Mala	1.00	Nejednako
Blagi_rast	Srednji_rast	Srednja	Srednja	1.00	Nejednako
Blagi_rast	Srednji_rast	Srednja	Velika	1.00	Nejednako
Blagi_rast	Srednji_rast	Velika	Mala	1.00	Nejednako
Blagi_rast	Srednji_rast	Velika	Srednja	1.00	Nejednako
Blagi_rast	Srednji_rast	Velika	Velika	1.00	Nejednako
Blagi_rast	Ostar_rast	Mala	Mala	1.00	Nejednako
Blagi_rast	Ostar_rast	Mala	Srednja	1.00	Nejednako
Blagi_rast	Ostar_rast	Mala	Velika	1.00	Nejednako
Blagi_rast	Ostar_rast	Srednja	Mala	1.00	Nejednako
Blagi_rast	Ostar_rast	Srednja	Srednja	1.00	Nejednako
Blagi_rast	Ostar_rast	Srednja	Velika	1.00	Nejednako
Blagi_rast	Ostar_rast	Velika	Mala	1.00	Nejednako
Blagi_rast	Ostar_rast	Velika	Srednja	1.00	Nejednako
Blagi_rast	Ostar_rast	Velika	Velika	1.00	Nejednako
Srednji_rast	Ostar_pad	Mala	Mala	1.00	Nejednako
Srednji_rast	Ostar_pad	Mala	Srednja	1.00	Nejednako
Srednji_rast	Ostar_pad	Mala	Velika	1.00	Nejednako
Srednji_rast	Ostar_pad	Srednja	Mala	1.00	Nejednako
Srednji_rast	Ostar_pad	Srednja	Srednja	1.00	Nejednako
Srednji_rast	Ostar_pad	Srednja	Velika	1.00	Nejednako
Srednji_rast	Ostar_pad	Velika	Mala	1.00	Nejednako
Srednji_rast	Ostar_pad	Velika	Srednja	1.00	Nejednako
Srednji_rast	Ostar_pad	Velika	Velika	1.00	Nejednako
Srednji_rast	Srednji_pad	Mala	Mala	1.00	Nejednako
Srednji_rast	Srednji_pad	Mala	Srednja	1.00	Nejednako
Srednji_rast	Srednji_pad	Mala	Velika	1.00	Nejednako
Srednji_rast	Srednji_pad	Srednja	Mala	1.00	Nejednako
Srednji_rast	Srednji_pad	Srednja	Srednja	1.00	Nejednako
Srednji_rast	Srednji_pad	Srednja	Velika	1.00	Nejednako
Srednji_rast	Srednji_pad	Velika	Mala	1.00	Nejednako
Srednji_rast	Srednji_pad	Velika	Srednja	1.00	Nejednako
Srednji_rast	Srednji_pad	Velika	Velika	1.00	Nejednako
Srednji_rast	Blagi_pad	Mala	Mala	1.00	Nejednako

IF				THEN	
Srednji_rast	Blagi_pad	Mala	Srednja	1.00	Nejednako
Srednji_rast	Blagi_pad	Mala	Velika	1.00	Nejednako
Srednji_rast	Blagi_pad	Srednja	Mala	1.00	Nejednako
Srednji_rast	Blagi_pad	Srednja	Srednja	1.00	Nejednako
Srednji_rast	Blagi_pad	Srednja	Velika	1.00	Nejednako
Srednji_rast	Blagi_pad	Velika	Mala	1.00	Nejednako
Srednji_rast	Blagi_pad	Velika	Srednja	1.00	Nejednako
Srednji_rast	Blagi_pad	Velika	Velika	1.00	Nejednako
Srednji_rast	Jednako	Mala	Mala	1.00	Nejednako
Srednji_rast	Jednako	Mala	Srednja	1.00	Nejednako
Srednji_rast	Jednako	Mala	Velika	1.00	Nejednako
Srednji_rast	Jednako	Srednja	Mala	1.00	Nejednako
Srednji_rast	Jednako	Srednja	Srednja	1.00	Nejednako
Srednji_rast	Jednako	Srednja	Velika	1.00	Nejednako
Srednji_rast	Jednako	Velika	Mala	1.00	Nejednako
Srednji_rast	Jednako	Velika	Srednja	1.00	Nejednako
Srednji_rast	Jednako	Velika	Velika	1.00	Nejednako
Srednji_rast	Blagi_rast	Mala	Mala	1.00	Nejednako
Srednji_rast	Blagi_rast	Mala	Srednja	1.00	Nejednako
Srednji_rast	Blagi_rast	Mala	Velika	1.00	Nejednako
Srednji_rast	Blagi_rast	Srednja	Mala	1.00	Nejednako
Srednji_rast	Blagi_rast	Srednja	Srednja	1.00	Nejednako
Srednji_rast	Blagi_rast	Srednja	Velika	1.00	Nejednako
Srednji_rast	Blagi_rast	Velika	Mala	1.00	Nejednako
Srednji_rast	Blagi_rast	Velika	Srednja	1.00	Nejednako
Srednji_rast	Blagi_rast	Velika	Velika	1.00	Nejednako
Srednji_rast	Srednji_rast	Mala	Mala	1.00	Jednako
Srednji_rast	Srednji_rast	Mala	Srednja	1.00	Nejednako
Srednji_rast	Srednji_rast	Mala	Velika	1.00	Nejednako
Srednji_rast	Srednji_rast	Srednja	Mala	1.00	Nejednako
Srednji_rast	Srednji_rast	Srednja	Srednja	1.00	Jednako
Srednji_rast	Srednji_rast	Srednja	Velika	1.00	Nejednako
Srednji_rast	Srednji_rast	Velika	Mala	1.00	Nejednako
Srednji_rast	Srednji_rast	Velika	Srednja	1.00	Nejednako
Srednji_rast	Srednji_rast	Velika	Velika	1.00	Jednako
Srednji_rast	Ostar_rast	Mala	Mala	1.00	Nejednako
Srednji_rast	Ostar_rast	Mala	Srednja	1.00	Nejednako
Srednji_rast	Ostar_rast	Mala	Velika	1.00	Nejednako
Srednji_rast	Ostar_rast	Srednja	Mala	1.00	Nejednako
Srednji_rast	Ostar_rast	Srednja	Srednja	1.00	Nejednako
Srednji_rast	Ostar_rast	Srednja	Velika	1.00	Nejednako
Srednji_rast	Ostar_rast	Velika	Mala	1.00	Nejednako
Srednji_rast	Ostar_rast	Velika	Srednja	1.00	Nejednako
Srednji_rast	Ostar_rast	Velika	Velika	1.00	Nejednako
Ostar_rast	Ostar_pad	Mala	Mala	1.00	Nejednako
Ostar_rast	Ostar_pad	Mala	Srednja	1.00	Nejednako
Ostar_rast	Ostar_pad	Mala	Velika	1.00	Nejednako
Ostar_rast	Ostar_pad	Srednja	Mala	1.00	Nejednako
Ostar_rast	Ostar_pad	Srednja	Srednja	1.00	Nejednako
Ostar_rast	Ostar_pad	Srednja	Velika	1.00	Nejednako
Ostar_rast	Ostar_pad	Velika	Mala	1.00	Nejednako
Ostar_rast	Ostar_pad	Velika	Srednja	1.00	Nejednako
Ostar_rast	Ostar_pad	Velika	Velika	1.00	Nejednako
Ostar_rast	Srednji_pad	Mala	Mala	1.00	Nejednako
Ostar_rast	Srednji_pad	Mala	Srednja	1.00	Nejednako
Ostar_rast	Srednji_pad	Mala	Velika	1.00	Nejednako
Ostar_rast	Srednji_pad	Srednja	Mala	1.00	Nejednako

IF				THEN	
Ostar_rast	Srednji_pad	Srednja	Srednja	1.00	Nejednako
Ostar_rast	Srednji_pad	Srednja	Velika	1.00	Nejednako
Ostar_rast	Srednji_pad	Velika	Mala	1.00	Nejednako
Ostar_rast	Srednji_pad	Velika	Srednja	1.00	Nejednako
Ostar_rast	Srednji_pad	Velika	Velika	1.00	Nejednako
Ostar_rast	Blagi_pad	Mala	Mala	1.00	Nejednako
Ostar_rast	Blagi_pad	Mala	Srednja	1.00	Nejednako
Ostar_rast	Blagi_pad	Mala	Velika	1.00	Nejednako
Ostar_rast	Blagi_pad	Srednja	Mala	1.00	Nejednako
Ostar_rast	Blagi_pad	Srednja	Srednja	1.00	Nejednako
Ostar_rast	Blagi_pad	Srednja	Velika	1.00	Nejednako
Ostar_rast	Blagi_pad	Velika	Mala	1.00	Nejednako
Ostar_rast	Blagi_pad	Velika	Srednja	1.00	Nejednako
Ostar_rast	Blagi_pad	Velika	Velika	1.00	Nejednako
Ostar_rast	Jednako	Mala	Mala	1.00	Nejednako
Ostar_rast	Jednako	Mala	Srednja	1.00	Nejednako
Ostar_rast	Jednako	Mala	Velika	1.00	Nejednako
Ostar_rast	Jednako	Srednja	Mala	1.00	Nejednako
Ostar_rast	Jednako	Srednja	Srednja	1.00	Nejednako
Ostar_rast	Jednako	Srednja	Velika	1.00	Nejednako
Ostar_rast	Jednako	Velika	Mala	1.00	Nejednako
Ostar_rast	Jednako	Velika	Srednja	1.00	Nejednako
Ostar_rast	Jednako	Velika	Velika	1.00	Nejednako
Ostar_rast	Blagi_rast	Mala	Mala	1.00	Nejednako
Ostar_rast	Blagi_rast	Mala	Srednja	1.00	Nejednako
Ostar_rast	Blagi_rast	Mala	Velika	1.00	Nejednako
Ostar_rast	Blagi_rast	Srednja	Mala	1.00	Nejednako
Ostar_rast	Blagi_rast	Srednja	Srednja	1.00	Nejednako
Ostar_rast	Blagi_rast	Srednja	Velika	1.00	Nejednako
Ostar_rast	Blagi_rast	Velika	Mala	1.00	Nejednako
Ostar_rast	Blagi_rast	Velika	Srednja	1.00	Nejednako
Ostar_rast	Blagi_rast	Velika	Velika	1.00	Nejednako
Ostar_rast	Srednji_rast	Mala	Mala	1.00	Nejednako
Ostar_rast	Srednji_rast	Mala	Srednja	1.00	Nejednako
Ostar_rast	Srednji_rast	Mala	Velika	1.00	Nejednako
Ostar_rast	Srednji_rast	Srednja	Mala	1.00	Nejednako
Ostar_rast	Srednji_rast	Srednja	Srednja	1.00	Nejednako
Ostar_rast	Srednji_rast	Srednja	Velika	1.00	Nejednako
Ostar_rast	Srednji_rast	Velika	Mala	1.00	Nejednako
Ostar_rast	Srednji_rast	Velika	Srednja	1.00	Nejednako
Ostar_rast	Srednji_rast	Velika	Velika	1.00	Nejednako
Ostar_rast	Ostar_rast	Mala	Mala	1.00	Jednako
Ostar_rast	Ostar_rast	Mala	Srednja	1.00	Nejednako
Ostar_rast	Ostar_rast	Mala	Velika	1.00	Nejednako
Ostar_rast	Ostar_rast	Srednja	Mala	1.00	Nejednako
Ostar_rast	Ostar_rast	Srednja	Srednja	1.00	Jednako
Ostar_rast	Ostar_rast	Srednja	Velika	1.00	Nejednako
Ostar_rast	Ostar_rast	Velika	Mala	1.00	Nejednako
Ostar_rast	Ostar_rast	Velika	Srednja	1.00	Nejednako
Ostar_rast	Ostar_rast	Velika	Velika	1.00	Jednako

Uvodna maska "Time explorera-a"



Program za vremensku interpolaciju i vremensku granulaciju

```
PRIVATE put,tablica,putanja,poljev,poljed,vrij  
close databases
```

```
IF EMPTY(THISFORM.text1.VALUE) OR EMPTY(THISFORM.text2.VALUE) OR  
EMPTY(THISFORM.text3.VALUE) OR EMPTY(THISFORM.text4.VALUE)
```

```
    THISFORM.text6.VALUE="Status: Potrebno je unijeti sve tražene vrijednosti !!!"  
    THISFORM.Refresh()  
    RETURN
```

```
ENDIF  
put=ALLTRIM(THISFORM.text1.VALUE)  
tablica=ALLTRIM(THISFORM.text2.VALUE)  
poljev=ALLTRIM(THISFORM.text3.VALUE)  
poljed=ALLTRIM(THISFORM.text4.VALUE)
```

putanja=put+tablica

USE &putanja IN 0

CREATE TABLE c:\radni\zaref (datum D(8),DanUT N(2),Tjedan N(5),Mjesec
N(2),Godina N(5),vrijednost N (20,9),flag C(1))

dat=CTOD(init.poc_datum)

THISFORM.text6.VALUE="Status: Formiram tablicu uzoraka..."

THISFORM.Refresh()

SELECT zaref

GO TOP

DO WHILE YEAR(dat)=init.godina

APPEND BLANK

REPLACE zaref.datum WITH dat

REPLACE zaref.Tjedan WITH WEEK(dat)

REPLACE zaref.Mjesec WITH MONTH(dat)

REPLACE zaref.DanUT WITH DOW(dat)

REPLACE zaref.Godina WITH YEAR(dat)

IF THISFORM.optiongroup1.option1.Value=1

REPLACE zaref.flag WITH "D"

ENDIF

IF THISFORM.optiongroup1.option2.Value=1

REPLACE zaref.flag WITH "T"

ENDIF

IF THISFORM.optiongroup1.option3.Value=1

REPLACE zaref.flag WITH "M"

ENDIF

IF THISFORM.optiongroup1.option1.Value=1

REPLACE zaref.flag WITH "G"

ENDIF

dat=dat+1

ENDDO

THISFORM.text6.VALUE="Status: Tablica uzoraka formirana!"

THISFORM.Refresh()

koi=DATE()

THISFORM.text6.VALUE="Status: Proces interpolacije u tijeku..."

THISFORM.Refresh()

SELECT zaref

GO top

DO WHILE NOT EOF()

koi=zaref.datum

SELECT &tablica

GO TOP

LOCATE FOR &poljed=koi

IF FOUND()

```

        vrij=&poljev

ELSE
    vrij=0

ENDIF
SELECT zaref
REPLACE zaref.vrijednost WITH vrij
SKIP 1
ENDDO
THISFORM.text6.VALUE="Status: Proces interpolacije završen !"
    THISFORM.Refresh()

THISFORM.text6.VALUE="Status: Proces granulacije u tijeku..."
    THISFORM.Refresh()

IF THISFORM.optiongroup1.option1.Value=1 && Dani
    SELECT datum,sum(vrijednost) as vrijednost, danut as indeks, tjedan as
ciklus,mjesec as mjesec FROM zaref ;
    GROUP BY datum ORDER BY datum ASC INTO TABLE c:\te\data\refiis
ENDIF

IF THISFORM.optiongroup1.option2.Value=1 && tjedni
    SELECT sum(vrijednost) as vrijednost, tjedan as indeks, mjesec as ciklus FROM
zaref;
    GROUP BY tjedan ORDER BY tjedan ASC INTO TABLE c:\te\data\refiis
ENDIF

IF THISFORM.optiongroup1.option3.Value=1 && mjeseci
    SELECT sum(vrijednost) as vrijednost, mjesec as ind, godina as cik FROM
zaref GROUP BY;
    mjesec ORDER BY mjesec ASC INTO TABLE y:\te\data\refiis
ENDIF

THISFORM.text6.VALUE="Status: Proces granulacije završen !"
    THISFORM.Refresh()

    THISFORM.command4.ENABLED=.T.

```

Transformacija u REFII model

```

PRIVATE put,tablica,putanja,polje,v1,v2,maks,mini
close databases

IF EMPTY(THISFORM.text1.VALUE) OR EMPTY(THISFORM.text2.VALUE) OR
EMPTY(THISFORM.text3.VALUE)

    THISFORM.text4.VALUE="Status: Potrebno je unijeti sve tražene vrijednosti !!!"
    THISFORM.Refresh()
    RETURN

```



```

ENDIF
put=ALLTRIM(THISFORM.text1.VALUE)
tablica=ALLTRIM(THISFORM.text2.VALUE)
polje=ALLTRIM(THISFORM.text3.VALUE)
putanja=put+tablica
rputanja=put+"razredi"

izvrsi.putanja=put
izvrsi.tablica=tablica

USE &putanja IN 0
USE &rputanja IN 0
SELECT &tablica
SELECT MAX(&polje) as MAKS, Min(&polje) as MINI FROM &tablica INTO CURSOR
ini
maks=ini.maks
mini=ini.mini

USE c:\te\data\refii IN 0 EXCLUSIVE
SELECT refii
ZAP

```

```

THISFORM.text4.VALUE="Status: Proces transformacije u REFII sintaksu u tijeku..."
THISFORM.Refresh()

```

```

SELECT &tablica
GO TOP
DO WHILE NOT EOF()
    THISFORM.text5.VALUE=STR(RECNO()/RECCOUNT()*100,8,2)+ " %"
    THISFORM.Refresh()
    v1=&polje
    SKIP 1
    v2=&polje

```

```

                SELECT REFII
                APPEND BLANK
REPLACE REFII.ref WITH izvrsi.ref(v1,v2)
REPLACE REFII.kut WITH;
izvrsi.kut(izvrsi.normiraj(v1,mini,maks),izvrsi.normiraj(v2,mini,maks))
REPLACE REFII.koef WITH;
izvrsi.koef(izvrsi.normiraj(v1,mini,maks),izvrsi.normiraj(v2,mini,maks))

                replace REFII.povrsina with izvrsi.povrsina(v1,v2)
                replace REFII.indeks WITH refiis.indeks
                REPLACE REFII.ciklus WITH refiis.ciklus
                REPLACE REFII.mjesecc WITH refiis.mjesecc
SELECT &tablica

```

ENDDO

```
SELECT REFII
GO BOTTOM
DELETE
THISFORM.text4.VALUE="Status: Proces transformacije u REFII sintaksu završen !"
```

```
cMessageTitle = 'Klasifikacija po razredima'
cMessageText = 'Da li želite klasificirati otklone kuteva po razredima?'
nDialogType = 4 + 32 + 256
nAnswer = MESSAGEBOX(cMessageText, nDialogType, cMessageTitle)
```

```
DO CASE
CASE nAnswer = 6
do svrstaj.prg
CASE nAnswer = 7
WAIT WINDOW 'Odustali ste od klasifikacije !'
ENDCASE
```

```
THISFORM.command4.ENABLED=.T.
THISFORM.Refresh()
```

Pomoćni objekti transformacije

```
mainlib.normiraj()
PARAMETERS v,minimum,maksimum
local vnovi
vnovi=(v-minimum)/(maksimum-minimum)
RETURN vnovi
```

```
mainlib.koeficijent()
PARAMETERS y1, y2
PRIVATE vrijednost,k
DO CASE
CASE y2-y1= 0
RETURN (0)
CASE y2-y1> 0

k=(y2-y1)
RETURN (k)
```

```
CASE y2-y1< 0
```

```
k=(y1-y2)
```

```
RETURN (k)
ENDCASE
```

```
mainlib.ref()
PARAMETERS k1, k2
DO CASE
CASE k2-k1= 0
RETURN ("E")
CASE k2-k1> 0
RETURN ("R")
CASE k2-k1< 0
RETURN ("F")
ENDCASE
```

```
mainlib.povrsina()
PARAMETERS x1, x2
PRIVATE p
p=((x1*1)+(x2*1))/2
RETURN (p)
```

svrstaj.prg

```
DO WHILE NOT EOF()
WAIT WINDOW "Klasificiram kuteve..." NOWAIT
m.iznos=refii.koef
m.oznaka=refii.ref
SELECT razredi
GO top
LOCATE FOR BETWEEN(m.iznos,razredi.dg,razredi.gg) AND razredi.ref=m.oznaka
IF FOUND()
m.razred=razredi.razred
SELECT refii
REPLACE refii.razred WITH ALLTRIM(m.razred)
ENDIF

SELECT REFII
WAIT WINDOW STR(RECNO()) NOWAIT
SKIP 1
ENDDO
WAIT WINDOW "GOTOVO !!!" NOWAIT
```

Modul za otkrivanje uzoraka u vremenskim serijama

```
class Uzorci:
def __init__(self):
self.data = []
self.stablo=[]
```

```

self.nova=[]
self.uzorak=[]
self.sloj0=[]
self.sloj1=[]
self.sloj2=[]
self.sloj3=[]
self.sloj4=[]
self.sloj5=[]
self.sloj6=[]
self.sloj7=[]
self.sloj8=[]

def add(self, x):
    self.data.append(x)
def addtwice(self, x):
    self.add(x)
    self.add(x)
def datoteka(self,dat):
    for line in open(dat):
        line=line.strip('\n')
        line=line.strip('""')
        self.add(line)

def kroz_listu(self,duzina):

    for x in range(len(self.data)-(duzina-1)):

        self.uzorak=[]
        for y in range(x,x+duzina):
            self.uzorak.append(self.data[y])
            if duzina==1:
                self.sito(self.data[y])

        if duzina==2:
            self.brisi_nefrekventne(self.sloj0,0.25,1)
            self.sloj0=self.filtriraj_listu(self.sloj0)
            self.medjuslojno_sito(self.sloj0,self.uzorak,self.sloj1)

        if duzina==3:
            self.brisi_nefrekventne(self.sloj1,0.1,2)
            self.sloj1=self.filtriraj_listu(self.sloj1)
            self.medjuslojno_sito(self.sloj1,self.uzorak,self.sloj2)

        if duzina==4:
            self.brisi_nefrekventne(self.sloj2,0.1,3)
            self.sloj2=self.filtriraj_listu(self.sloj2)
            self.medjuslojno_sito(self.sloj2,self.uzorak,self.sloj3)

        if duzina==5:
            self.brisi_nefrekventne(self.sloj3,0.1,4)

```

```

        self.sloj3=self.filtriraj_listu(self.sloj3)
        self.medjuslojno_sito(self.sloj3,self.uzorak,self.sloj4)

    if duzina==5:
        self.brisi_nefrekventne(self.sloj4,0.1,4)
        self.sloj4=self.filtriraj_listu(self.sloj4)
        self.medjuslojno_sito(self.sloj4,self.uzorak,self.sloj5)

def prolaz(self):
    brojac=1
    while brojac<=6:
        self.kroz_listu(brojac)
        brojac=brojac+1

def sito(self,uzorak):
    indikator=0
    for c in range(len(self.sloj0)):
        if self.sloj0[c][1]==uzorak:
            indikator=1
            self.sloj0[c][2]=self.sloj0[c][2]+1
    if indikator==0:
        pomoc=[]
        p1=[uzorak]
        p1.append
        pomoc=['0',uzorak,1,p1]
        self.sloj0.append(pomoc)

def brisi_nefrekventne(self,l,ponavljanje,duz):
    for b in range(len(l)):
        if (float(l[b][2])/(len(self.data)-(duz-1)))<ponavljanje:
            l[b][2]=0

def filtriraj_listu(self,l):
    Inova=[]
    for b in range(len(l)):
        if l[b][2]<>0:
            Inova.append(l[b])
    return Inova

def medjuslojno_sito(self,lista_upper,uzorak,lista_izlaz):
    indikator=0
    for c1 in range(len(lista_izlaz)):
        if lista_izlaz[c1][1]==uzorak[-1] and lista_izlaz[c1][3]==uzorak:
            indikator=1
            lista_izlaz[c1][2]=lista_izlaz[c1][2]+1

    if indikator==0:
        for b1 in range(len(lista_upper)):

```

```

        if lista_upper[b1][1]==uzorak[-2] and
lista_upper[b1][3]==uzorak[0:(len(uzorak)-1)]:
        pomoc=[]
        pomoc=[uzorak[-2],uzorak[-1],1,uzorak]
        lista_izlaz.append(pomoc)

```

Objekt za otkrivanje sezonskih oscilacija

```

IF NOT USED("RAZREDI")
    USE c:\te\data\RAZREDI IN 0
ENDIF

```

```

IF NOT USED("REFII")
    USE c:\te\data\refii IN 0
ENDIF

```

```

CREATE TABLE uzorci(indeks N(6),indreset N(6), frek N(6))

```

```

SELECT razredi
I=RECCOUNT()

```

```

DIMENSION resetka(I)
j=1
DO while NOT EOF()
    resetka(j)=razredi.razred
    WAIT WINDOW resetka(j) nowait
    j=j+1
    SKIP 1
ENDDO

```

```

SELECT refii
GO top
DO WHILE NOT EOF()
m.indeks=indeks
m.razred=refii.razred

```

```

FOR j=1 TO I
    IF ALLTRIM(m.razred)=ALLTRIM(resetka(j))
        WAIT WINDOW resetka(j)+" "+m.razred + " "+STR(j) nowait
        m.indreset=j
        EXIT
    ENDIF
ENDFOR

SELECT uzorci
GO TOP

```

```

LOCATE FOR (uzorci.indeks=m.indeks AND uzorci.indreset=j)
IF FOUND()
    replace frek WITH frek+1
ELSE
    APPEND BLANK
    replace indeks WITH m.indeks
    replace indreset WITH m.indreset
    replace frek WITH frek+1
ENDIF

```

```

SELECT refii
SKIP 1
ENDDO

```

```

SELECT indeks, count(*) as broj FROM refii GROUP BY indeks into TABLE broj
SELECT uzorci.indeks, uzorci.indreset, uzorci.frek,broj FROM broj,uzorci WHERE
uzorci.indeks=broj.indeks INTO TABLE uzorci1
SELECT indeks,indreset, frek/broj*100 as pouzdanost FROM uzorci1 INTO TABLE
uzorci2 ORDER BY indeks

```

```

SELECT * from refii WHERE indeks=5 or indeks=6 ORDER BY ciklus,indeks INTO
TABLE zauzor

```

Kod za otkrivanje događajnosti

```

A=['4', '1', '1']
B=['1', '4', '4']
C=['4', '4', '1']
D=['4', '1', '4']
event=[]
for br in range(len(stablo.dogadjaj)):
    if stablo.dogadjaj[br]==A:
        event.append("A")
    elif stablo.dogadjaj[br]==B:
        event.append("B")
    elif stablo.dogadjaj[br]==C:
        event.append("C")
    elif stablo.dogadjaj[br]==D:
        event.append("D")
    else:
        pass

    stablo.sloj0=[]
    stablo.sloj1=[]
    stablo.sloj2=[]
    stablo.sloj3=[]
    stablo.sloj4=[]
    stablo.sloj5=[]

```

```

        stablo.sloj6=[]
        stablo.sloj7=[]
        stablo.sloj8=[]
    stablo.data=event
    stablo.prolaz()
    print "cvor", stablo.sloj0
    print "prvi sloj ", stablo.sloj1
    print "drugi sloj ", stablo.sloj2
    print "treći sloj ", stablo.sloj3
    print "četvrti sloj ", stablo.sloj4

```

Program za traženje motiva

```

    Lista[]
    lista_izlaznih_indeksa[]
    stablo.kroz_listu(2)
    Lista=stablo.data
    Trazeni_motiv=[ [1,1], [1,2] ]

    for c1 in range(len(Lista)):
        if Lista[c1][-2]==Trazeni_motiv[1][1] and Lista[c1][0]==Trazeni_motiv[1][2] and
        Lista[c1+1][-2]==Trazeni_motiv[2][1] and Lista[c1+1][-1]==Trazeni_motiv[2][2] :
            pom1.append(Lista[c1][1])
            pom1.append(Lista[c1][2])
            pom1.append(Lista[c1+1][1])
            pom1.append(Lista[c1+2][2])
            lista_izlaznih_indeksa.append(pom1)

```