

## Data mining sustavi- kako i čime stići do cilja

Da kojim slučajem ovaj tekst čitate 1996 godine, vjerojatno vam se pojam *data warehousinga* ne bi činio toliko bliskim kao što je to slučaj sada. Ako promatramo genezu prihvaćanja *data warehouse* sustava na hrvatskom tržištu, tada je za očekivati da će i *data mining* sustavi trebati prijeći isti trnoviti put kako bi jednog dana u hrvatskim poslovnim okvirima bili prihvaćeni u punopravno članstvo nečega što percipiramo kao sastavni dio BI sustava.

Kada se jednog dana ipak uputite u avanturu dubinskog kopanja vaših podataka u potrazi za skrivenim grumenjima znanja, korisno je imati smjernice u kojem smjeru učiniti prvi korak prije nego što uronite u dubine vlastitih podataka.

Jedan od uvriježenih pogrešnih stereotipa svodi se na odgađanje provedbe *data mining* analiza, jer ne postoji izgrađeno skladište podataka, iako ono nije preduvjet za njihovo provođenje. Skladište podataka samo olakšava provođenje *data mining* analiza.

Prilikom donošenja odluke o tome kako će ove analize postati sastavni dio vašeg poslovanja vjerojatno će ipak komponenta ulaganja u softver imati presudnu ulogu. U skladu s tim možda ćete se prepoznati u tri kategorije izjavnih rečenica.

### **Situacija: "Ne želim uložiti ni kune u softver"**

Čak i u slučaju donošenja ovakve vrste odluke ipak postoji nada da će se i u vašim podacima čuti odjeci kopanja po podacima.

Ulaganje od 0 kuna podrazumijeva korištenje *open source* sustava za *data mining*. Temeljno gorivo koje pokreće razvoj ovakvih vrsta sustava (Orange, Weka, GeNIE) je beskrajni entuzijizam njihovih autora, te se dobrim dijelom razvijaju u okviru različitih sveučilišta. Ova rješenja kvalitetom često puta ne zaostaju za komercijalnim proizvodima, te se isporučuju bilo kao skup objektno orijentiranih programskih biblioteka, bilo kao sustavi sa korisničkim sučeljem, ili pak kao kombinacija jednog i drugog. Jedna od temeljnih prednosti korištenja ovakvih vrsta sustava proizlazi iz činjenice što se u njima često puta pojavljuju vrlo kvalitetna inovativna rješenja proizašla temeljem znanstvenih istraživanja. Da bi se ona efikasno iskoristila, potreban je vrlo visok stupanj stručnosti za njihovo korištenje, što podrazumijeva dubinsko poznavanje metoda *data mininga*, jer korisnici ne mogu računati na tehničku podršku prilikom njihovog korištenja u onoj mjeri kao što je to slučaj sa komercijalnim softverskim proizvodima. Često puta je potrebno dorađivati izvorne kodove ili pisati dodatne programske module kako bi sustave prilagodili konkretnim analitičkim potrebama, ili omekšala "krutost" izvornog oblika sustava. Prilagodba se u najvećem broju slučajeva odnosi na postizanje kompatibilnosti sa raspoloživim izvorima podataka te prilagodbi sustava za rješavanje specifičnih problema proizašlih iz poslovne prakse. No bez obzira na sve navedene prepreke krajnji rezultati nakon prilagodbe mogu biti vrlo impresivni, i što je najvažnije vrlo fleksibilni. Ovakav pristup preporuča se u poslovnoj okolini koja raspolaže sa jakim stručnim kadrom za *data mining*, gdje se pojavljuje česta potreba za *ad hoc data mining* analizama, kao i potreba za razvoj nestandardnih aplikativnih rješenja baziranih na otkrivanju znanja iz baza podataka.

### **Situacija: "Ne želim previše uložiti u softver"**

Ovakva odluka neminovno vas vodi do kupovanja niza analitičkih alata za *data mining*, pri čemu se naglasak prilikom kupnje stavlja s obzirom na djelatnost vašeg poduzeća i željenu problematiku kojom se bavite u okviru analiza. Nezaobilazni komercijalni alat za temeljne

analize koji se može proširivati s obzirom na problematiku analize je SPSS, koji već u osnovnoj verziji uz široku lepezu statističkih modela sadrži metode poput hijerarhijskog i *k-mean* klasteriranja. Ovisno o analitičkim potrebama, moguće je dokupiti modele kao što su primjerice "survival modeli", regresijski modeli, stabla odlučivanja, modeli za analizu vremenskih serija, pročišćavanje podataka i slično.

Osim SPSS-a, moguće je kupiti alate poput Hugin (Bayesove mreže), FuzzyTecha (fuzzy logika), NeuroTran (neuronske mreže), Nenet (samoorganizirajuće mape), te ih međusobno povezati s ciljem provođenja složenih analitičkih zadataka.

Ovdje je također dominantna uloga visokog stupnja stručnosti u poznavanju metodologije *data mininga*. Za razliku od prethodnog slučaja, ovdje se može računati na visoki stupanj tehničke potpore proizvođača softvera, pri čemu se ne smije zanemariti problematika povezivanja različitih softverskih koncepcija u harmoničnu cjelinu, koja u konačnici mora rezultirati rješavanju specifičnih vrsta problema. Prednost ovakvog pristupa također je fleksibilnost u izgradnji analitičkih modela, manji stupanj "krutosti" softvera nego što je to slučaj sa situacijom u kojoj se koristi *open source*, te se samim time smanjuje potreba programske dorade sustava.

### **Situacija: "Želim integrirano rješenje i potpunu konzultantsku uslugu"**

Ako se odlučite za ovu soluciju u domeni *data mininga*, tada će vaša odluka zasigurno pasti na sustave za *data mining* kakve nude SAS (Enterprise miner) i SPSS (Clementine). Osim što ćete uložiti u kvalitetna softverska rješenja, dobit ćete i stručnu konzultantsku pomoć ovisno o cilju zbog kojeg uvodite ovakvu vrstu sustava. Na ovaj način također možete implementirati specifične sustave koji se oslanjaju na *data mining* metodologiju poput sustava za *churn analize*, *fraud detection* i slično. Ovakav pristup rješavanja problematike *data mininga* unutar vašeg poduzeća implicira značajna ulaganja, koja su nemjerljiva po iznosima u odnosu na prethodno spomenute slučajeve.

### **Ne zavaravajte se da ste kupovinom softvera riješili sve probleme**

Iako je za sve segmente BI –a jedan od presudnih faktora tehnologija, u *data miningu* više negoli u bilo kojem drugom segmentu BI-a dominantnu ulogu ima čovjek kao glavni resurs. Bez obzira na to što postoje razvijeni softverski alati koji se temelje na sofisticiranim matematičkim metodama, oni su samo sredstvo pomoću kojeg se dolazi do analitičkih ciljeva, pri čemu ovdje kreativnost dolazi do punog izražaja. Dodatna kompleksnost ovog područja proizlazi iz činjenice da je izgrađene modele potrebno neprestano nadograđivati pa čak i u potpunosti izmijeniti s obzirom na vremenske odmake u provođenju analiza.

Dakle, ako razmišljate o provođenju *data mining* analiza kako bi riješili određene tipove problema u svom poduzeću, ne razmišljajte koji softver kupiti, već da li imate dostupnog stručnjaka (bilo unutar poduzeća, tržišta radne snage ili u vidu vanjskog konzultanta) koji može uspješno obaviti taj posao. Spremnost poduzeća prilikom korištenja ove metodologije svakako moramo promatrati i u svjetlu upućenosti gdje nam ove analize mogu pomoći u rješavanju praktičnih problema iz perspektive poslovanja vlastitog poduzeća.