

© Prof. dr. sc. Goran Klepac

www.goranklepac.com

goran(*at*)goranklepac.com

Kako u potpunosti iscijediti informacijsku naranču

Zvezdana plovidba

Kao što bi rekao Krleža „... i s najjeftinijom busolom lakše je ploviti nego po zvijezdama, pogotovo kad je oblačno pak nema zvijezda...“. Parafrazirajući ovu misao, data warehouse sustavi, koji su postali naša svakodnevnica, poduzeća promatraju poput zvijezda u noći, pokušavajući se orijentirati na nemirnim tržišnim vodama nastojeći zadržati ispravan put i zaobilazeći opasnosti.

No, objektivno gledajući postojeća rješenja, koja se često puta oslanjaju na sustav izvještaja iz data warehousea, ponekad nisu ništa više nego jeftina busola (ako se držimo ove alegorije), koja donekle služi svrsi te je neosporno kako ju je nužno imati.

Problemi telekomunikacijske kompanije

Pretpostavimo da ste telekomunikacijska kompanija koja pruža usluge mobilne telefonije, i da ste upravo shvatili kako imate problema sa rastućim trendom prekida ugovornih odnosa. Vjerojatno biste nakon te spoznaje (iako biste sigurno imali odjel koji bi se bavio praćenjem tih trendova), tražili izvještaje o stopama prekida ugovornih odnosa u posljednjih nekoliko mjeseci, stopama prekida ugovornih odnosa po regijama, stopama prekida ugovornih odnosa prema segmentima (ako imate razrađene segmente) Ovakvi izvještaji bi vam mogli dati naznaku gdje je problem, no vrlo teško će vam dati potpuni odgovor, a još manje ćete iz njih razraditi strategiju akcija u cilju sprečavanja prekida ugovornih odnosa (osim u nekim iznimnim situacijama).

Naredni korak bio bi zasigurno usmjeren na višedimenzijsku analizu (OLAP) tipa „stopa prekida ugovornih odnosa u posljednjih nekoliko mjeseci po regijama, po segmentima ...“.

Ovaj potez bi vam zasigurno dao neke dodatne informacije i rasvijetlio određene aspekte trendova koje želite razumjeti. Pretpostavimo da ste otkrili kako je rastuća stopa prekida ugovornih odnosa zabilježena na području sjevera Hrvatske i to od strane pretplatnika. Sada biste vjerojatno tražili generiranje čitavog niza izvještaja baziranog na ovom specifičnom otkrivenom segmentu. Generalno govoreći, radili biste *ad hoc* reporting vjerojatno pomoću OLAP-a nad tako izdvojenim setom podataka koji se je pokazao najučestaliji prilikom prekida ugovornih odnosa. Nakon niza izgeneriranih izvještaja na zahtjev ovi postupci bi vam možda dodatno rasvijetlili situaciju i oboružani takvim spoznajama krenuli biste u razradu strategije sprečavanja daljnjih prekida ugovornih odnosa.

Zanatski odrađen posao

Na prvi pogled, sve je korektno odrađeno, no da li stvarno tako?

Princip u kojem promatrate koncentracije s obzirom na neki incijalni kriterij (npr. stopa prekida ugovornih odnosa s obzirom na regiju) može biti upitan. Pretpostavimo da se koncentrirate na regiju koja ima pretpostavimo najvišu stopu prekida ugovornih odnosa od recimo 3 % mjesečno, dok ostale regije imaju stopu prekida ugovornih odnosa u rasponu od 0,1-0,2 % što vam nije alarmantno . Pažnju će vam vjerojatno privući regija sa najvišom stopom prekida ugovornih odnosa, te ćete vjerojatno dalje nastaviti analizirati tu regiju. Možda ćete promatrati profile korisnika koji su prekinuli ugovorne odnose unutar te regije, možda ostale korištene produkte i slično.

U čemu je problem ?

Osnovni problem proizlazi iz činjenice što ste temeljno particioniranje proveli uz određenu pretpostavku kako je određeni kriterij u ovom slučaju – regija značajan za prekid ugovornih odnosa. Sa druge strane, uočivši povećanu koncentraciju unutar jedne regije maksimalno ste se zaokupili sa tom regijom zanemarujući ostalu populaciju. Možda je koncentracija unutar te regije značajno viša, u odnosu na ostale regije, no ako sumarno promatrate stopu prekida unutar ostalih regija, tada ona na ostatku populacije nosi puno veću težinu.

Tradicionalni pristup analitici pomoću OLAP-a i reportinga podrazumijeva eksperta/analitičara kao glavnog scenarista analize. To ne mora nužno biti loše, no da li postoji bolja alternativa.

Zašto dodatno ne uposliti strojeve ?

Prikazni / uobilčajeni pristup oslanja se na intuiciju eksperta, alternativni pristup bio bi korištenje data mining algoritama, koji mjere utjecaj pojedinih varijabli tipa regija, trajanje ugovornog odnosa, iznos računa i sl. na prekid ugovornih odnosa. Na taj način moguće je istestirati utjecaj velikog broja varijabli (pretpostavimo stotinjak i više) na prekid ugovornog odnosa.

Vratimo se na naš prethodni primjer, u kojem nam je teoretski moglo promaknuti kako najveći utjecaj na prekid ugovornog odnosa ima stagnacija u korištenju ugovorene usluge, korisnici dospjeli putem specifičnih kampanji unutar određenih dobnih skupina. Ove karakteristike ne moraju biti dominantne samo unutar tradicionalnim načinom uočene regije. To može biti pravilo koje se proteže kroz cijeli portfelj, a zavarni činjenicom male stope prekida ugovornih odnosa na razini određenih regija, percipiramo kako nam one nisu interesantne za daljnju analizu.

Zlatni oblutci skrivenog znanja

Zajedničke karakteristike u ovakvim situacijama, mogu predstavljati plamičak koji na duži rok može buknuti poput plamena, te regionalno niska stope prekida ugovornih odnosa ne znači da te regije ne pokazuju pravilnosti po nekim drugim kriterijima s obzirom na prekide ugovornih odnosa. Upravo u uvoju činjenici leži snaga primjene data mining metoda- fokus na bitno, a ne ono što se na prvi pogled čini bitno. Temeljem otkrivenih bitnih karakteristika, nad cjelokupnim portfeljem korisnika moguće je puno kvalitetnije razviti poslovnu strategiju smanjivanja trenda prekida ugovornih odnosa. Umjesto nagađanja što je bitno uz neizvjestan ishod, u prilici smo natjerati strojeve da to odrade puno efikasnije od nas.

Gdje su prednosti ?

Prednost ovog pristupa, očituje se u činjenici kako je u kratkom vremenu, algoritamaski moguće izmjeriti utjecaj velikog broja faktora na prekide ugovornih odnosa, te se fokusirati se na bitne

aspekte u nastojanju razrade strategije sprečavanja ovih trendova. Sa druge strane, na temelju takvih proračuna, moguće je kreirati prediktivne modele, koji će računati vjerojatnosti prekida ugovornog odnosa za svaki pojedini subjekt, te tako izdvojiti najrizičnije i najprofitabilnije kao ciljnu skupinu prema kojoj treba preventivno djelovati.

Ovakav pristup omogućava maksimalno iskorištavanje ekspertnog znanja prilikom generiranja izvedenih varijabli potencijalno relevantnih za mjerenje prekida ugovornih odnosa. Ova tvrdnja proizlazi iz činjenice kako je mjerenje utjecaja automatizirano. Uloga analitičara svodi se na poznavanje poslovne strane problema, primjerice kakve su bile promjene u strukturi portfelja korisnika, da li se značajno izmjenila struktura korisnika i kada kako bi se utjecaji na prekide ugovornih odnosa promatrali prema primjerice tim segmentima. Dodatne aktivnosti analitičara bile bi usmjerene na osmišljavanje potencijalno korisnih (izvedenih) biheviorističkih varijabli poput intenzitet korištenja usluge i slično. U konačnici se mjere utjecaji i ovih varijabli na prekid ugovornih odnosa s ciljem pronalaženje uzroka temeljnog problema. Iskustvo je pokazalo kako upravo biheviorističke varijable nose najviše informacija vezane uz prekide ugovornih odnosa.

Da li ste spremi prihvatiti izazov ?

Tradicionalni pristup analitici koji se temelji na data warehouse sustavima, svakako može biti koristan za donošenje poslovnih odluka, posebice u segmentu operativnog odlučivanja. Korištenje data mining metoda i tehnologije, doprinosi tzv. „efektu maksimalnog cijedenja naranče“. To bi jednostavno rečeno značilo, kako tradicionalni sustavi reportinga i OLAP još uvijek ostavljaju podosta podatkovnog skoka u informacijskoj naranči, i to onog najhranjivijeg, koji i te kako može osnažiti poslovni organizam. Problem je jedino u tome što ponekad moramo srušiti određene mitove kako bismo krenuli dalje nekim novim „revolucionarnijim“ putevima u poslovnoj analitici. Još je Krleža ustvrdio kako revolucija nije mit nego pobjeda nad mitom. Upravo ta spoznaja može nas motivirati da poslovnu analitiku uvedemo u jednu novu produktivniju dimenziju.