

**UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA**

DIPLOMSKO DELO

MAJA SANCIN

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

DIPLOMSKO DELO

**UPORABA PODATKOVNEGA RUDARJENJA PRI ANALIZI NAKUPNIH NAVAD
KUPCEV BENCINSKIH SERVISOV PODJETJA PETROL D.D. LJUBLJANA**

Ljubljana, avgust 2008

MAJA SANCIN

IZJAVA

Študentka MAJA SANCIN izjavljam, da sem avtorica tega diplomskega dela, ki sem ga napisala pod mentorstvom MAG. ALEŠA POPOVIČA, in dovolim njegovo objavo na fakultetnih spletnih straneh.

V Ljubljani, dne 21.8.2008

Podpis: _____

KAZALO

UVOD.....	1
1 OPREDELITEV IN POMEN PODATKOVNEGA RUDARJENJA	2
2 POMEMBNI POJMI	3
3 ŠEST NALOG PODATKOVNEGA RUDARJENJA	4
3.1 Klasifikacije.....	4
3.2 Ocenjevanje	4
3.3 Napovedovanje	4
3.4 Opisovanje.....	5
3.5 Asociacije	5
3.6 Razvrščanje v skupine	5
4 PODROČJA UPORABE PODATKOVNEGA RUDARJENJA	6
4.1 Podatkovno rudarjenje v poslovnih vedah	7
4.1.1 Izdelava profilov kupcev	7
4.1.2 Segmentacija	8
4.1.3 Odzivnost strank.....	8
4.1.4 Upravljanje s tveganji.....	8
4.1.5 Aktiviranje kupca	8
4.1.6 Navzkrižna prodaja in modeli »več enakih izdelkov«	8
4.1.7 Analiza nakupne košarice.....	9
4.1.8 Obraba ali »mešalnik«.....	9
5 TEHNIKE PODATKOVNEGA RUDARJENJA	9
5.1 Najbližji sosed	9
5.2 Asociativna pravila.....	10
5.3 Odločitveno drevo	11
5.4 Razvrščanje v skupine	12
5.5 Umetne nevronske mreže	12
6 PROCES PODATKOVNEGA RUDARJENJA.....	13
7 ANALIZA PODATKOV IZ ANKETE PODJETJA PETROL d.d.	15
7.1 Opis podjetja Petrol d.d.	15
7.2 Opis ankete	16
7.3 Orodje Visual Studio 2005	17
7.4 Analiza podatkov.....	18
SKLEP	22
LITERATURA IN VIRI	24
PRILOGE	

KAZALO SLIK

Slika 1: Odločitveno drevo za ugotavljanje kreditne sposobnosti.....	11
Slika 2: Preprosta nevronska mreža	13
Slika 3: Proces podatkovnega rudarjenja (CRISP-DM model).....	14
Slika 4: Profil najboljših strank	20

KAZALO PRILOG

PRILOGA 1: Vprašalnik	1
PRILOGA 2: Seznam podatkov iz baze podatkov	4
PRILOGA 3: Demografski podatki anketirancev	7
PRILOGA 4: Nakupno obnašanje anketirancev.....	9
PRILOGA 5: Analiza vpliva demografskih in socioekonomskih značilnosti kupcev na potrošnjo	10
PRILOGA 6: Analiza profila najboljših kupcev	12
PRILOGA 7: Analiza vpliva mnenj strank z različnimi nakupnimi obnašanji o kakovosti za Petrol	13
PRILOGA 8: Asociativna pravila: vpliv mnenja o zaupanju prodajalcem na potrošnjo samo goriva.....	14
PRILOGA 9: Slovarček slovenskih prevodov tujih izrazov	15

UVOD

Zmožnosti za ustvarjanje in shranjevanje podatkov se iz dneva v dan povečujejo. Eksplozivna rast shranjenih podatkov in čedalje ostrejša konkurenca v globaliziranem svetu ustvarjata potrebo po novih tehnikah in avtomatiziranih orodjih, ki nam lahko pomagajo pri pretvarjanju obsežne količine podatkov v informacije in znanje za učinkovito odločanje. Pri tem nam je v veliko pomoč podatkovno rudarjenje, s katerim odkrivamo informacije, vzorce in povezave v velikih količinah podatkov. Pridobljeno znanje lahko nato uporabimo za boljše (poslovno) odločanje in za zagotavljanje konkurenčne prednosti.

V diplomskem delu obravnavam podatkovno rudarjenje in njegovo uporabnost. Predstavila bom naloge in glavne tehnike podatkovnega rudarjenja, nato pa s pomočjo orodja Visual Studio 2005 analizirala podatke iz ankete o zadovoljstvu strank bencinskih servisov podjetja Petrol d.d. Ljubljana (v nadaljevanju Petrol) ter opisala postopke in ugotovitve. V največji meri se bom usmerila v najboljše kupce Petrola, saj ti dajejo največ dobička. Poleg tega bom pozorna na vplive, ki so pomembni za večjo ali dodatno potrošnjo.

Namen diplomskega dela je v analizi literature s področja podatkovnega rudarjenja, uporabi pridobljenega znanja na primeru iz prakse in s tem v pridobivanju uporabnih informacij iz baze podatkov podjetja Petrol. Cilj diplomskega dela je na podlagi teoretičnega znanja analizirati bazo podatkov Petrola in pridobiti koristne informacije, ki so lahko podlaga za nadaljnje odločanje (predvsem na področju trženja in prodaje) v podjetju Petrol. Pri tem je moj cilj analizirati ciljne kupce in vplive na večjo potrošnjo. Na podlagi tega lahko podjetje sestavi trženjski splet in se usmeri v bolj donosne kupce.

Raziskovalni pristop gre pri mojem delu iz teoretičnega na praktično in iz splošnega na specifično. Najprej sem analizirala splošno literaturo o podatkovnem rudarjenju, nakar sem za vsako področje pregledala specifične članke (iskala sem jih v znanstvenih bazah člankov). Teoretično znanje sem nato skušala tudi praktično uporabiti.

V praktičnem delu diplomske naloge sem si postavila nekaj vprašanj s sovpadajočimi hipotezami, pri čemer sem želela odkriti prave odgovore na zastavljena vprašanja. Prva hipoteza pri odkrivanju vpliva demografskih in socioekonomskih spremenljivk na potrošnjo je, da te vplivajo na potrošnjo na prodajnih mestih Petrola. Druga hipoteza pa je, da v največji meri vpliva na potrošnjo dohodek gospodinjstva. Hipoteza pri drugem vprašanju je rezultat prejšnjega vprašanja: tipični najboljši kupec je samski, nima vzdrževanih članov in je moškega spola. Pri tretjem vprašanju je hipoteza ta, da boljše mnenje o kakovosti Petrolovih prodajnih mest vpliva na nakup tudi izdelkov iz dodatne ponudbe Petrola.

Diplomska naloga je razdeljena na sedem poglavij, pri čemer jih je prvih šest teoretičnih, sedmo poglavje pa predstavlja praktičen del diplomske naloge. V prvem poglavju opredeljujem pojem podatkovnega rudarjenja in razložim, zakaj ima dandanes tako velik

pomen. Drugo poglavje posvečam razlagi pomembnih pojmov in izrazov. V tretjem poglavju opisujem šest nalog podatkovnega rudarjenja. V četrtem poglavju predstavim razne primere uporabe podatkovnega rudarjenja v znanosti. Največji poudarek dam opisu uporabnosti podatkovnega rudarjenja v poslovnem svetu. V petem poglavju opredelim pet glavnih tehnik podatkovnega rudarjenja, s katerimi lahko opravljamo naloge podatkovnega rudarjenja. Šesto poglavje namenjam procesu podatkovnega rudarjenja na podlagi CRISP-DM modela. V sedmem poglavju predstavim celotno analizo nakupnih navad kupcev Petrolovih bencinskih servisov, pri čemer najprej opišem podjetje, anketo in uporabljeno orodje podatkovnega rudarjenja, nato pa samo izvedbo in ugotovitve analize.

1 OPREDELITEV IN POMEN PODATKOVNEGA RUDARJENJA

Podatkovno rudarjenje lahko opredelimo na več načinov, saj gledajo razni avtorji na to iz različnih zornih kotov. Han in Kamber (2000, str. 6) na primer trdita, da je podatkovno rudarjenje izkopavanje znanja iz velike količine podatkov. Hand, Mannila in Smith (2001, str. 6) menijo, da je podatkovno rudarjenje »analiza (pogosto velike) serije opazovalnih podatkov, ki je namenjena pridobivanju nepričakovanih povezav in povzemanju vsebine podatkov na nove načine, ki so jasni in uporabni lastniku podatkov«. Cabena, Hadjinian, Stadler, Verthes in Zanassi (1998, str. 12) definirajo podatkovno rudarjenje kot interdisciplinarno vedo, ki združuje tehnike strojnega učenja, prepoznavanja vzorcev, statistike, baz podatkov in vizualizacije, da bi izvela informacije iz velikih baz podatkov.

Če povzamem razne definicije, je podatkovno rudarjenje sistematično in avtomatizirano iskanje informacij, vzorcev, povezav in trendov z raziskovanjem v veliki količini podatkov. Ciljne informacije, ki jih želimo s tem pridobiti, morajo biti tehtne, uporabne in morajo dati neko konkurenčno prednost. Pred analizo take informacije niso očitne, vidne ali poznane. Analitik torej skuša, s pomočjo podatkovnega rudarjenja, pretvoriti množice podatkov v pomembno znanje, ki bo uporabno za (poslovno) odločanje in ki bo pomagalo pridobivati nove priložnosti. Podatkovno rudarjenje se opira na statističnih metodah, strojnem učenju (ang. *Machine Learning*), umetni inteligenci (ang. *Artificial Intelligence*, AI) in na vizualizacijskih orodjih (ang. *Data Visualization Techniques*) (Sumathi, 2006, str. 5-10).

Velik pomen podatkovnega rudarjenja danes in v prihodnosti ponazarja že dejstvo, da količina zbranih podatkov eksponentno narašča. Pri tem si lahko postavimo vprašanje, kaj se posameznik lahko sploh nauči iz teh podatkov. Naraščajoče in ogromne količine podatkov so neobvladljive le z uporabo tradicionalnih načinov analiziranja podatkov. Zato so orodja za podatkovno rudarjenje pri odkrivanju znanja iz podatkov v veliko pomoč (Dürr, 2003, str. 3).

Larose (2005, str. 4) našteva naslednje razloge za znatno razvijanje področja podatkovnega rudarjenja:

- eksponentno večanje zbiranja podatkov;
- podjetja uporabljajo podatkovna skladišča, ta pa predstavljajo zanesljiv vir podatkov;
- naraščajoča dostopnost in uporaba interneta;

- naraščajoči pritisk tekmovalnosti v globaliziranem svetu;
- napredki v tehnologiji in povečanje prostora na trdem disku.

2 POMEMBNI POJMI

S podatkovnim rudarjenjem ustvarjamo modele ali ugotavljamo vzorce. **Model** je splošna predstavitev izvirnih podatkov na visoki ravni. Opredelimo ga na podlagi določenih parametrov, ki jih lahko ocenimo z danimi podatki. Napovedovalne modele lahko uporabljamo za napovedovanje ali klasifikacijo vrednosti opazovane spremenljivke, opisne modele pa za razvrščanje v skupine in za opisovanje (Laxman & Sastry, 2006, str. 175).

Medtem ko ima model splošen značaj, so vzorci specifični. **Vzorec** je lokalna struktura, ki daje specifično izjavo o nekaj spremenljivkah ali nekaj podatkih (Laxman & Sastry, 2006, str. 175). Vzorec nam na primer pokaže kombinacijo dogodkov z visoko korelacijo.

Poleg razlike med modelom in vzorcem je pri podatkovnem rudarjenju potrebno dobro poznavanje spremenljivk. **Spremenljivka** (ali atribut) je lastnost opazovane enote. Pomembno je, da analitik primerno določi vrste spremenljivk, da jih lahko potem poznajo tudi algoritmi strojnega učenja. Le tako lahko delajo pravilne primerjave, se naučijo pravih konceptov itd. Poleg tega lahko na tak način samostojno obravnavajo manjkajoče vrednosti in iščejo veljavne vrednosti.

Spremenljivke lahko ločimo po tem, koliko informacij vključujejo v sebi (Pyle, 2003, str. 314):

- **nominalne spremenljivke** so le označbe ali imena, uporabljena za identificiranje posameznih entitet. Na primer so lahko to imena, serijske številke, kode itd. S takimi spremenljivkami ne moremo računati.
- **kategorične (diskretne) spremenljivke** identificirajo skupino entitet. Entitete, ki spadajo v določeno kategorijo, imajo določene skupne lastnosti, ki so značilne za to skupino. Na primer spadajo vsi študentje v kategorijo ljudi.
- **ordinalne spremenljivke** so kategorije, ki jih lahko racionalno naštejemo v določenem vrstnem redu. Primer za tako kategorijo vključuje vrstni red vroče > milo > hladno. Pri tem poznamo vrstni red, ne poznamo pa točne razdalje med vrednostmi. Teh vrednosti zato ne moremo seštevati ali odštevati.
- **intervalne spremenljivke** so ordinalne spremenljivke, pri katerih lahko tudi merimo razdaljo med vrednostmi. Torej niso le razvrščene, a so tudi merjene v fiksni in enakih enotah. Primer intervalne spremenljivke je temperatura v stopinjah Celzija. Računanje razlike takih vrednosti je smiselno, seštevaje pa ni.
- **razmernostne spremenljivke** so spremenljivke, pri katerih obstaja veljavno razmerje in neka spodnja meja. S takimi števili ravnamo kot z realnimi števili in torej lahko z njimi računamo. Razdalja je dober primer razmernostne spremenljivke.

Spremenljivke klasificiramo tudi na podlagi tega, ali so enostavne ali binarne. **Enostavne spremenljivke** (ang. *Unary Attributes*) v celoti označujejo lastnosti posameznega segmenta. Nekatere enostavne spremenljivke definiramo po videzu segmenta. Primer takih atributov so barve (rdeča, zelena, modra itd.), splošna struktura (peščena, zrnata itd.) ali oblika (okrogla, kvadratna itd.). Vsi rdeči segmenti imajo podoben videz (ne glede na obliko) in vsi okrogli segmenti imajo podobno obliko (ne glede na videz). **Binarne spremenljivke** (ang. *Binary Attributes*) so bolj kompleksne, saj je njihov element sestavljen iz dveh segmentov. Primer so zebrine črno-bele črte, ki so sestavljene iz dveh segmentov, ki delijo podobni videz in obliko (Ferrari & Ziesserman, 2007, str. 3).

3 ŠEST NALOG PODATKOVNEGA RUDARJENJA

S šestimi nalogami podatkovnega rudarjenja, ki jih opisujem v tem poglavju, lahko rešujemo veliko ekonomskih, intelektualnih, poslovnih in drugih problemov.

V splošnem lahko razdelimo naloge podatkovnega rudarjenja v dve skupini: v usmerjene in v neusmerjene naloge. Usmerjene naloge imajo kot cilj na podlagi preteklih podatkov odkriti vrednost določene ciljne spremenljivke. V to skupino spadajo klasifikacije, ocenjevanja in napovedovanja. Neusmerjene naloge navajajo lastnosti podatkov, ki so v bazi podatkov. Pri tem ne ciljamo na določeno posebno spremenljivko, a želimo razkriti strukturo v podatkih. (Berry & Linoff, 2004, str. 8). V to skupino nalog spadajo asociacije in razvrščanje v skupine.

3.1 Klasifikacije

Klasificiranje (ang. *Classification*) je proces pridobivanja skupine modelov (ali funkcij), ki opisujejo in ločijo razrede podatkov ali koncepte. Cilj klasificiranja je uporabiti model za napovedovanje razreda za nove prej neopredeljene objekte. Model razredov pridobimo s pomočjo podatkov namenjenih učenju (ang. *Training Data*). Prikažemo ga lahko v raznih oblikah, kot na primer v obliki »če – potem« pravil, v obliki odločitvenega drevesa, nevronske mreže ali matematične formule (Han & Kamber, 2000, str. 15). Primer klasifikacije je delitev kandidatov za kredit v nizko, srednje ali visoko rizične.

3.2 Ocenjevanje

Ocenjevanje (ang. *Estimation*) in klasificiranje sta si zelo podobni. Razlika je le v tem, da je pri ocenjevanju ciljna spremenljivka numerična in ne kategorična. Prihodnje vrednosti ocenjujemo na podlagi vrednosti zgodovinskih podatkov (Larose, 2005, str. 12). Poudarek je na porazdelitvenih trendih, ki se opirajo na razpoložljive podatke (Han & Kamber, 2000, str. 15). Za ocenjevanje vrednosti sta najbolj primerni tehniki podatkovnega rudarjenja nevronske mreže in regresijski modeli (Berry & Linoff, 2004, str. 10).

3.3 Napovedovanje

Napovedovanje (ang. *Prediction*) je zelo podobno klasificiranju in ocenjevanju. Pri napovedovanju klasificiramo na podlagi nekega napovedanega prihodnjega obnašanja ali na

podlagi napovedane prihodnje vrednosti. Edini način preveritve pravilnosti klasifikacije je čakanje, da se napovedano dejansko zgodi ali ne (Berry & Linoff, 2004, str. 10). Primer je napovedovanje povečanja prometnih nesreč s smrtnim izidom, če se bodo v prihodnjem letu povečale omejitve hitrosti (Larose, 2005, str. 13).

Napovedovanje ločujemo od klasificiranja in ocenjevanja zato, ker so pri napovedovanju potrebne dodatne aktivnosti za določanje časovne povezave med vhodno in ciljno spremenljivko. Zgodovinske podatke uporabimo za izgradnjo modela, ki razlaga sedanje opazovano obnašanje. Ko v model damo sedanje vhodne podatke, dobimo napoved za prihodnje obnašanje (Berry & Linoff, 2004, str. 10).

3.4 Opisovanje

Včasih je cilj podatkovnega rudarjenja enostavno razumevanje ljudi, izdelkov ter drugih razredov in konceptov. Za to je najbolje, da jih opisujemo na kratek, jedrnat in precizen način. Taka opisovanja (ang. *Description*) lahko dobimo s karakterizacijo, tako da opišemo ciljni razred. Lahko pa ga tudi primerjamo z drugimi primerjalnimi razredi. Karakterizacija je strnjen opis glavnih značilnosti ciljnega razreda podatkov. Podatke le specifičnega razreda lahko dobimo s poizvedbo v bazi podatkov. Izhod opisovanja je lahko predstavljen v več oblikah: v grafih v obliki tortic, v stolpčnih diagramih, linijskih diagramih, večdimenzionalnih kockah podatkov, v večdimenzionalnih tabelah, v obliki pravil itd (Han & Kamber, 2000, str. 13-14).

3.5 Asociacije

Analiza asociacij (ang. *Association Analysis*) pomeni odkrivanje spremenljivk, ki »gredo skupaj«. Pri tem želimo izmeriti zvezo med dvema ali več spremenljivkami. Asociativna pravila, ki jih bom bolje predstavila v naslednjem poglavju, so v »če – potem« obliki in imajo še mero podpore in zaupanja (Larose, 2005, str. 17). Z razliko od klasificiranja lahko uporabimo asociativna pravila tudi v primerih, ko ni označenih nobenih razredov in ko nobena struktura ne velja za zanimivo. Z asociativnimi pravili napovedujemo vrednosti (ne le razredov) in lahko več kot eno vrednost atributa naenkrat. Zato obstaja več asociativnih pravil kot pravil klasificiranja. Ker bi bilo asociativnih pravil preveč in bi nekatera bila neuporabna, nastavimo omejitve podpore in zaupanja spremenljivke in s tem določimo prag smiselnosti pravila (Spletni tečaj podatkovnega rudarjenja KDnuggets, 2008)

Analizo asociacij uporabljamo predvsem pri analizi nakupne košarice ali pri analizi podatkov iz transakcij (Han & Kamber, 2000, str. 13-14).

3.6 Razvrščanje v skupine

Razvrščanje v skupine (ang. *Clustering*) je segmentiranje heterogene populacije v več homogenih skupin. Razlika med razvrščanjem v skupine in klasifikacijami je ta, da pri razvrščanju v skupine želimo analizirati objekte, za katere ne poznamo razredov. Pri klasifikacijah spada vsak objekt v določen razred, ki ga dobimo s pomočjo modela, ki smo

ga razvili s podatkovnim rudarjenjem po zgodovinskih podatkih. Pri razvrščanju v skupine nimamo predhodno določenih razredov ali primerov. Oblikujemo jih na podlagi medsebojne podobnosti tako, da razvrščamo podatke na podlagi dimenzij (kot sta na primer starost ali regija), ki jih sami določimo. Tako dobimo skupine predmetov (ali ljudi), ki imajo podobne značilnosti (Berry & Linoff, 2004, str. 11). Cilj razvrščanja v skupine ni klasificirati ali oceniti določeno ciljno vrednost na podlagi modela, a je cilj le segmentirati podatke (Larose, 2005, str. 16).

Naloga razvrščanja v skupine je ponavadi prvi korak v raziskovanjih s pomočjo podatkovnega rudarjenja. Rezultate te naloge uporabljajo potem analitiki kot vhode za nadaljnja raziskovanja z različnimi tehnikami podatkovnega rudarjenja (Larose, 2005, str. 17).

4 PODROČJA UPORABE PODATKOVNEGA RUDARJENJA

Tehnike podatkovnega rudarjenja uporabljajo analitiki iz več področij znanosti: iz poslovnih ved, astronomije, medicine, molekularne biologije, geologije in še bi lahko naštevala.

V medicini lahko podatkovno rudarjenje uporabljamo, na primer za raziskovanje bolezni, določanje zdravil bolnikom in diagnostiko. Ramon et al. (2007, str. str. 243–256) v svojem članku opisujejo uporabnost podatkovnega rudarjenja pri bolnikih, ki so potrebni intenzivne nege. Pri njih so vitalne funkcije pod stalno kontrolo in v primeru nevarnosti jih podpirajo zdravljenja in mehanske naprave. Avtorji trdijo, da bi orodja za podatkovno rudarjenje lahko pomagala kliničnim zdravnikom za hitrejše odkrivanje problemov pri pacientih. Izziv bi bil v prepoznavanju vzorcev, ki naznanjajo na posamezne probleme (torej v odkrivanju problema in opisovanju diagnoze), in v čimprejšnjem napovedovanju vzorcev v veliki količini podatkov. Na podlagi tega bi lahko takoj pravilno ukrepali.

Podatkovno rudarjenje je koristno tudi pri ugotavljanju prevar. Na primer lahko, s pomočjo podatkovnega rudarjenja, razkrijemo zlorabe kreditnih kartic. Nekatere transakcije s kreditnimi karticami so neveljavne in pri teh ni problemov. Obstajajo pa tudi transakcije, ki so formalno korektne, a gre v bistvu za sleparije. S povečanjem števila izdanih kreditnih kartic in z napredki v elektronskem poslovanju postaja ugotavljanje prevar čedalje bolj pomembno. Pri tem znanje strokovnjakov ni dovolj, saj se metode prevar spreminjajo in je zato potrebno stalno učenje. Pri detekciji prevar s kreditnimi karticami so umetne nevronske mreže najbolj uspešna tehnika podatkovnega rudarjenja, ki zasači zlorabe na podlagi učenja po zgodovinskih podatkih. Uporabna so tudi asociativna pravila, s katerimi lahko izvemo, v katerih primerih je verjetnost prevare največja (Brause, Langsdorf & Hepp, 1999, str. 103-106).

Tudi na področju ugotavljanja in napovedovanja kriminalnih dejanj so orodja za podatkovno rudarjenje v veliko pomoč. Uporabljajo jih zvezne agencije (kot na primer ameriška CIA in FBI) v borbi zoper prometnim prekrškom, spolnim zlorabam, kraji, prevaram, požigom,

preprodaji mamil in nasilnim zločinom. S tehniko asociativnih pravil in z razvrščanjem v skupine skušajo na primer identificirati kriminalce, ki na internetu širijo ilegalna sporočila. Z uporabo nevronske mreže pa lahko napovejo zločine tako, da analizirajo podatke, kot so čas, lokacija, prevozno sredstvo, naslov, fizične značilnosti in lastnina (Hisnchun, Wingyan, Xu, Yi Qin & Chau, 2004, str. 50-55).

Na znanstvenem področju ima podatkovno rudarjenje velik pomen. V biologiji pomaga na primer pri identifikaciji na tisoče genov v vsakem genomu. V biomedicini uporabljajo podatkovno rudarjenje za analizo kemičnih lastnosti zdravil in s tem pospešujejo odkritja na tem področju ter znižujejo stroške raziskovanj. V astronomiji so ustvarili popolnoma avtomatiziran sistem katalogiranja oddaljenih nebeških objektov, ki jih z vizualnim nadzorom ne moremo zaznati. Tehnike podatkovnega rudarjenja so se izkazale na vseh zgoraj omenjenih področjih (in še na drugih) in s svojo »spretnostjo« konkurirale strokovnjakom ali jih celo prekosile (Witten & Frank, 2005, str. 28).

Ker je diplomsko delo usmerjeno v uporabo podatkovnega rudarjenja na področju poslovnih ved, bom to podrobneje obravnavala v naslednjem podpoglavju. Znanje o primerih uporabe podatkovnega rudarjenja je nujno za določanje cilja analize in za izvedbo analize. V splošnem podjetja uporabljajo podatkovno rudarjenje za odkrivanje pomembnih informacij, s katerimi lahko bolje sprejemajo odločitve, zmanjšajo tveganja in ustvarjajo konkurenčno prednost.

4.1 Podatkovno rudarjenje v poslovnih vedah

Podatkovno rudarjenje je v poslovnih vedah zelo uporabno. Največjo dodano vrednost ima v splošnem na področju trženja in prodaje, pri katerem lahko veliko količino podatkov pretvorimo v uporabno znanje. Vseeno so tehnike pomembne tudi na drugih področjih (na primer v financah in zavarovalništvu). V nadaljevanju predstavljam več primerov uporabe podatkovnega rudarjenja v poslovnem svetu.

4.1.1 Izdelava profilov kupcev

Poznavanje stranke (kupca) je dandanes zelo pomembno. Izdelava profilov kupcev (ang. *Profile Analysis*) je odličen način spoznavanja strank in vključuje določitev skupnih značilnosti populacije, za katero se zanimamo. Pri tem vzamemo v poštev razne demografske značilnosti (kot na primer spol, starost, zakonski stan itd.) in druge karakteristike, ki so pomembne za posel (na primer povprečno višino prodaje). Profil stranke lahko ugotavljamo z analizo odvisnosti, z razvrščanjem v skupine in z opisovanjem konceptov. Na podlagi profila stranke se tržnik odloči o strategijah in taktikah zadovoljevanja strank. Izdelovanje profila stranke je posebej uporabno pri ciljnih tržnih segmentih (Shaw, Subramaniam, Tan & Welge, 2001, str. 133).

4.1.2 Segmentacija

Segmentirati trg pomeni »določiti in opisati različne skupine kupcev, ki imajo različne potrebe in preference« (Kotler, 2003, str. 279). Segmentacija (ang. *Segmentation*) je pomembna za osnovno razumevanje sedanjih kupcev. Stranke lahko razdelimo na podlagi ene ali več dimenzij. Podobnosti v vsakem segmentu naznanjajo podobna nakupna obnašanja. Informacije o segmentih so uporabne za učinkovito doseganje vseh kupcev ali za ciljno trženje (Huang, Tzeng & Ong, 2007, str. 313).

4.1.3 Odzivnost strank

Prvi korak pri povečanju učinkovitosti trženjskih kampanj s ciljnim trženjem je izdelava modela odzivnosti strank (ang. *Response Model*). S tem želimo predvideti, kdo bo dovzeten na določeno ponudbo izdelka ali storitve. Model lahko naredimo na podlagi preteklih podatkov o obnašanju strank, ali na podlagi podobnih nadomestnih podatkov (Parr Rud, 2001, str. 7-11).

4.1.4 Upravljanje s tveganji

Na področju upravljanja s tveganji (ang. *Risk Management*) lahko s pomočjo podatkovnega rudarjenja napovemo verjetnost, da bo določeno pričakovanje izpolnjeno ali ne (Parr Rud 2001, str. 7). Na primer lahko v zavarovalništvu napovemo možne izgube, ki nam jih bo povzročala posamezna stranka. Na tak način lahko posamezniku zaračunamo različno zavarovalno premijo glede na njegovo tveganost (ang. *Variable Pricing*). V bančnem sektorju lahko segmentiramo posojilojemalce glede na stopnjo tveganja, na podlagi predhodnih posojil (Zanasi, 2003, str. 7).

4.1.5 Aktiviranje kupca

Modeli aktiviranja napovedujejo, ali bo neka oseba postala aktivna stranka ali ne (Parr Rud 2001, str. 7). Aktiviranje strank (ang. *Activation*) poteka tako, da podjetje najprej sledi potencialni stranki v vseh korakih, ki so potrebni, da bo ta postala aktivna stranka. Velika razlika je na primer v tem, ali je oseba le obiskovalec spletne strani določene banke ali odpre kasneje račun in še vplaga v delnice. V vsakem koraku podjetje definira sprožila, ki omogočajo kontaktiranje strank v pravem času, s pravim sporočilom in po pravi poti (internetu, elektronski pošti, klicnem centru ali pošti). Na tak način podjetje avtomatizira dialog s potencialnimi kupci, kar poveča verjetnost aktivacije stranke (Million Handshakes, 2008)

4.1.6 Navzkrižna prodaja in modeli »več enakih izdelkov«

Modeli navzkrižne prodaje (ang. *Cross-Sell Models*) napovedujejo verjetnost, da bo določena stranka kupila še kak drugi izdelek ali storitev istega podjetja. Z modeli »več enakih izdelkov« (ang. *Up-Sell Models*) želimo napovedati verjetnost, da bo stranka kupila večkrat enak proizvod ali storitev podjetja. Z uporabo tovrstnih modelov želi podjetje ohraniti sedanje kupce, kar je najlažji način za ustvarjanje prihodkov (Parr Rud, 2001, str. 7-11).

Navzkrižno prodajo je na primer učinkovito uporabilo podjetje Tesco iz Velike Britanije, ki je v glavnem trgovsko podjetje, a se je v koncu 20. stoletja diverzificiralo (vstopilo je tudi na področje financ). S pomočjo kartice zvestobe so v podjetju zbirali podatke iz transakcij v prodajalnah. Podatke so potem analizirali s pomočjo podatkovnega rudarjenja in pridobljene informacije uporabili tudi pri prodaji finančnih storitev. Navzkrižna prodaja jim je omogočila pridobivanje strank za finančne storitve s 50% nižjimi stroški od drugih finančne institucij (Mukund, 2003, str. 9).

4.1.7 Analiza nakupne košarice

Izbira nakupne košarice je odločitveni proces, pri katerem kupec izbere artikle iz množice kategorij izdelkov v istem priходу po nakupih. Ključna značilnost izbire nakupne košarice je v medsebojni odvisnosti zelenih zvez med artikli v končni košarici. Z analizo nakupne košarice (ang. *Market Basket Analysis*) želimo izvedeti, katere izdelke kupci kupujejo skupaj (Russel & Petersen, 2000, str. 367–368). Take informacije lahko potem podjetje uporablja za več namenov: pri načrtovanju razporeditve izdelkov, za dodeljevanje popustov le na enega od izdelkov, ki se kupujejo skupaj, za dajanje kuponov pri nakupu ujemajočega izdelka in tako dalje (Witten & Frank, 2005, str. 27).

4.1.8 Obraba ali »mešalnik«

Veliko podjetij vabi kupce s cenejšimi začetnimi izdelki ali storitvami, nakar te podražijo in upajo na lojalnost kupcev (na primer se to dogaja na področju telekomunikacij). Kupci večkrat niso lojalni in hodijo od konkurenta do konkurenta. Obraba (ang. *Attrition*) pomeni zmanjšanje uporabe izdelka ali storitve. »Mešalnik« (ang. *Churn*) pa je zamenjava ponudnika izdelka ali storitve (Parr Rud 2001, str. 11). S pomočjo podatkovnega rudarjenja lahko podjetja zaustavijo odhajanje kupcev h konkurentom tako, da vnaprej odkrijejo določene spremembe v vzorcih, ki naznanjajo na možen odhod kupca, in ga nato skušajo obdržati z določenimi akcijami (Witten & Frank, 2005, str. 27).

5 TEHNIKE PODATKOVNEGA RUDARJENJA

Obstaja več tehnik (algoritmov oziroma metod) podatkovnega rudarjenja, s katerimi lahko opravljamo v tretjem poglavju opisanih šest nalog podatkovnega rudarjenja. Izbira tehnike je v praksi odvisna od poslovnega primera. V tem poglavju predstavljam pet najbolj znanih algoritmov podatkovnega rudarjenja: tehniko najbližjega sosedu, asociativna pravila, odločitvena drevesa, razvrščanje v skupine in nevronske mreže.

5.1 Najbližji sosed

Tehniko najbližjega sosedu (ang. *Nearest Neighbors*) v največji meri uporabljamo za klasificiranje, čeprav je uporabna tudi za ocenjevanje in napovedovanje. Pri tej tehniki dobimo klasifikacijo za neklasificirano opazovalno enoto tako, da jo primerjamo z najbolj podobnim zapisom v shranjenih zgodovinskih podatkih (Larose, 2005, str. 96). Tehnika najbližjega sosedu temelji torej na ideji podobnosti in se zgleduje po človeškemu načinu

razmišljanja na podlagi spomina (*Memory-Based Reasoning*, MBR). Ko ljudje razmišljamo, primerjamo spremenljivke neznane opazovane enote s spremenljivkami že znanih podobnih dogodkov in tako sklepamo vrednost neznanke (Berry & Linoff, 2004, str. 257). Na podoben način analitik napoveduje s tehniko najbližjega sosed »vrednost spremenljivke za neko (opazovano) enoto tako, da pogleda vrednost iste spremenljivke za tiste enote, ki so »najbližje« opazovani enoti« (Kovačič, Jaklič, Indihar Štemberger & Groznik, 2004, str. 250).

Pri uporabi tehnike najbližjega sosed potrebujemo serijo analognih zgodovinskih podatkov. Najlažja metoda iskanja najbližjega sosed je ta, da najprej poiščemo razdaljo med iskano vrednostjo in posameznimi zapisi v seriji podatkov. Razdaljo lahko izračunamo z izbrano funkcijo razdalje (ang. *Distance Function*). Zapis, do katerega ima obravnavana entiteta najmanjšo razdaljo, je najbližji sosed. Na podlagi njegovih podatkov nato dobimo rešitev problema. V slučaju, ko obstaja več kandidatov za najbližjega sosed (več zapisov se nahaja na podoben razdalji do obravnavane entitete) in nam ti dajejo različne rezultate, lahko kandidate podrobneje obravnavamo s tehtanim glasovanjem (ang. *Weighted Voting*) (Berry & Linoff, 2004, str. 257).

Tehniko najbližjega sosed lahko na primer uporabi zdravnik, ko nov bolnik potrebuje zdravilo. Zdravnik mora bolniku predpisati zdravilo, s katerim so ozdraveli bolniki s podobnim profilom. Primerja lahko starost, simptome, spol in druge spremenljivke in na podlagi tega dobi najbolj podobnega preteklega bolnika. Ta je v tem primeru »najbližji sosed«. Zdravilo, ki ga je zdravnik njemu predpisal, bo sedaj predpisal tudi novemu pacientu.

5.2 Asociativna pravila

S tehniko asociativnih pravil iščemo attribute ali značilnosti, ki »gredo skupaj«. To pomeni, da skušamo razkriti asociacije med temi atributi tako, da odkrijemo pravila, s katerimi lahko izmerimo zvezo med dvema ali več atributi. Ta tehnika je predvsem uporabna pri analizi nakupne košarice (Larose, 2005, str. 180).

Asociativna pravila imajo naslednjo obliko: če »predhodni dogodki«, potem »posledica«. Poleg tega merijo še podporo in zaupanje pravila (Larose, 2005, str. 180). **Podpora** (ang. *Support*) je verjetnost, da katera koli transakcija vsebuje predhodni dogodek in posledico. Torej predstavlja pogostost pojavljanja pravila. **Zaupanje** (ang. *Confidence*) je verjetnost, da tista transakcija, ki vsebuje predhodni dogodek, vsebuje tudi posledico (Kovačič et al., 2004, str. 250). Torej meri veljavnost pravila. Podpora je manjša ali kvečjemu enaka zaupanju.

Primer asociativnega pravila je lahko naslednji: »Če kupec kupi brezalkoholno pijačo, potem kupi tudi čipse.« Pri tem pravilu je podpora 5%, zaupanje pa 70%.

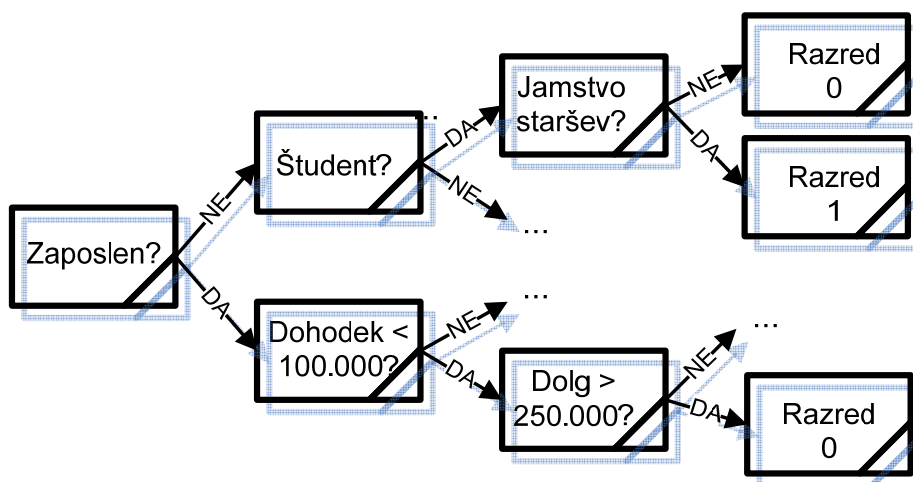
Ob ugotavljanju pravil analitik določi minimalno podporo in minimalno zaupanje pravila. Omejitve nastavi zato, ker bi bilo v nasprotnem slučaju visoko število pravil, ki le redkokdaj veljajo.

Problem analize nakupne košarice je ta, da pri tem pridemo večkrat do neuporabnih spoznanj. Primer je lahko pravilo: »če kupec kupi lonec, potem kupi tudi pokrovko«. Tako spoznanje je samo po sebi umevno in zato ni zanimivo. Analitik cilja vedno na prej nepoznana pravila, ki so skrita v podatkih in ki so lahko uporabna za nadaljnje odločanje in poslovanje, s katerimi lahko podjetje doseže konkurenčno prednost.

5.3 Odločitveno drevo

Odločitveno drevo ali drevo odločanja (ang. *Decision Tree*) je tehnika podatkovnega rudarjenja namenjena klasificiranju podatkov v ločene skupine ali veje. Odločitvena drevesa so v grafični obliki (v obliki drevesa) in prikazujejo neko odločitveno situacijo. Pri tem drevo začne s koreniko in se potem cepi. Vsako vozlišče predstavlja neko odločitveno vprašanje, kjer primerjamo vrednost spremenljivke z določeno konstanto. Na primer se lahko vprašamo, ali je vrednost spremenljivke večja ali manjša od konstante. Listi predstavljajo tisti delež populacije, ki je dosegla posamezen list. (Witten & Frank, 2005, str. 102). S pomočjo odločitvenih dreves skušamo torej deliti opazovane podatke v medsebojno izključujoče podskupine. Pri tem iščemo zakonitosti, ki jih lahko uporabimo v kasnejših odločitvenih situacijah. Kvaliteta drevesa je odvisna od izbire spremenljivk, ki najbolje delijo niz podatkov. Niz podatkov razcepimo zaporedoma v podnize, dokler naslednja delitev ne bi dala nobenih statistično pomembnih razlik ali dokler podskupine niso premajhne, da bi bila smiselna nadaljnja delitev. Drevo nato predstavlja model, s katerim lahko klasificiramo nove opazovane objekte in njihove spremenljivke. Prednost odločitvenih dreves je v tem, da nazorno prikazujejo pridobljeno znanje in da se z njimi na enostaven način izvleče pravila za razvrščanje v »če – potem« obliki (Kirkos, Spathis & Manolopoulos, 2004, str. 998-999).

Slika 1: Odločitveno drevo za ugotavljanje kreditne sposobnosti



Vir: H. Dürr, *Anwendungen des Data Mining in der Praxis*, 2005, str. 4.

Drevesa odločitve so posebno dobra pri ugotavljanju segmentov populacije, ki se obnašajo na tak način, kot mi želimo (Parr Rud, 2001, str. 19). Na primer lahko analiziramo značilnosti naših kupcev, ki so unovčili poslan kupon za določen izdelek.

Na Sliki 1 (na strani 11) je prikazan primer odločitvenega drevesa za ugotavljanje kreditne sposobnosti kupcev. Razred 0 predstavlja stranke, ki niso kreditno sposobne, medtem ko so kupci iz Razreda 1 kreditno sposobni. Kot lahko opazimo, zaposleni z dohodki pod 100.000 denarnih enot in z dolgom nad 250.000 denarnih enot niso vredni kredita. Obratno velja za študente, za katere jamčijo starši.

5.4 Razvrščanje v skupine

Z razvrščanjem v skupine (ang. *Clustering*) dobimo skupine predmetov (ali ljudi), ki imajo podobne značilnosti. Naš cilj je, da se skupine med sabo čim bolj razlikujejo, znotraj skupine pa želimo doseči čim višjo homogenost.

Pri tehniki razvrščanja v skupine najprej dobimo diagram, ki kaže kako enote padejo v skupine. V najbolj preprostem slučaju lahko to preprosto naredimo tako, da upodobimo enote na podlagi dveh dimenzij in logično razdelimo sliko ter tako dobimo skupine (Witten & Frank, 2005, str. 81).

Tehnika razvrščanja v skupine je zelo uporabna pri segmentiranju trga. Pri tem, glede na določene dimenzije (kot so lahko spol, starost, dohodek itd.) razdelimo naše kupce v skupine in oblikujemo tipični profil vsake skupine. To nam lahko daje mnoge koristi. Prvo kot prvo lahko odločevalec uporabi poseben trženjski splet, da z večjo natančnostjo cilja na manjšo skupino trga. Na tak način lahko razporedi vire na bolj uspešen in učinkovit način. Poleg tega sili segmentacija trga v bližji odnos med kupci in podjetjem. Posebna korist segmentacije je še ta, da odločevalec lahko lažje določi konkurenčne strategije (na primer strategijo diferenciacije, strategijo nizkih stroškov ali strategijo fokusa) (Huang et al., 2007, str. 313).

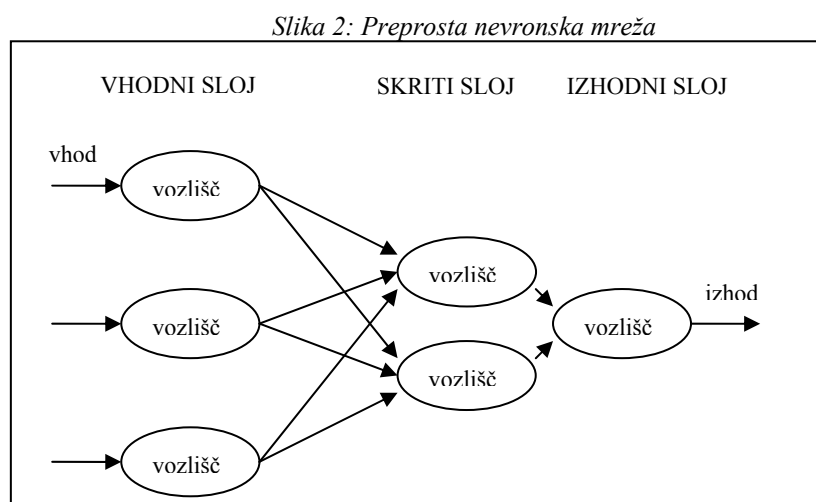
5.5 Umetne nevronske mreže

Inspiracija za umetne nevronske mreže (ang. *Artificial Neural Networks*) je bilo spoznanje kompleksnega delovanja sistema učenja živalskih možganov, ki jih sestavljajo povezani nizi nevronov. Čeprav ima nevron relativno preprosto strukturo, lahko goste mreže povezanih nevronov opravljajo zahtevne naloge učenja, kot so na primer klasifikacije in prepoznavanja vzorcev. Umetne nevronske mreže predstavljajo poskus na osnovni ravni posnemanja vrste nelinearnega učenja, ki je značilna za nevronske mreže, ki jih ustvarja narava (Larose, 2005, str. 128-129).

Na Sliki 2 (na strani 13) je prikazan enostaven primer nevronske mreže. Nevronska mreža lahko vodi le v eno smer in ne omogoča vrtenja podatkov. Sestavljena je iz dveh ali več

slojev, čeprav ima večina tri sloje: vhodni sloj, skriti sloj in izhodni sloj. Nekatere nevronske mreže imajo več skritih slojev, čeprav je ponavadi dovolj le eden. Nevronska mreža je popolno povezana, kar pomeni, da je vsako vozlišče sloja povezano z vsemi vozlišči naslednjega sloja. Vsaka povezava med vozlišči ima svojo utež v vrednostih med nič in ena (Larose, 2005, str. 131-132). Uteži predstavljajo relativno moč posameznih vrednosti v vozliščih in so rezultat učenja nevronske mreže.

Da se lahko nevronska mreža uči, moramo preskrbeti skupek zgodovinskih vhodnih podatkov in ustreznih poznanih vrednosti izhodov. Nevronska mreža se uči tako, da skuša postopno razumeti povezavo med vhodnimi in izhodnimi podatki. Iterativno uravnava uteži za vsak testni primer, dokler povprečna napaka na kvadrat med ocenjeno in dejansko vrednostjo se ne nahaja na predhodno določenem zadovoljivem nivoju (De Barcelos Tronto, Simoes da Silva & Sant'Anna, 2008, str. 359).



Vir: D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, 2005, str. 132, slika 7.2.

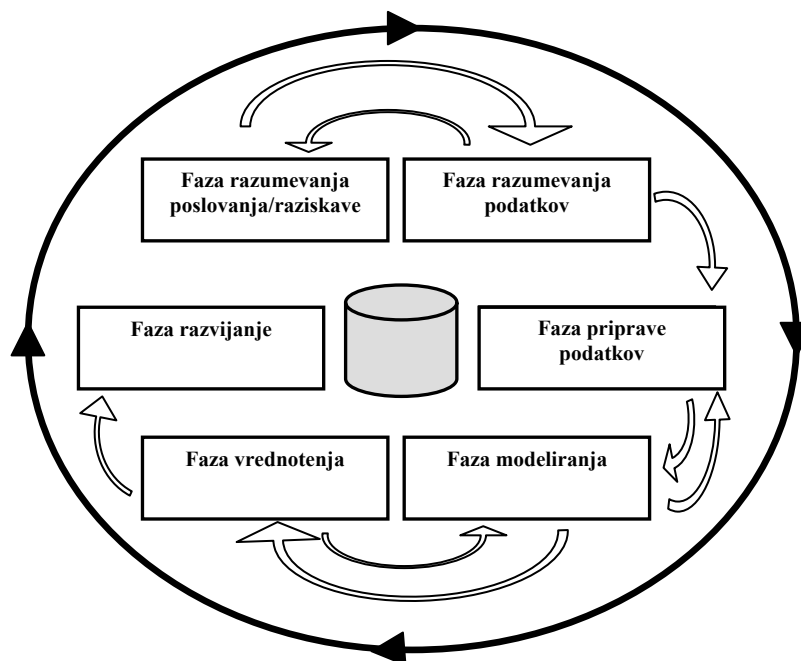
Vhodni sloj sprejema vhode iz niza podatkov (na primer vrednosti atributov) in jih preprosto poda naprej. V naslednji fazi se vhodni podatki pretvorijo s pomočjo transformacijske funkcije (ki je običajno seštevek zmnožkov uteži in vrednosti spremenljivk) in gredo v naslednji sloj. Postopek se ponavlja, dokler podatki ne pridejo iz izhodnega sloja. Izhod iz izhodnega sloja predstavlja rešitev problema.

6 PROCES PODATKOVNEGA RUDARJENJA

V literaturi lahko dobimo več različnih modelov procesa podatkovnega rudarjenja. V nadaljevanju predstavljam CRISP-DM model (*Cross-Industry Standard Process for Data Mining*), ki so ga leta 1996 razvili analitiki iz DaimlerChrysler-ja, SPSS-a in NCR-ja. Njihov cilj je bil sestaviti standardni in objektivni model. Uporaben je na različnih industrijskih področjih in ohranja nevtralen pogled do orodij za podatkovno rudarjenje.

Iz Slike 3 je razvidno, da proces podatkovnega rudarjenja sestoji iz šestih faz. Vrstni red faz ni strogo določen, a se je vedno potrebno premikati ali naprej ali nazaj med različnimi fazami, kar je odvisno od izhoda iz predhodne faze. Večkrat se lahko zgodi, da se je za doseganje izboljšav potrebno vračati na predhodne faze. Puščice na sliki predstavljajo najbolj pomembne povezave med fazami. Model je iterativen, saj se večkrat zgodi, da rešitev raziskovalnega problema vodi v zanimanje za nadaljnje raziskave. Poleg tega je v novih projektih priporočljivo kot vhod uporabiti spoznanja in izkušnje iz prejšnjih raziskav.

Slika 3: Proces podatkovnega rudarjenja (CRISP-DM model)



Vir: D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, 2005, str. 6, slika 1.1.

Na spletni strani CRISP-DM modela (<http://www.crisp-dm.org/Process/index.htm>) lahko najdemo opise posameznih faz:

1. faza: razumevanje poslovanja ali raziskave

V prvi fazi smo usmerjeni v razumevanje ciljev projekta in zahtev iz poslovnega zornega kota. To potem pretvorimo v definicijo problema podatkovnega rudarjenja. Poleg tega moramo pripraviti preliminarno strategijo za doseganje zadanih ciljev.

2. faza: razumevanje podatkov

Zbrati moramo podatke in jih raziskovati, da pridemo v stik z njimi in da jih lahko bolje razumemo. Poleg tega je potrebno oceniti kakovost podatkov.

3. faza: priprava podatkov

Faza priprave podatkov je precej mučna. Podatki prihajajo iz različni virov podatkov in posledično obstajajo razlike v časovnih dobah, glavnih ključih, napakah, načinih shranjevanja, agregaciji podatkov itd. Zato je potrebno podatke zbrati, jih integrirati in prečistiti. Iz začetnih surovih podatkov moramo pripraviti končni skupek podatkov, ki ga

bomo uporabljali v vseh naslednjih fazah. Izbrati moramo vse tabele, enote in spremenljivke, ki jih želimo analizirati in ki so primerni za našo analizo. Če je potrebno, moramo pretvoriti nekatere spremenljivke v drugo obliko. Prečistiti moramo surove podatke, da so pripravljeni za orodja modeliranja.

4. faza: modeliranje

V četrti fazi moramo izbrati in aplicirati primerno tehniko modeliranja. Večkrat se zgodi, da je potrebno uporabiti več tehnik modeliranja za reševanje istega problema. Z namenom optimiziranja rezultatov moramo nato preveriti nastavitve parametrov. Če določena tehnika zahteva spremembe v podatkih, se lahko vrnemo v fazo priprave podatkov.

5. faza: vrednotenje

Na začetku te faze imamo že izgrajen model (ali več modelov), ki z vidika analize podatkov zgleda visoko kvaliteten. Preden uporabimo model, moramo bolj podrobno ovrednotiti njegovo kvaliteto in učinkovitost. Zagotoviti moramo, da lahko z modelom dejansko dosežemo zadane poslovne ali raziskovalne cilje in da nismo pozabili česa pomembnega. Na koncu te faze se moramo odločiti o uporabi rezultatov podatkovnega rudarjenja.

6. faza: razvijanje

Projekt se ponavadi ne konča takrat, ko je model ustvarjen. Tudi ko je cilj projekta le povečanje znanja, je potrebno to znanje primerno organizirati in predstaviti, da se ga lahko potem uporabi. Zadnja faza projekta je lahko preprosta (ko na primer naredimo le poročilo) ali kompleksna (ko na primer implementiramo ponavljajoč proces podatkovnega rudarjenja). Kakor koli že, moramo na koncu razumeti, katera dejanja bomo morali izvesti, da bomo kreiran model tudi dejansko uporabili.

7 ANALIZA PODATKOV IZ ANKETE PODJETJA PETROL d.d.

Podjetje Petrol spremlja od leta 1999 zadovoljstvo svojih kupcev (Blažič, 2008, str.13). Za namene analize navad kupcev podjetja Petrol, sem prejela bazo podatkov z rezultati ankete iz leta 2001. V spodnjih podpoglavjih opisujem problematiko, uporabljeno orodje podatkovnega rudarjenja in cilje ter rezultate analize.

7.1 Opis podjetja Petrol d.d.

Podjetje Petrol spada med največja slovenska podjetja: je največje slovensko energetska podjetje, največji slovenski uvoznik in podjetje z najvišjimi prihodki v Sloveniji. Poleg tega spada med največje slovenske trgovske družbe in ponuja delo več kot 500 zaposlenim (Spletna stran podjetja Petrol, 2008).

Osrednja dejavnost podjetja je trgovanje z naftnimi derivati, zemeljskim in utekočinjenim plinom ter električno energijo. Pri tem podjetje ustvarja nekaj več kot 80% vseh svojih prihodkov. Preostalih nekaj manj kot 20% prihodkov pridobi s trgovanjem z blagom za

široko porabo in s storitvami. Na tem perspektivnem področju ima Petrol 10% realno letno stopnjo rasti (Spletna stran podjetja Petrol, 2008).

Podjetje Petrol stremi k poslovni odličnosti in spremlja ter spoštuje tržne zahteve s področja storitev, informacijske tehnologije in varovanja okolja (Spletna stran podjetja Petrol, 2008). V zadnjem desetletju skuša s pomočjo raziskav zagotoviti celovite informacije za vse faze nakupne aktivnosti potrošnikov. Raziskave dajejo informacije na primer o zadovoljstvu potrošnikov, o njihovem odzivu na oglaševanja in pospeševanja prodaje, o njihovem vedenju na področju oskrbe z energenti itd. Poleg tega v podjetju raziskujejo tudi potrebe ter zadovoljstvo drugih deležnikov podjetja (zaposlenih, delničarjev, medijev in javnosti) in značilnosti trgov. Vse te informacije so namenjene sprejemanju poslovnih odločitev (predvsem na področju trženja) in omogočajo vpeljavo strategije upravljanja odnosov s strankami (ang.: *Customer Relationship Management, CRM*) (Blažič, 2008, str. 5-19).

Petrol ima po Sloveniji 307 črpalk in ima po številu črpalk 65% tržni delež (Poročila podjetja Petrol, 2007). Glavni konkurent podjetja je avstrijski OMV, ki ponuja tudi prestižno gorivo OMV Super 100. Poleg tega so na trgu prisotna še podjetja MOL, Agip in Shell. V letu 2001 (podatke, ki jih uporabljam v diplomski nalogi segajo v leto 2001) je imel Petrol 280 črpalk po Sloveniji in s tem kar 72% tržni delež glede na število črpalk (Poročila podjetja Petrol, 2001). Kljub skoraj monopolnem položaju se je Petrol zavedal nevarnosti konkurence ter pomembnosti zadovoljstva kupcev in je sproti spremljal dogajanje na trgu.

Podjetje Petrol nudi svojim kupcem kartico zvestobe. "Kartica Magna je namenjena brezgotovinskemu plačevanju blaga in storitev tako na Petrolovih bencinskih servisih v Sloveniji, na Hrvaškem in Bosni in Hercegovini ter v skladiščih, kot tudi na več kot tisoč drugih prodajnih mestih po vsej Sloveniji" (Spletna stran podjetja Petrol, 2008). Imetnikom nudi številne ugodnosti in druge dodatne storitve, ki jih omogoča Petrol skupaj s partnerji.

7.2 Opis ankete

V moji diplomski nalogi sem za analizo dobila podatke o Petrolu iz ankete, ki se je odvijala telefonsko od 26. do 29. junija 2001. Novejših podatkov nisem prejela, saj so podatki iz raziskav za podjetja skrajno zaupni in poleg tega raziskave niso pogoste. Anketa, ki jo uporabljam pri svoji analizi, je bila namenjena merjenju zaznane kakovosti za podjetje Petrol in ocenjevanju poslovanja z Magna kartico. Za ocenjevanje so uporabili Likartovo petstopenjsko lestvico. Pri vprašanjih glede kakovosti Petrola so anketiranci odgovarjali z ocenami od 1 (sploh se ne strinjam) do 5 (popolnoma se strinjam). Pri vprašanjih v zvezi z Magna kartico, so vprašani lahko dajali ocene od 1 (popolnoma nezadovoljen) do 5 (popolnoma zadovoljen). Možno je bilo torej tudi delno strinjanje in delno zadovoljstvo (ocena 3). Poleg tega je vprašalnik dopuščal izbiro »brez odgovora« ali »ne morem odgovoriti«.

Vprašanja iz ankete so bila razdeljena v štiri sklope: v sklop merjenja kakovosti za Petrol, v sklop demografskih in socioekonomskih vprašanj, v sklop merjenja pričakovanj in v sklop ocene poslovanja z Magno. Poleg tega so v bazi podatkov shranjeni tudi razni drugi kvantitativni podatki, kot so potrošnja z Magno in število obiskanih Petrolovih prodajnih mest. Vprašalnik ter natančnejši seznam uporabljenih podatkov se nahajata v Prilogi 1 in v Prilogi 2.

V anketi je sodelovalo 201 imetnikov Magna kartice (63% od teh je bilo moških in 37% je bilo žensk). Povprečna starost je bila 43 let. V Prilogi 3 so nazorno prikazani demografski statistični podatki anketirancev. Iz grafov je razvidno, da je skupina anketirancev iz raznih zornih kotov precej homogena. Na primer je večina vprašanih odgovorila, da so samski (81%) in da so (so)lastniki stanovanja, v katerem živijo (80%).

Da bi se pred podatkovnim rudarjenjem bolje seznanila s podatki, sem najprej naredila temeljne statistične analize s pomočjo orodja Excel. Opazila sem, da je zadovoljstvo strank glede kakovosti Petrola na povprečno zelo visoki ravni. Pri skoraj vseh vprašanjih je povprečna raven zadovoljstva nad 4. Torej lahko trdim, da so imetniki Magna kartice v splošnem zadovoljni s Petrolovimi storitvami. Edino pri vprašanjih o tem, ali je razumljivo, da zaradi velike količine dela Petrolovi prodajalci ne morejo takoj pomagati ali ne morejo posvetiti veliko časa, je povprečna ocena malenkost pod 4. Odgovori na vprašanja so zelo homogeni, kar je v bistvu ovira za ugotavljanje asociativnih pravil. Če bi bile vrednosti spremenljivk bolj razpršene, bi bilo pridobivanje uporabnih povezav v podatkih dosti lažje.

7.3 Orodje Visual Studio 2005

Pri analizi podatkov sem uporabljala orodje za Microsoft Visual Studio 2005, ki ima vključeno komponento za podatkovno rudarjenje *Analysis Services*. Orodje Visual Studio 2005 sem izbrala zato, ker omogoča uporabo mnogo tehnik podatkovnega rudarjenja (asociativna pravila, regresijsko analizo, drevesa odločanja, razvrščanje v skupine, umetne nevronske mreže in analizo časovnih zaporedij) in projektiranje ter izdelovanje modelov za podatkovno rudarjenje. Na trgu je dosti orodij, ki omogočajo le del teh tehnik, Visual Studio 2005 pa se mi zdi dovolj popoln za mojo analizo. Orodje podpira tudi večdimenzionalne baze podatkov (OLAP kocke) (Spletna stran mreže razvijalcev Microsoft, 2008).

Bazo podatkov sem dobila v orodju SPSS, ki je združljivo z orodjem za podatkovno rudarjenje Clementine. Povezava med Visual Studiom 2005 in bazo podatkov v SPSS-ju ni bila uspešna, zato sem bazo podatkov izvozila v MS Access 2003. Na žalost je baza podatkov le dvodimenzionalna, kar onemogoča uporabo vseh zmogljivosti Visual Studia 2005 (večdimenzionalne baze podatkov imajo več hierarhij in lahko z zvijanjem ter vrtanjem v globino pridobimo dosti uporabnih informacij).

7.4 Analiza podatkov

Pred analizo podatkov sem si postavila nekaj vprašanj in sovpadajočih hipotez. S pomočjo podatkovnega rudarjenja sem želela odkriti prave odgovore na zastavljena vprašanja.

Vprašanje 1: Ali demografski in socioekonomski podatki vplivajo na potrošnjo na prodajnih mestih Petrola? Če da, kateri?

- **Podvprašanje 1.1: Kateri demografski in socioekonomski podatki vplivajo na potrošnjo goriva?**
- **Podvprašanje 1.2: Kateri demografski in socioekonomski podatki vplivajo na potrošnjo dodatne ponudbe Petrola?**

Vprašanje je po mojem mnenju pomembno za boljše razumevanje kupcev in za nadaljnje trženjske akcije. Podjetje se s takimi informacijami lahko usmerja v bolj donosne kupce in izpušča manj donosne.

Prva hipoteza je v tem primeru ta, da demografski in socioekonomski podatki vplivajo na potrošnjo na prodajnih mestih Petrola. Druga hipoteza pa je, da v največji meri vpliva na potrošnjo dohodek gospodinjstva. Poleg potrditve hipoteze sem želela z analizo dobiti še druge značilnosti, ki vplivajo na celotno potrošnjo, na potrošnjo le goriva in na potrošnjo dodatne ponudbe. Te je le s pomočjo logičnega razmišljanja težko oceniti. Analizo sem izvedla s pomočjo tehnike odločitvenih dreves.

Pri izdelavi modela sem kot napovedne spremenljivke (*Predict Only*) določila potrošnjo v zadnjih šestih mesecih z Magna kartico na Petrolovi prodajnih mestih, potrošnjo goriva v zadnjih šestih mesecih z Magna kartico na Petrolovi prodajnih mestih in potrošnjo dodatne ponudbe v zadnjih šestih mesecih z Magna kartico na Petrolovi prodajnih mestih. Da bi izvedela, katere demografske in socioekonomske spremenljivke vplivajo na potrošnjo, sem izbrala kot vhode (*Input*) vse možne vplivne spremenljivke: mesečni dohodek gospodinjstva, spol, starost, izobrazbo, zakonski stan, število vzdrževanih članov in stanovanjske razmere.

V Prilogi 5 je na Sliki 10 odločitveno drevo, ki prikazuje na levi strani celotno populacijo in se nato proti desni deli glede na najbolj pomembne spremenljivke. Svetloba lista kaže na gostoto populacije v listu: najbolj temni listi imajo največji del populacije. Visual Studio 2005 prikaže le spremenljivke, ki v večji meri vplivajo na napovedovalno spremenljivko (zato v prikazu ni vseh spremenljivk). Najbolj pomembni vpliv na napovedovalno spremenljivko imajo vhodi, ki so bližji listu s celotno populacijo.

Iz rezultata analize lahko opazimo, da je prva hipoteza potrjena, druga pa ne. Demografski podatki res vplivajo na višino potrošnje, saj je orodje Visual Studio 2005 delilo populacijo in ustvarilo odločitveno drevo. Ko razčlenitev glede na vplive vhodov ni smiselna, celotna populacija ostane v prvem listu. Druga hipoteza je ovržena, saj orodje za podatkovno

rudarjenje ni delilo populacije glede na dohodek. To pomeni, da ta vhodna spremenljivka nima pomembnega vpliva na potrošnjo.

Kateri je lahko razlog, da razlike v dohodkih ne vplivajo na potrošnjo na Petrolovih prodajnih mestih? Glavni razlog sta po mojem mnenju narava ponujenih izdelkov in skoraj monopolni položaj Petrola v letu 2001. Znana ekonomista Samuelson in Nordhaus (2002, str. 64) trdita, da je »za življenjske potrebščine, npr. hrano, gorivo [...], povpraševanje togo (neelastično)«. Take dobrine ljudje kupijo tudi, ko se cena poviša ali je dohodek nižji.

Na celotno potrošnjo v največji meri vpliva zakonski stan, nato pa še število vzdrževanih članov in stanovanjske razmere (glej Sliko 10 v Prilogi 5). Kupci, ki največ potrošijo, so samski in nimajo vzdrževanih članov.

Na potrošnjo za nakup goriva na Petrolovih prodajnih mestih vpliva v največji meri število vzdrževanih članov in nato še spol (glej Sliko 11 v Prilogi 5). Za gorivo največ zapravijo moški, ki nimajo vzdrževanih članov.

Pri potrošnji za nakup dodatne Petrolove ponudbe spol ni pomembna spremenljivka (glej Sliko 12 v Prilogi 5). Največji je vpliv zakonskega stanu, pri čemer ločene ali ovdovele osebe porabijo manj denarja v te namene.

2. Vprašanje: Katere so demografske značilnosti najboljših Petrolovih strank?

Najboljši kupci so tisti, ki potrošijo največ denarja. Z nedonosnimi kupci lahko zgubljam veliko nepotrebnih energij in sredstev. Pomembno je, da poznamo ciljne kupce, zato sem želela izvedeti, kateri je profil tipičnega dobrega kupca. Hipoteza je bila pri tem vprašanju rezultat prejšnjega vprašanja: tipični najboljši kupec je samski, nima vzdrževanih članov in je moškega spola. Poleg teh kvalitativnih podatkov sem želela izvedeti, v kolikšni meri to drži in če obstajajo še kake druge značilnosti najboljših kupcev Petrola.

Da bi dobila najboljše kupce, sem se zanašala na Paretovo 20/80 pravilo, ki pravi, da 20% najboljših kupcev daje 80% prihodkov. Za to sem najprej izvozila podatke v Excel in razporedila kupce od tistega, ki na prodajnih mestih Petrola je v zadnjih šestih mescih pred anketo največ potrošil vse do najslabšega kupca. Nato sem preverila Paretovo 20/80 pravilo in prišla do zaključka, da najboljših 20% kupcev prinašajo Petrolu 50% prihodkov. Čeprav ne dosegajo Paretovega pravila, so ti kupci najbolj pomembni. Podatke o 20% kupcev sem izvozila v Accessovo bazo podatkov. Nato sem kreirala povezavo na poizvedbo o najboljših kupcih (ki v tem primeru predstavljajo celotno populacijo). Za tehniko podatkovnega rudarjenja sem določila tehniko razvrščanja strank, saj ta na zelo nazoren način pokaže značilnosti populacije.

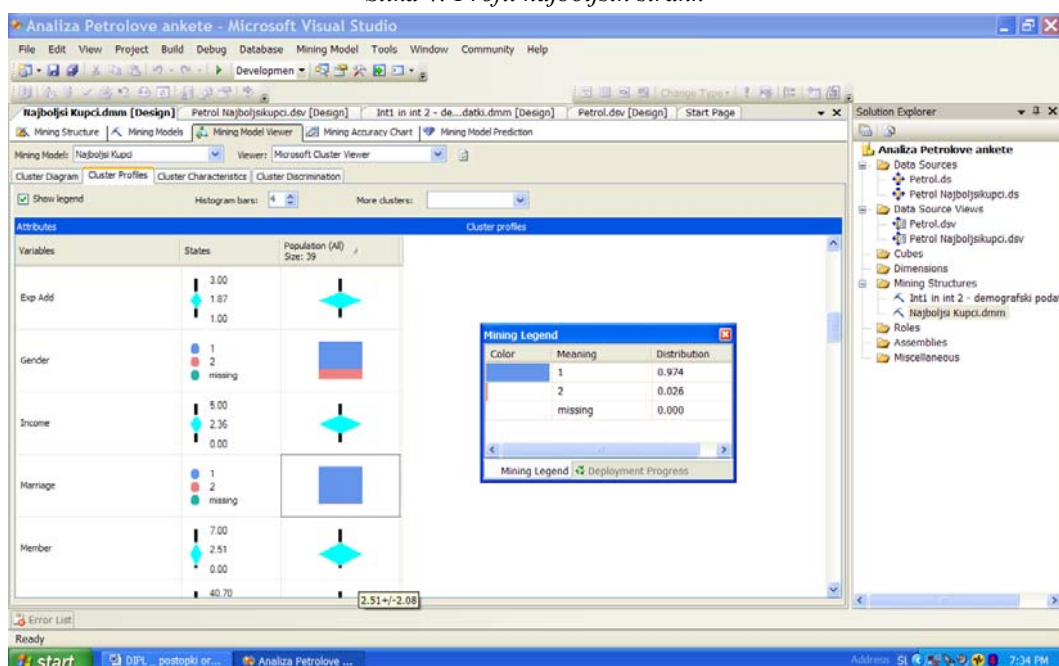
Da bi lahko videla vse značilnosti populacije, sem kot vhodne spremenljivke vključila vse spremenljivke, razen identifikacijske številke, saj je to glavni ključ. Določila sem še, katerega tipa je posamezna spremenljivka, in zagnala sem proces.

Kot lahko vidimo na Sliki 4, je hipoteza skoraj v celoti potrjena. Kar 97,4% najboljših kupcev Petrola je samskih in 74,4% jih je moškega spola. Število vzdrževanih članov pa ni značilna lastnost najboljših kupcev (variira med 0 in 3 in povprečje je 1,4). 89,7% najboljših kupcev je (so)lastnikov stanovanja, v katerem živijo. Večina (64,1%) jih je končala srednjo šolo, 30,8% pa višjo ali visoko šolo.

Med demografske in socioekonomske podatke, pri katerih je vrednost spremenljivke razpršena, spadata še starost in dohodek gospodinjstva. Starost se giblje med 27 in 60 let, povprečna starost pa je 44 let.

Mnenja najboljših kupcev o Petrolovi prodajni mestih so v splošnem zelo dobra (ta podatek velja tudi za druge kupce). Prav tako je večina zadovoljna z načinom obračuna transakcij z Magna kartico.

Slika 4: Profil najboljših strank



Vir: Lastna izdelava

Na podlagi analize odgovorov strank, sem zasledila nekatere zanimive vzorce obnašanja strank, ki so po mojem mnenju uporabni za nadaljnje poslovanje in odločanje v podjetju Petrol. Ker je pri podatkovnemu rudarjenju bistvo prav odkrivanje prej neznanih, a uporabnih informacij, bom povzela svoje ugotovitve. Prva ugotovitev je naslednja: 76% najboljših kupcev (sploh) se ne strinja, da na Petrolovi prodajni mestih kupuje le gorivo, in se (popolnoma) strinja, da ob nakupu goriva vedno kupi še kako malenkost. Istočasno se večina (sploh) ne strinja, da se pogosto na Petrolovi prodajni mestih ustavi le za nakup izdelkov dodatne ponudbe. To lahko pomeni, da je pri nakupovanju na Petrolovi prodajni

mestih za najboljše kupce Petrola glavni cilj nakup goriva in je nakup dodatna ponudbe impulzivni. Lahko pa tudi pomeni, da kupec želi izdelek iz dodatne ponudbe, in kupi pri tem tudi gorivo. Zaradi tega je dodatna ponudba bistvena za povečanje prihodkov podjetja Petrol.

Drugi pomembni vzorec obnašanja je po mojem mnenju ta, da večina populacije (89,7%) dobiva obvestila o ugodnostih, namenjenih imetnikom Magna kartice, a da le 17,9% najboljših kupcev izkorišča ugodnosti. Pri tem bi svetovala podjetju Petrol, naj raziskuje razloge za tako nizek odstotek, naj analizira želje in potrebe kupcev ter naj na podlagi tega ukrepa. Mogoče dane ugodnosti nimajo večje vrednosti v očeh kupcev in jih je zato potrebno prilagoditi.

3. vprašanje: Ali stranke, ki na Petrolovi prodajni mestih kupujejo le gorivo, imajo podpovprečno mnenje o kakovosti za Petrol? In tiste stranke, ki kupujejo tudi druge izdelke ali hodijo večkrat le po izdelke dodatne ponudbe na Petrolova prodajna mesta, imajo nadpovprečno mnenje o kakovosti za Petrol?

Nekatere stranke se strinjajo ali popolnoma strinjajo, da na Petrolovi prodajnem mestu kupujejo samo gorivo (glej Prilogo 4). Pri tem me zanima, katera so mnenja teh strank o kakovosti Petrolovi prodajni mestih. Rada bi izvedela, ali mnenje o kakovosti za Petrol vpliva na dejstvo, da kupci kupujejo le gorivo, kupujejo tudi malenkosti iz dodatne ponudbe ali gredo do Petrola tudi za dodatne nakupe. Odgovor na to vprašanje lahko pomaga Petrolu pri večanju prihodkov s povišanjem zadovoljstva strank na določenih ključnih področjih. Moja hipoteza je pri tem ta, da boljše mnenje o kakovosti vpliva na nakup tudi drugih izdelkov.

Da bi pregledala, katera mnenja v največji meri vplivajo na zgoraj omenjene odločitve, sem izbrala tehniko odločitvenega drevesa. Kot vir podatkov sem uporabila celotno Petrolovo bazo podatkov. Med napovedne spremenljivke sem dala mnenja o naslednjih vprašanjih: ali kupci na Petrolovem prodajnem mestu kupujejo samo gorivo, ali ob nakupu goriva vedno kupijo še kako malenkost in ali se pogosto ustavljajo na Petrolovi prodajni mestih le za nakup izdelkov iz dodatne ponudbe. Med vhodne spremenljivke sem vključila vse odgovore na vprašanja o kakovosti za Petrol (glej vprašalnik v Prilogi 1).

Rezultati analize so pokazali, da na odločitev o nakupu izdelka iz dodatne ponudbe (z gorivom ali brez) ne vplivajo mnenja o kakovosti za Petrol.

Iz Slike 15 v Prilogi 7 lahko opazimo, da je mnenje o zaupanju prodajnim osebam Petrola (SQ_12) edina spremenljivka, ki ima določen vpliv na odločitev o nakupu goriva brez dodatnih izdelkov.

Da bi bolje raziskovala korelacijo med tema dvema spremenljivkama, sem jo preizkusila tudi s tehniko asociativnih pravil (glej Prilogo 8). Pri tem sem prišla do naslednjih pravil:

- če se kupec popolnoma strinja, da lahko zaupa prodajalcem, potem se sploh ne strinja, da na Petrolovi prodajnih mestih kupuje le gorivo (z zaupanjem 50,1% in podporo 16,3%);
- če se kupec (popolnoma) ne strinja, da lahko zaupa prodajnim osebam, potem se delno strinja ali strinja, da na Petrolovi prodajnih mestih kupuje samo gorivo (z zaupanjem 50,0% in podporo 37,9%).

SKLEP

Podatkovno rudarjenje je ob naraščanju količine shranjenih podatkov v veliko pomoč pri pridobivanju uporabnih informacij za poslovno odločanje. V teoretičnem delu sem razložila definicijo in pomen te vede ter opisala razne možnosti uporabe v praksi. Poleg tega sem predstavila šest nalog, glavnih pet tehnik in proces podatkovnega rudarjenja.

Znanje o zgoraj omenjenih konceptih sem uporabila za dejansko analizo nakupnih navad kupcev na Petrolovi bencinskih servisih. Z orodjem za podatkovno rudarjenje Visual Studio 2005 sem analizirala podatke iz ankete o zadovoljstvu strank, ki jo je podjetje Petrol izvedlo leta 2001.

Najprej sem analizirala vplive demografskih in socioekonomskih spremenljivk na višino potrošnje na Petrolovi prodajnih mestih. Na podlagi tega lahko v Petrolu bolje poznajo svoje kupce in bolj učinkovito sprejemajo odločitve iz področja trženja in prodaje. Prišla sem do zaključka, da dohodek ne vpliva na potrošnjo. Na celotno potrošnjo v največji meri vpliva zakonski stan, nato pa še število vzdrževanih članov in stanovanjske razmere. Največ potrošijo samski kupci, ki nimajo vzdrževanih članov. Na potrošnjo za nakup goriva na Petrolovi prodajnih mestih vpliva v največji meri število vzdrževanih članov in nato še spol, pri čemer za gorivo največ zapravijo moški, ki nimajo vzdrževanih članov. Pri potrošnji za nakup dodatne Petrolove ponudbe je vpliv zakonskega stanu, pri čemer ločene ali ovdovele osebe porabijo manj denarja v te namene.

V diplomskem delu sem izdelala profil 20% najboljših strank. Ti prinašajo podjetju 50% prihodkov in so zato ciljni kupci. Tipični ciljni kupec podjetja Petrol je samski moški s srednjo ali visoko izobrazbo, ki je (so)lastnik stanovanja, v katerem živi. V take kupce mora podjetje vlagati največ energij in sredstev ter prilagoditi trženjski splet njihovim potrebam in željam. Ker velika večina ciljnih kupcev ne izkorišča ugodnosti Magna kartice, podjetju priporočam, naj analizira želje ciljnih strank in temu skuša prilagoditi ugodnosti.

V zadnjem delu analize sem želela izvedeti, če obstaja vpliv mnenj o kakovosti za Petrol na to, ali kupci kupujejo na Petrolovi prodajnih mestih le gorivo, tudi izdelke iz dodatne ponudbe in na to, ali gredo včasih na Petrol le za nakup izdelkov iz dodatne ponudbe. Taka informacija je pomembna, saj lahko pomaga Petrolu pri večanju prihodkov s povišanjem zadovoljstva strank na določenih ključnih področjih. Rezultati analize so pokazali, da na odločitev o nakupu izdelka iz dodatne ponudbe (z gorivom ali brez) ne vplivajo mnenja o

kakovosti za Petrol. Edina spremenljivka, ki ima določeno stopnjo korelacije z odločitvijo o nakupu goriva brez dodatnih izdelkov, je mnenje o zaupanju prodajnim osebam Petrola. Pri tem se kupec, ki se popolnoma strinja, da lahko zaupa prodajalcem, sploh ne strinja, da na Petrolovi prodajnih mestih kupuje le gorivo. Če pa se kupec (popolnoma) ne strinja, da lahko zaupa prodajnim osebam, potem se delno strinja ali strinja, da na Petrolovi prodajnih mestih kupuje samo gorivo.

V diplomskem delu sem dosegla zastavljene cilje. Na podlagi teoretičnega znanja sem raziskala bazo podatkov Petrola. Analizirala sem ciljne kupce in vplive na večjo potrošnjo. Pridobila sem koristne informacije, ki bi jih lahko podjetje Petrol uporabljalo za nadaljnje odločanje in na podlagi katerih bi lahko sestavilo trženjski splet ter se usmerilo v bolj donosne kupce (ob predpostavki, da bi bili podatki ažurni).

LITERATURA IN VIRI

1. *Attracting Profitable Customers through Intelligent Dialogs: Banco Popular Case Study 2004 [Million Handshakes]*. Najdeno 25. julija 2008 na spletnem naslovu <http://www.millionhandshakes.com/MHPPProdWeb/Home.htm>.
2. Berry, M. J. A. & Linoff, G. S. (2004). *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management*. Indianapolis: Wiley Publishing.
3. Blažič, T. (2008). *Vzpostavitev sistema tržnih raziskav za podporo poslovnim odločitvam – Primer Petrol*. Ljubljana: Podjetje Petrol.
4. Cabena, P., Hadjinian, P., Stadler, R., Verthes, J. & Zanassi, A. (1998). *Discovering Data Mining: From Concept to Implementation*. New Jersey: Prentice Hall.
5. Brause, R., Langsdorf, T. & Hepp, M. (1999). Neural Data Mining for Credit Card Fraud Detection. *IEEE Computer Society*, 11 (1), 103-106.
6. De Barcelos Tronto, I. F., Simoes da Silva, J. D. & Sant'Anna, N. (2008). An investigation of artificial neural networks based prediction systems in software project management. *The Journal of Systems and Software*, 81 (3), 356–367.
7. Dürr, H. (2003). *Anwendungen des Data Mining in der Praxis*. Ulm: Universität Ulm.
8. Ferrari, A & Ziesserman, A. (2007). *Learning Visual Attributes*. Najdeno 22. julija 2008 na spletnem naslovu <http://www.robots.ox.ac.uk/>.
9. Han, J. & Kamber, M. (2000). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
10. Hand, D., Mannila, H. & Smyth, P. (2001). *Principles of Data Mining*. London: The MIT Press.
11. Hisnchun, C., Wingyan, C., Xu, J. J., Yi Qin, G. W. & Chau, M. (2004). Crime Data Mining: A General Framework and Some Examples. *Computer*, 37 (4), 50-56.
12. Huang, J. J., Tzeng, G. H. & Ong, C. S. (2007). Marketing segmentation using support vector clustering. *Expert Systems with Applications*, 32 (2), 313–317.
13. Kirkos, E., Spathis, C., Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32 (4), 995–1003.
14. Kotler, P. (2004). *Management trženja*. (slovenski prevod 11. izdaje) Zagreb: GV Založba.
15. Kovačič, A., Jaklič, J., Indihar Štemberger, M. & Groznik, A. (2004). *Prenova in informatizacija poslovanja*. Ljubljana: Ekonomska fakulteta.
16. Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons.
17. Laxman, S. & Sastry, P. S. (2006). A Survey of Temporal Data Mining. *Sadhana*, 31 (2), 173-198.
18. *Letno poročilo skupine Petrol in družbe Petrol d.d., Ljubljana za leto 2007*. Najdeno 25. julija 2008 na spletnem naslovu <http://www.petrol.si/media/lp.petrol.07-slo.pdf>.

19. *Letno poročilo skupine Petrol in družbe Petrol d.d. za leto 2001*. Najdeno 25. julija 2008 na spletnem naslovu http://www.petrol.si/letna_porocila/2001/slo/.
20. Mukund, A. (2003). *Tesco: The Customer Relationship Manager Champion*. Hyderabad: ICFAI Center for Management Research.
21. Parr Rud, O. (2001). *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management*. New York: John Wiley & Sons.
22. Pyle, D. (2003). *Business Modeling and Data Mining*. San Francisco: Morgan Kaufmann Publishers.
23. Ramon, J. et al. (2007). Mining data from intensive care patients. *Advanced Engineering Informatics*, 21 (3), 243–256.
24. Russel, G. J. & Petersen, A. (2000). Analysis of Cross Category Dependence in Market Basket Selection. *Journal of Retailing*, 76 (3), 367-392.
25. Samuelson, P. A. & Nordhaus, W. D. (2002). *Ekonomija*. (slovenski prevod 6. izdaje) Zagreb: GV Založba.
26. Shaw, M. J., Subramaniam, C., Tan, G. T. & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31 (1), 127–137.
27. *Spletna stran CRISP-DM modela*. Najdeno 14. julija 2008 na spletnem naslovu <http://www.crisp-dm.org/Process/index.htm>.
28. *Spletna stran mreže razvijalcev Microsoft (MSDN – Microsoft Developer Network)*. Najdeno 24. julija 2008 na spletnem naslovu <http://msdn.microsoft.com/it-it/library/ms174861.aspx>.
29. *Spletna stran podjetja Petrol*. Najdeno 22. julija 2008 na spletnem naslovu <http://www.petrol.si>.
30. *Spletni tečaj podatkovnega rudarjenja KDnuggets*. Najdeno 7. julija 2008 na spletnem naslovu http://www.kdnuggets.com/data_mining_course/index.html.
31. Sumathi, S. & Sivanandam, S. N. (2006). *Introduction to Data Mining and its Applications*. Berlin: Springer.
32. Witten, I. H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann Publishers.
33. Zanasi, A. (2003). *Analisi dei Dati e Data Mining*. Modena: Facoltà di Scienze della Comunicazione.

PRILOGA 1: Vprašalnik

Merjenje zaznane kakovosti za Petrol

Navedli vam bomo nekaj trditev, vi pa nam povejte, v kolikšni meri se z njimi strinjate:

1. Oprema Petrolovih prodajnih mest je moderna.
2. Petrolova prodajna mesta delujejo privlačno.
3. Petrolova prodajna mesta so čista.
4. Petrolovi prodajalci so primerno oblečeni.
5. Petrolovi prodajalci so primerno urejeni.
6. Na Petrolovih prodajnih mestih kupujem samo gorivo.
7. Ob nakupu goriva vedno kupim še kako malenkost.
8. Na Petrolovih prodajnih mestih se pogosto ustavim samo zaradi nakupa izdelkov dodatne ponudbe (tj. vsega, kar ni gorivo).
9. Dodatna ponudba Petrola (tj. vse, kar ni gorivo) je zelo pestra.
10. Petrolovi prodajalci so vedno prijazni.
11. Petrolovi prodajalci mi znajo dobro svetovati.
12. Petrolovim prodajalcem lahko zaupam.
13. Ker imajo Petrolovi prodajalci veliko dela, je popolnoma razumljivo, da mi ne morejo takoj pomagati.
14. Ker imajo Petrolovi prodajalci veliko dela, je popolnoma razumljivo, da mi ne morejo posvetiti veliko časa.
15. Nerealno je pričakovati, da bodo Petrolovi prodajalci ustregli mojim posebnim željam.
16. Petrolovi prodajalci mi navadno niso pripravljeni pomagati.
17. Na blagajni Petrolovega prodajnega mesta sem hitro na vrsti.
18. Lokacija Petrolovega prodajnega mesta, ki ga najpogosteje obiščem, mi popolnoma ustreza.
19. Delovni čas Petrolovega prodajnega mesta, ki ga najpogosteje obiščem, mi popolnoma ustreza.
20. Nerealno je pričakovati, da bodo vsa Petrolova prodajna mesta odprta 24 ur na dan.

Merjenje pričakovanj

1. Izmed naštetih lastnosti izberite tri, ki so za vas ob obisku prodajnega mesta na črpalki najpomembnejše:
 - Oprema prodajnega mesta.
 - Čistoča prodajnega mesta
 - Urejenost prodajalcev.
 - Prijaznost prodajalcev.
 - Pestrost ponudbe.
 - Hitrost postrežbe pri blagajni.

- Lokacija črpalke.
- Delovni čas črpalke.
- 2. Ali je za vas pomembna še kakšna lastnost, ki je nismo našli?
- Da.
- Ne.

Ocena poslovanja z Magno:

1. Ali ste imetnik Magna kartice?
 - Da.
 - Ne.
2. Kako dolgo ste že imetnik Magna kartice (v letih)?
3. Ali ste v letošnjem letu dobili kakšno obvestilo o ugodnostih, namenjenih imetnikom Magna kartice?
 - Da.
 - Ne.
4. Ali ste v letošnjem letu izkoristili katero od ugodnosti, namenjenih imetnikom Magna kartice?
 - Da.
 - Ne.
5. V kolikšni meri ste zadovoljni z obstoječim pisnim načinom obveščanja o ugodnostih, namenjenih imetnikom Magna kartice?
6. V kolikšni meri ste zadovoljni z načinom obračuna opravljenih transakcij?
7. Ste že kdaj poklicali Magna center s pritožbo?
 - Da.
 - Ne.

Demografska in socioekonomska vprašanja:

1. Spol:
 - Moški.
 - Ženski.
2. Starost (v letih): _____
3. Izobrazba:
 - Nedokončana osnovna šola.
 - Končana osnovna šola.
 - Končana srednja šola.
 - Diploma višje ali visoke stopnje.
 - Magisterij ali doktorat.
4. Zakonski stan:
 - Samski.
 - Poročen oziroma živeč s partnerjem.

- Ločen ali ovdovel.
- 5. Dohodki članov gospodinjstva v prejšnjem mesecu:
 - Do 100.000 SIT.
 - Od 100.000 do 200.000 SIT.
 - Od 200.000 do 400.000 SIT.
 - Od 400.000 do 600.000 SIT.
 - Od 600.000 do 800.000 SIT.
 - Od 800.000 do 1.000.000 SIT.
 - Brez dohodkov.
- 6. Število vzdrževanih družinskih članov: _____
- 7. Stanovanjske razmere:
 - Lastnik ali solastnik hiše ali stanovanja.
 - Najemnik hiše ali stanovanja.
 - Podnajemnik v hiši ali stanovanju.

PRILOGA 2: Seznam podatkov iz baze podatkov

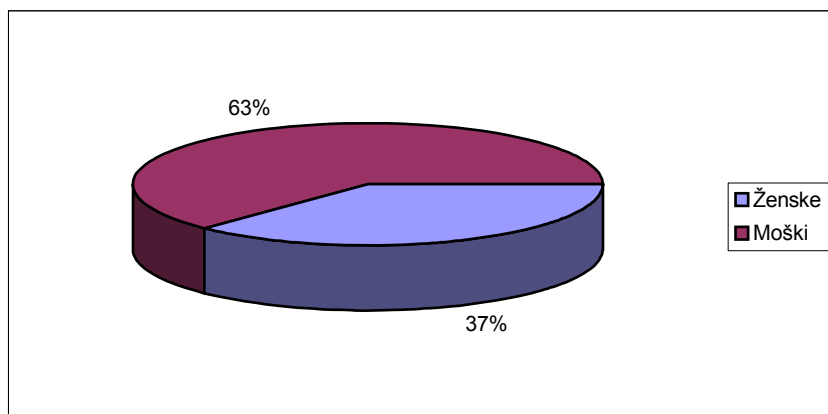
Koda spremenljivke	Opis spremenljivke
GENDER	Spol anketiranca
AGE	Starost anketiranca
RES_OWN	Stanovanjske razmere
ENROL	Leto včlanitve v <i>Magna</i> program
MEMBER	Dolžina članstva v <i>Magna</i> programu (v letih)
EDUCA	Izobrazba
MARRIAGE	Zakonski stan
DEPEND	Število vzdrževanih družinskih članov
INCOME	Dohodki članov gospodinjstva v prejšnjem mesecu (maj 2001)
EXP_1	Prva najbolj pomembna značilnost pri obisku prodajnega mesta.
EXP_2	Prva najbolj pomembna značilnost pri obisku prodajnega mesta.
EXP_3	Prva najbolj pomembna značilnost pri obisku prodajnega mesta.
EXP_ADD	Ali je za vas pomembna še kakšna lastnost, ki je nismo našeli?
EXP_DET	Če da, katera lastnost, ki je nismo našeli, je za vas pomembna?
B_INFO	Ali ste v letošnjem letu dobili kakšno obvestilo o ugodnostih, namenjenih imetnikom Magna kartice?
B_USE	Ali ste v letošnjem letu izkoristili katero od ugodnosti, namenjenih imetnikom Magna kartice?
S_INFO	V kolikšni meri ste zadovoljni z obstoječim pisnim načinom obveščanja o ugodnostih, namenjenih imetnikom Magna kartice?
S_TRAN	V kolikšni meri ste zadovoljni z načinom obračuna opravljenih transakcij?
COMPL	Ste že kdaj poklicali Magna center s pritožbo?

Koda spremenljivke	Opis spremenljivke
SQ_1	Oprema Petrolovih prodajnih mest je moderna.
SQ_2	Petrolova prodajna mesta delujejo privlačno.
SQ_3	Petrolova prodajna mesta so čista.
SQ_4	Petrolovi prodajalci so primerno oblečeni.
SQ_5	Petrolovi prodajalci so primerno urejeni.
SQ_6	Na Petrolovih prodajnih mestih kupujem samo gorivo.
SQ_7	Ob nakupu goriva vedno kupim še kako malenkost.
SQ_8	Na Petrolovih prodajnih mestih se pogosto ustavim samo zaradi nakupa izdelkov dodatne ponudbe (tj. vsega, kar ni gorivo).
SQ_9	Dodatna ponudba Petrola (tj. vse, kar ni gorivo) je zelo pestra.
SQ_10	Petrolovi prodajalci so vedno prijazni.
SQ_11	Petrolovi prodajalci mi znajo dobro svetovati.
SQ_12	Petrolovim prodajalcem lahko zaupam.
SQ_13	Ker imajo Petrolovi prodajalci veliko dela, je popolnoma razumljivo, da mi ne morejo takoj pomagati.
SQ_14	Ker imajo Petrolovi prodajalci veliko dela, je popolnoma razumljivo, da mi ne morejo posvetiti veliko časa.
SQ_15	Nerealno je pričakovati, da bodo Petrolovi prodajalci ustregli mojim posebnim željam.
SQ_16	Petrolovi prodajalci mi navadno niso pripravljeni pomagati.
SQ_17	Na blagajni Petrolovega prodajnega mesta sem hitro na vrsti.
SQ_18	Lokacija Petrolovega prodajnega mesta, ki ga najpogosteje obiščem, mi popolnoma ustreza.
SQ_19	Delovni čas Petrolovega prodajnega mesta, ki ga najpogosteje obiščem, mi popolnoma ustreza.
SQ_20	Nerealno je pričakovati, da bodo vsa Petrolova prodajna mesta odprta 24 ur na dan.

Koda spremenljivke	Opis spremenljivke
TOTAL	Celotni znesek, ki ga je anketiranec potrošil s kartico Magna v zadnjih šestih mesecih.
T_EXT	Celotni znesek, ki ga je anketiranec s kartico Magna v zadnjih šestih mesecih v prodajnih mestih, ki niso lastnina Petrola.
T_INT	Celotni znesek, ki ga je anketiranec potrošil na Petrolovih prodajnih mestih s kartico Magna v zadnjih šestih mesecih.
T_INT1	Celotni znesek, ki ga je anketiranec potrošil za gorivo na Petrolovih prodajnih mestih s kartico Magna v zadnjih šestih mesecih.
T_INT2	Celotni znesek, ki ga je anketiranec potrošil za dodatno blago na Petrolovih prodajnih mestih s kartico Magna v zadnjih šestih mesecih.
POS	Število Petrolovih prodajnih mest, ki jih je imetnik kartice obiskal v zadnjih šestih mesecih.
VISIT1	Število obiskov imetnika kartice na prvem najbolj obiskanem Petrolovem prodajnem mestu.
VISIT2	Število obiskov imetnika kartice na drugem najbolj obiskanem Petrolovem prodajnem mestu.
VISIT3	Število obiskov imetnika kartice na tretjem najbolj obiskanem Petrolovem prodajnem mestu.
SUM2	Seštevek VISIT1 in VISIT2.
SUM3	Seštevek VISIT1, VISIT2 in VISIT3.
SUMT	Celotno število obiskov imetnika kartice na vseh prodajnih mestih Petrola v zadnjih šestih mesecih.
SHARE1	Odstotek VISIT1 v SUMT.
SHARE2	Odstotek SUM2 v SUMT.
SHARE3	Odstotek SUM3 v SUMT.

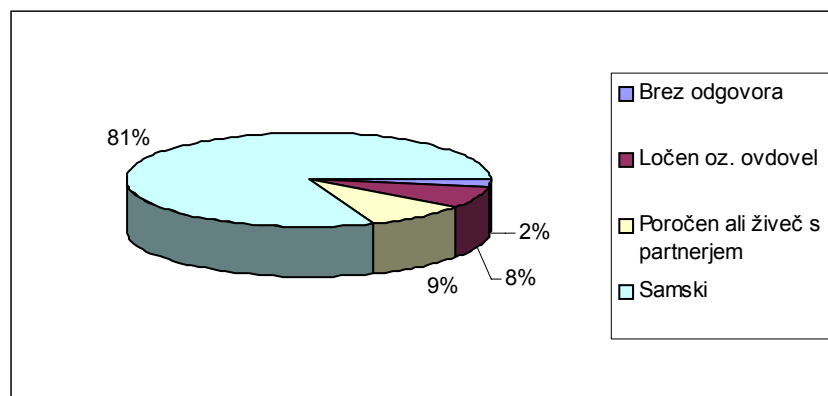
PRILOGA 3: Demografski podatki anketirancev

Slika 1: Razporeditev anketirancev po spolu



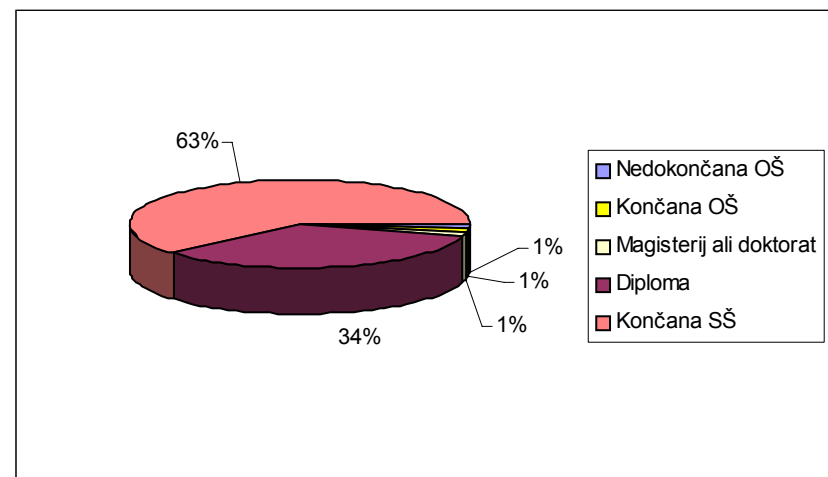
Vir: Lastna izdelava

Slika 2: Razporeditev anketirancev po zakonskem stanju



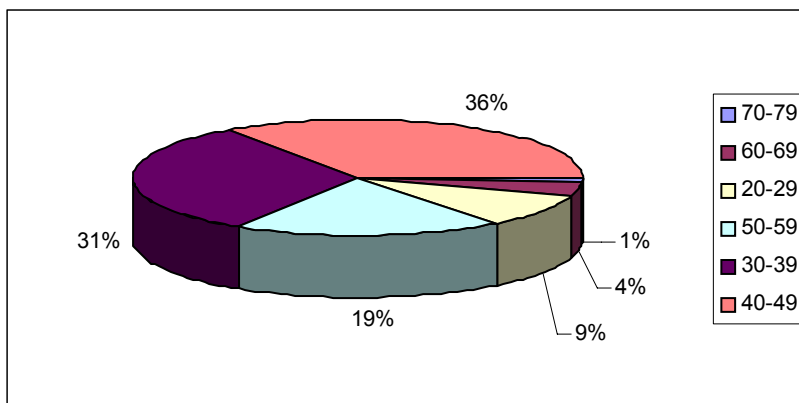
Vir: Lastna izdelava

Slika 3: Razporeditev anketirancev po izobrazbi



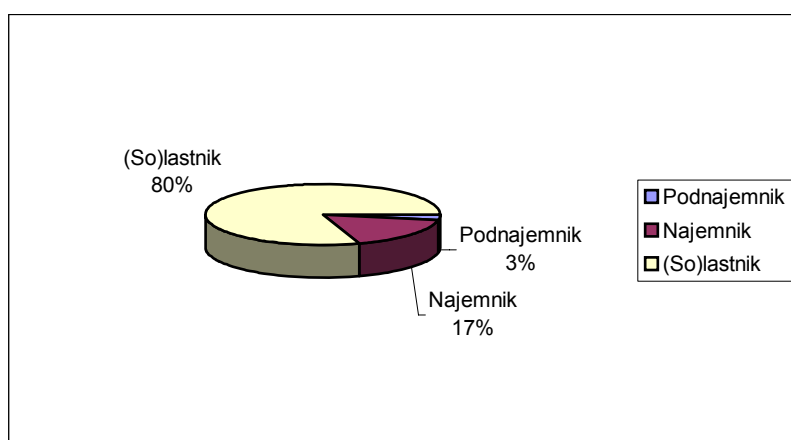
Vir: Lastna izdelava

Slika 4: Razporeditev anketirancev po starosti



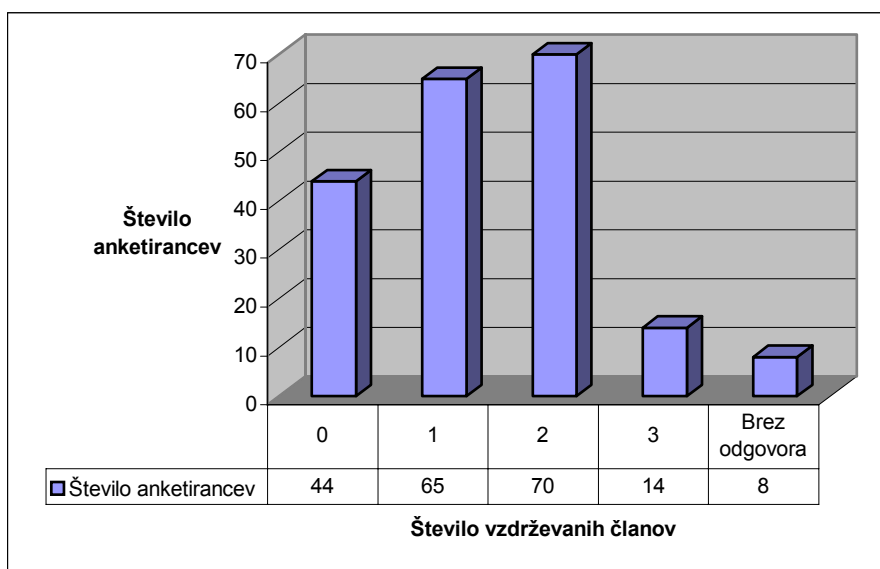
Vir: Lastna izdelava

Slika 5: Razporeditev anketirancev glede na stanovanjske razmere



Vir: Lastna izdelava

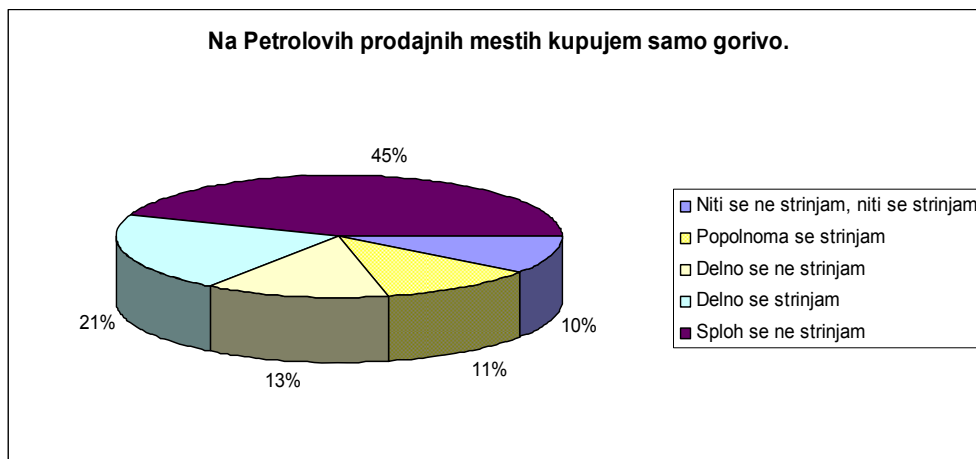
Slika 6: Število vzdrževanih družinskih članov



Vir: Lastna izdelava

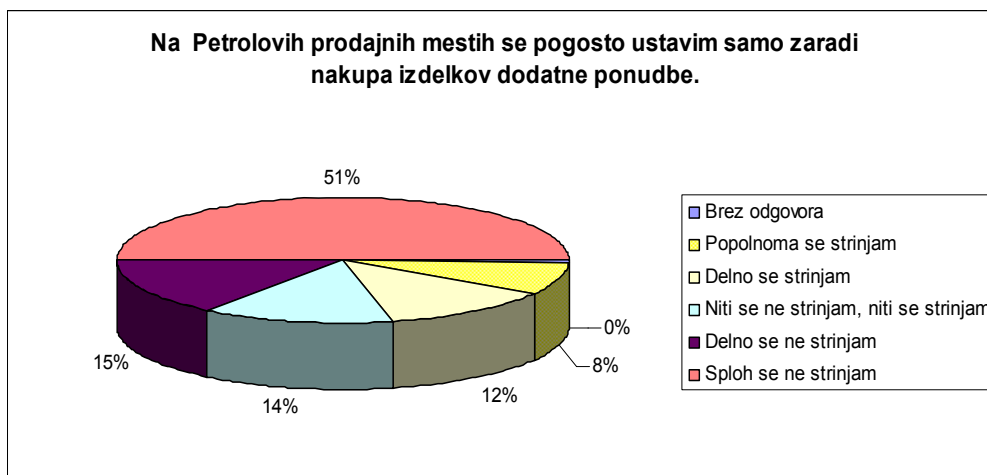
PRILOGA 4: Nakupno obnašanje anketirancev

Slika 7: Delitev kupcev na podlagi tega, ali kupujejo samo gorivo



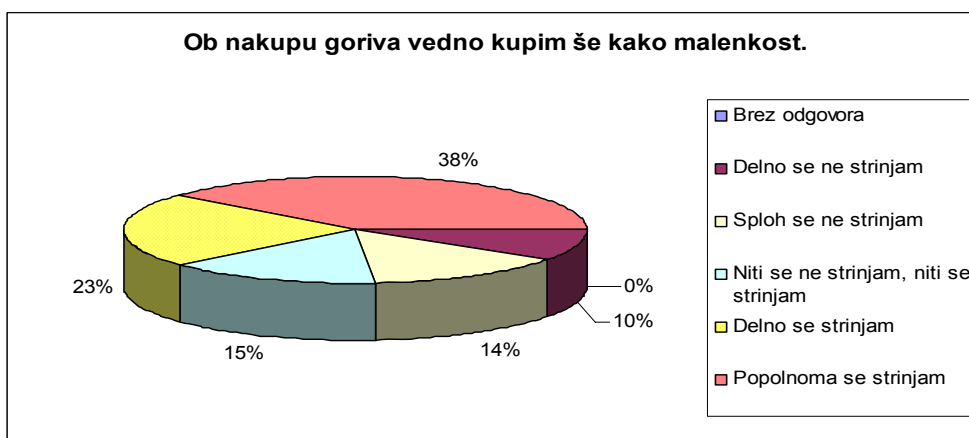
Vir: Lastna izdelava

Slika 8: Delitev kupcev na podlagi tega, ali pogosto kupujejo le izdelke dodatne ponudbe



Vir: Lastna izdelava

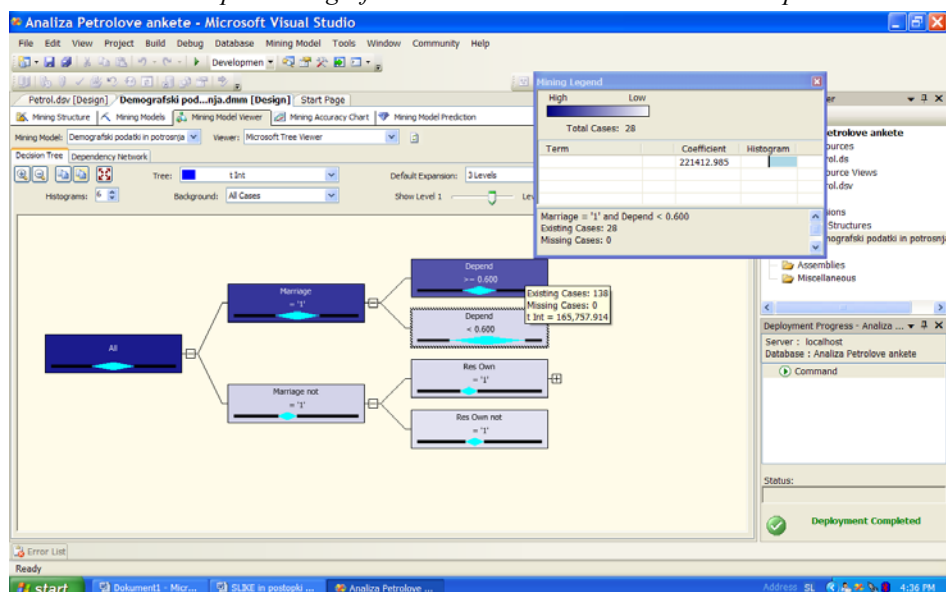
Slika 9: Delitev kupcev glede tega, ali kupujejo ob gorivu še kako malenkost



Vir: Lastna izdelava

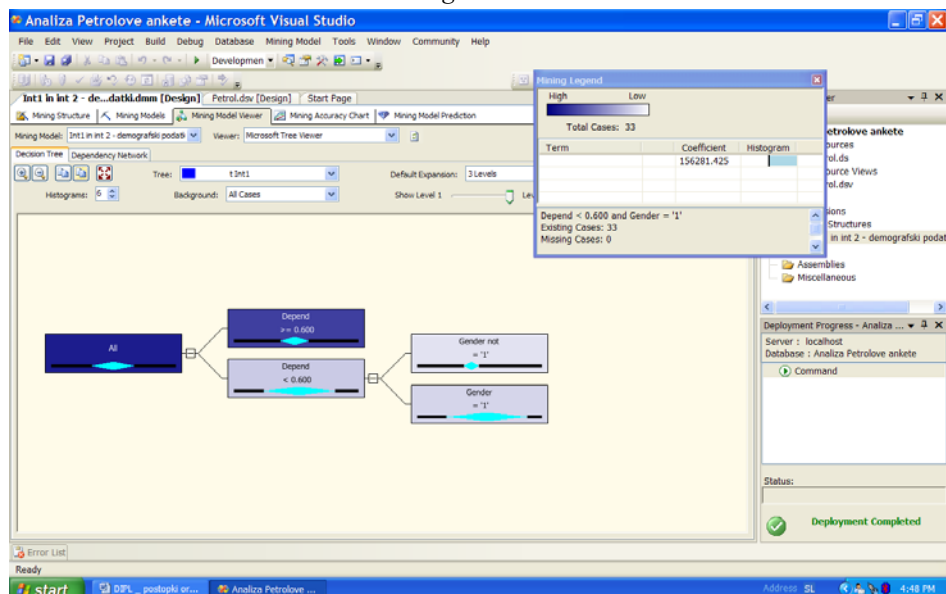
PRILOGA 5: Analiza vpliva demografskih in socioekonomskih značilnosti kupcev na potrošnjo

Slika 10: Odločitveno drevo: vpliv demografskih in socioekonomskih značilnosti kupcev na celotno potrošnjo



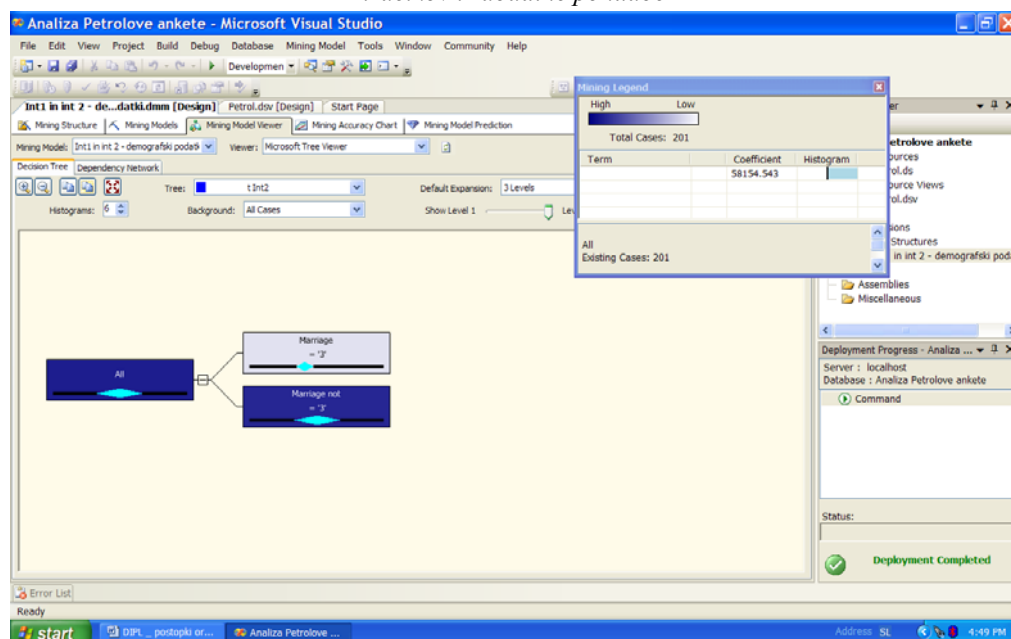
Vir: Lastna izdelava

Slika 11: Odločitveno drevo: vpliv demografskih in socioekonomskih značilnosti kupcev na potrošnjo za nakup goriva



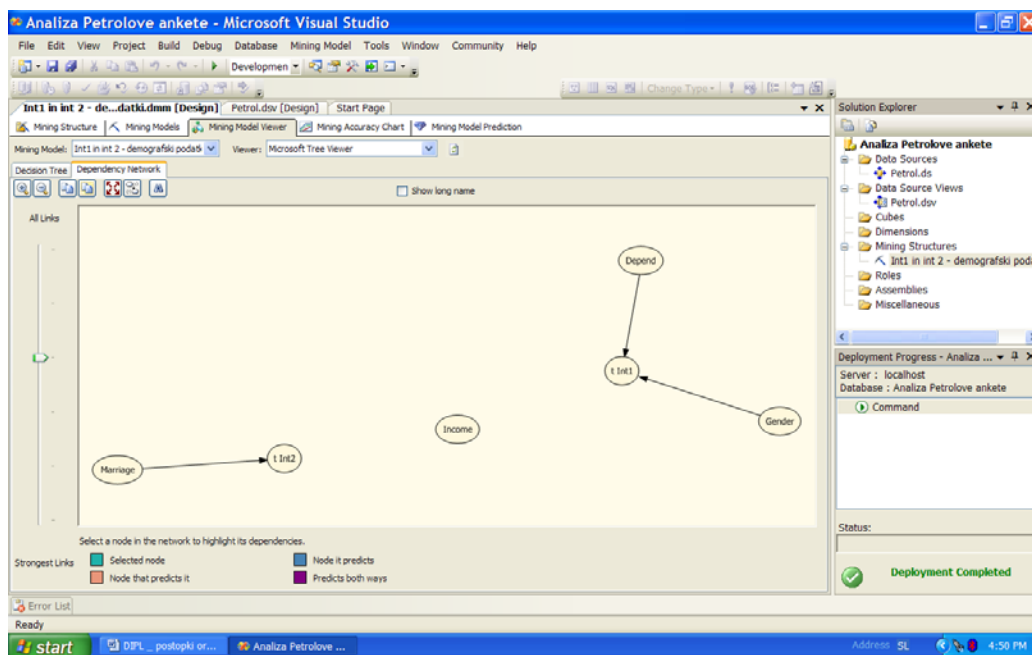
Vir: Lastna izdelava

Slika 12: Odločitveno drevo: vpliv demografskih in socioekonomskih značilnosti kupcev na potrošnjo za nakup izdelkov iz dodatne ponudbe



Vir: Lastna izdelava

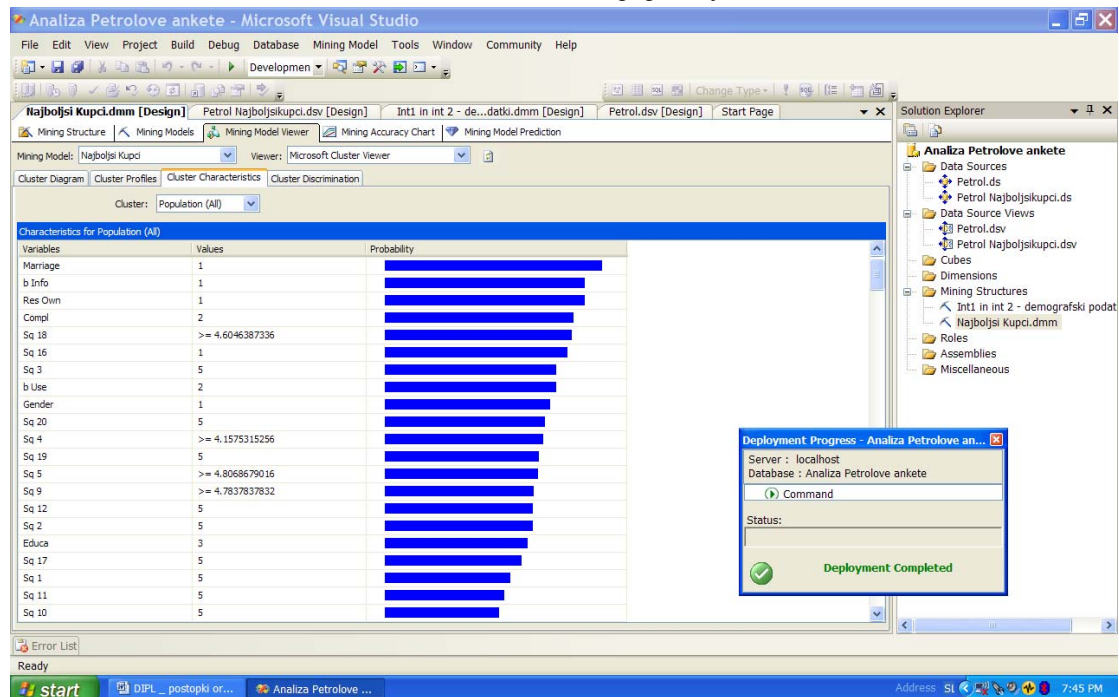
Slika 13: Mreža odvisnosti



Vir: Lastna izdelava

PRILOGA 6: Analiza profila najboljših kupcev

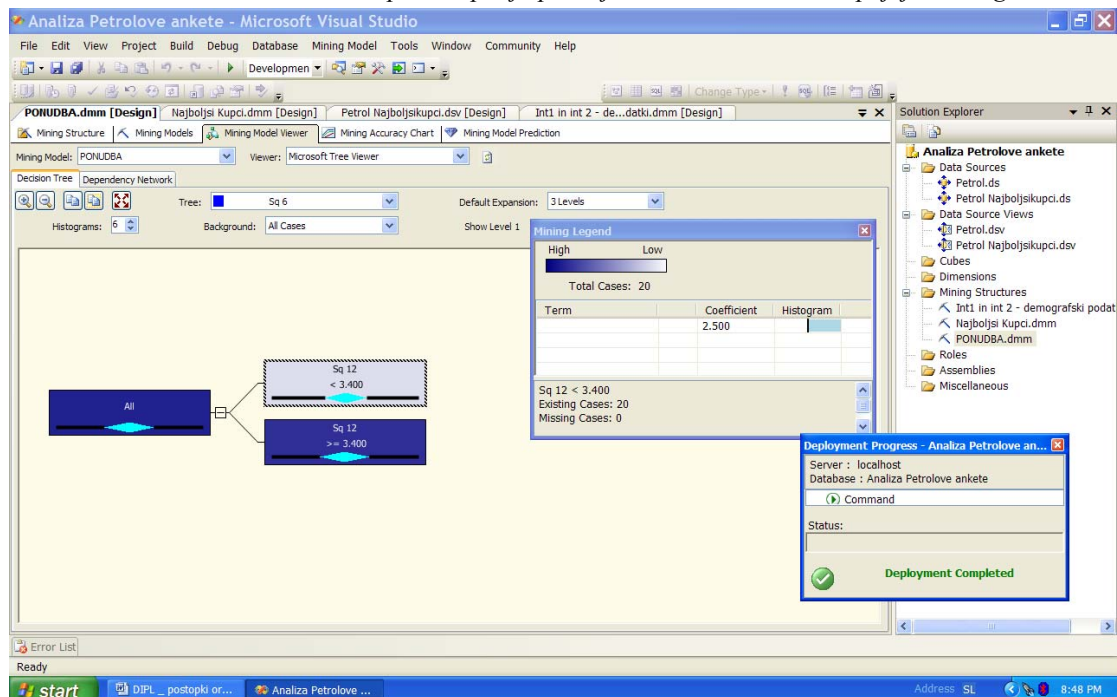
Slika 14: Značilnosti populacije



Vir: Lastna izdelava

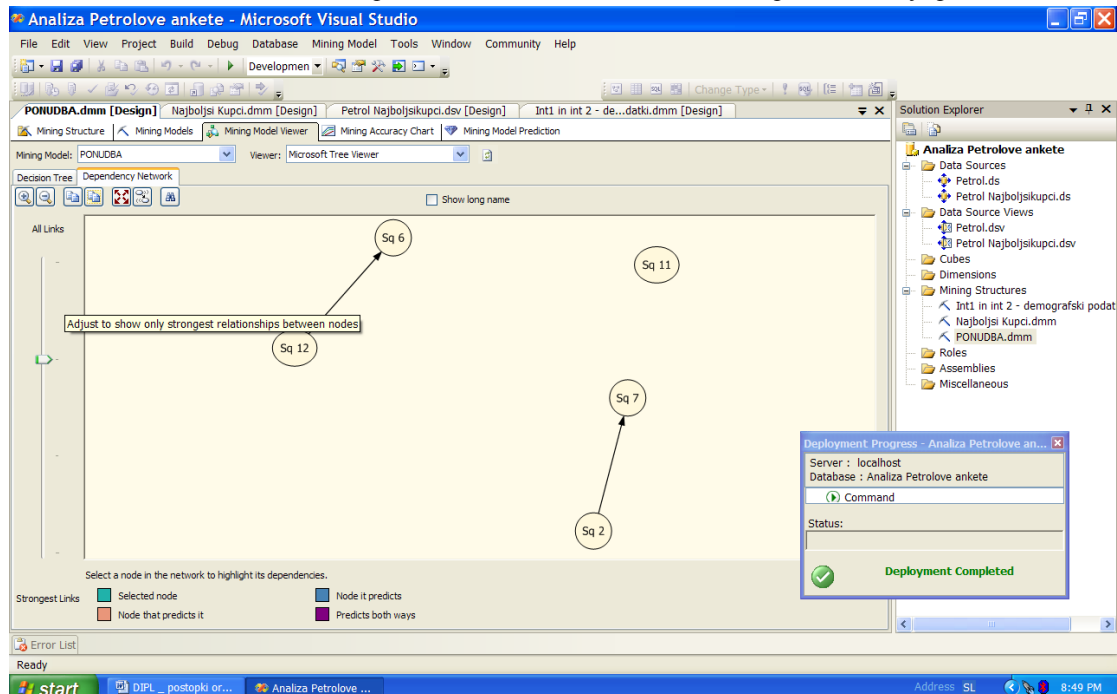
PRILOGA 7: Analiza vpliva mnenj strank z različnimi nakupnimi obnašanji o kakovosti za Petrol

Slika 15: Odločitveno drevo: vpliv zaupanja prodajalcem na stranke, ki kupujejo samo gorivo



Vir: Lastna izdelava

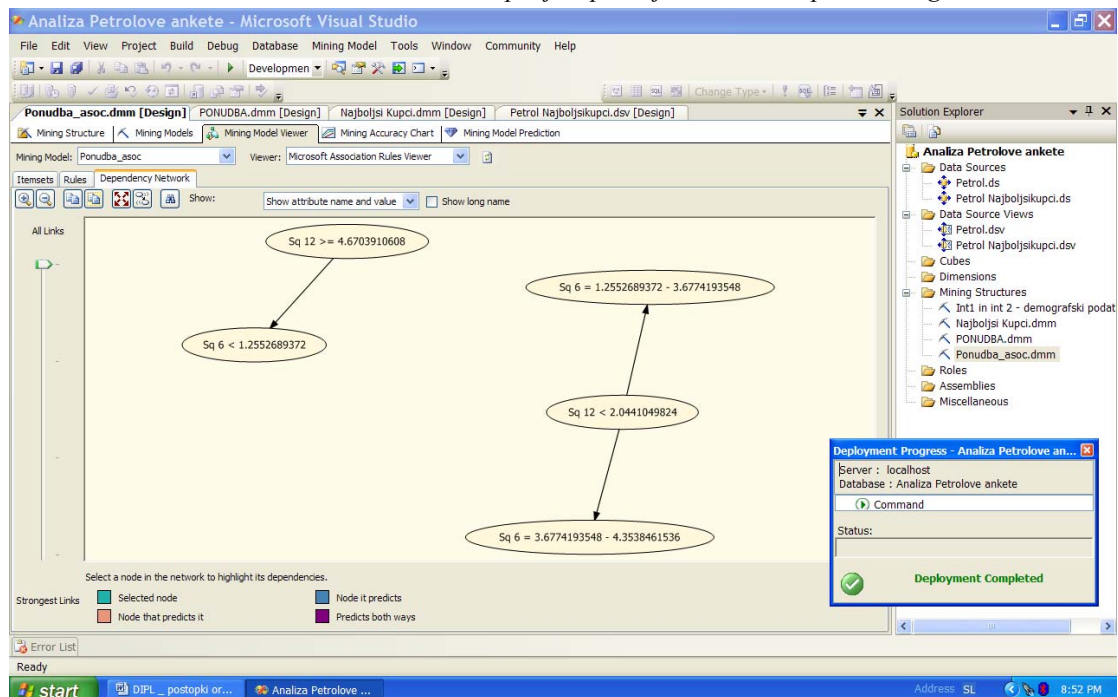
Slika 16: Mreža odvisnosti: Vpliv ocen kakovosti za Petrol na nakupno obnašanje potrošnikov



Vir: Lastna izdelava

PRILOGA 8: Asociativna pravila: vpliv mnenja o zaupanju prodajalcem na potrošnjo samo goriva

Slika 17: Mreža odvisnosti med zaupanjem prodajalcem in nakupom samo goriva



Vir: Lastna izdelava

Slika 18: Asociativna pravila

Probability	Importance	Rule
0.508	0.163	Sq 12 >= 4.6703910608 -> Sq 6 < 1.2552689372
0.500	0.330	Sq 12 < 2.0441049824 -> Sq 6 = 1.2552689372 - 3.6774193548
0.500	0.379	Sq 12 < 2.0441049824 -> Sq 6 = 3.6774193548 - 4.3538461536

Vir: Lastna izdelava

PRILOGA 9: Slovarček slovenskih prevodov tujih izrazov

ANGLEŠKI IZRAZ	SLOVENSKI IZRAZ
Activation	aktiviranje strank
Artificial Intelligence	umetna inteligenca
Artificial Neural Network	umetna nevronska mreža
Association Analysis	analiza asociacij
Association Rules	asociativna pravila
Attrition	obraba
Binary Attributes	binarne spremenljivke
Categorical Variable	kategorična spremenljivka
Churn	mešalnik
Classification	klasificiranje
Clustering	razvrščanje v skupine
Confidence	zaupanje
Customer Relationship Management	upravljanje odnosov s strankami
Cross-Selling	navzkrižna prodaja
Data Mining	podatkovno rudarjenje
Data Visualization Techniques	vizualizacijska orodja
Decision Tree	odločitveno drevo
Description	opisovanje
Distance Function	funkcija razdalje
Estimation	ocenjevanje
Input	vhod
Interval Variable	intervalna spremenljivka
Machine Learning	strojno učenje
Market Basket Analysis	analiza nakupne košarice
Market Segmentation	segmentiranje trga
Nearest Neighbors	najbližji sosed
Nominal Variable	nominalna spremenljivka
Ordinal Variable	ordinalna spremenljivka
Predict Only	napovedna spremenljivka
Prediction	napovedovanje
Profile Analysis	izdelava profilov kupcev
Ratio Variable	razmernostna spremenljivka
Response Model	model reakcije strank
Risk Management	upravljanje s tveganji
Support	podpora
Unary Attributes	enostavne spremenljivke
Up-Selling	več enakih izdelkov
Weighted Voting	tehtano glasovanje