

Primena Data Mining-a u cilju poboljšanja politike zapošljavanja u malim i srednjim maloprodajnim preduzećima

Ulfeta A. Marovac, Univerzitet u Novom Pazaru, ul. Vuka Karadžića bb

Sadržaj — Cilj ovog rada je da se uradi analiza uticaja starosne strukture prodavaca u malim i srednjim preduzećima primenom odgovarajućeg softverskog paketa i tehnike data mining-a. Na osnovu analize određen je kriterijum pri zapošljavanju novih radnika u maloprodajnim preduzećima. Izvršena je analiza zavisnosti kvaliteta prodaje od starosne strukture prodavaca na uzorku od 414 različitih maloprodajnih firmi čije je poslovanje praćeno polugodišnjim izveštajima u vremenskom periodu od tri godine (2004-2006). Istraživanje je vršeno neuralnim klasterovanjem jednom od tehnika data mining-a. Kao rezultat dobijamo šablone u obliku klastera sa odgovarajućim karakteristikama koje mogu opisivati dobru ili lošu prodaju i odgovarajućom starosnom strukturom prodavaca koja je za to odgovorna.

Ključne reči — Data mining, Neuralno klasterovanje, Neto zarada, Dugoročna dugovanja, Trenutna dugovanja, Stalna imovina

I. UVOD

USPEŠNOST malih i srednjih preduzeća u velikoj meri zavisi od starosne strukture zaposlenih a posebno od prodavaca u maloprodajnim objektima. Ova preduzeća nemaju posebne službe koje se bave upravljanjem kadrova i strategijom zapošljavanja i po pravilu se oslanjaju na iskustvo vlasnika preduzeća i opšte-prihvaćena mišljenja koja nisu uvek tačna. Pogrešan izbor zaposlenih može imati veoma negativan uticaj na poslovanje preduzeća a da vlasnici nisu svesni uzroka loših poslovnih rezultata. Mnoga istraživanja i postojanje savremenih softverskih paketa često nisu dostupna malim i srednjim preduzećima iako bi njihova primena mogla da ima značajan uticaj na ostvarivanje dobrih poslovnih rezultata.

Univerzitet u Novom Pazaru je realizator projekta "Razvoj softverskih sistema za podršku poslovanja malih i srednjih preduzeća" koje finansira Ministarstvo za nauku. Cilj projekta je da se koriste poznata softverska rešenja i razvijaju sopstveni softverski proizvodi koji će moći da budu korišćeni od strane malih i srednjih preduzeća u cilju poboljšanja poslovanja. Jedan od ključnih problema je politika zapošljavanja i primene savremenih softverskih rešenja u donošenju odluke za pravilan izbor prodavaca u maloprodaji. Jedna od aktivnosti projekta je da se izvrši istraživanje u cilju pronalazjenja zavisnosti kvaliteta prodaje od životnog doba prodavca tehnikama data

Ovaj rad finansiralo je Ministarstvo nauke Republike Srbije, Projekat tehnološkog razvoja TR-13012.

minig-a. Ovaj rad je nastao kao želja da se pokaže da je primena data mining tehnika jedan neizostavan deo u modernom poslovanju malih i srednjih preduzeća. Data mining treba razlikovati od automatske obrade podataka i primene statističkih alata. Data mining se razlikuje od klasičnih statističkih metoda po tome što se ne odvija po unapred određenim pravilima već pokazuje kreativnost u analizi podataka i na taj način može da otkrije nova neočekivana pravila [1].

II. OPIS POLAZNIH PODATAKA

Kako još uvek nije dostupna realna baza o poslovanju preduzeća u Novom Pazaru rad je realizovan nad modelom i bazom podataka o poslovanju maloprodajnih preduzeća iz različitih zemlja koja je učesnicima u projektu bila dostupna preko interneta. Baza na kojoj se vrši ispitivanje sastoji se od šest tabela u kojima se nalaze podaci o sedištu firme, vrsti artikala koji se u njoj prodaju, struktura prodaje po pripadnosti prodavaca određenom starosnom dobu i polugodišnji izveštaji o prodaji za prodavce (prihodi, troškovi, dobit, i ostali podaci).

III. DEFINICIJE

U stručnoj terminologiji se koriste različiti pojmovi, na primer umesto preduzeća se koristi privredno društvo, ali bi tada mnogi odomaćeni termini kao što je mala i srednja preduzeća bio rogovatan u radovima kao što je ovaj. Radi lakšeg praćenja ovog rada koristiće se terminologija koja je razumljivija vlasnicima malih i srednjih preduzeća kako bi mogli da koriste i primenjuju rezultate ovog rada:

- Stalna imovina – sva imovina koja pripada jednom preduzeću izuzev robe koju prodaje.
- Neto zarada – razlika ukupnog prihoda i ukupnih troškova daje neto zaradu.
- Trenutna dugovanja – sve trenutno neizmirene obaveze kao što su dug za robu, porez, troškovi za vodu, struju, i ostale dažbine.
- Dugoročna dugovanja – na primer dugovanja u obliku dugoročnih kredita.

IV. SOFTVERSKI ALAT

Alat koji je korišćen za istraživanje je softverski paket DB2 Intelligent Miner firme IBM. Ovaj alat nam omogućava korišćenje niza tehnika data mining-a, statističkih metoda itd., kao i mogućnost vizuelnog

predstavljanja rezultata koji su dobijeni [2]. Baza je obrađivana takođe u IBM-ovom alatu DB2.

V. IZBOR METODA DATA MINING

Za rešenje ovog problema koristili smo klasterovanje. Problem bi mogao da se reši i metodom klasifikacije međutim problem je naći algoritam za određivanje klasa. U našem zadatku bi se javio problem na koji način preslikati neke argumente u klasu koja određuje kvalitet prodaje. S obzirom da nemamo unapred definisane klase nama će više odgovarati da slične podatke grupišemo u klaster koji naravno nisu unapred definisani a onda da procenjujemo kvalitet prodaje pripadnika klastera. Kako radimo sa neprekidnim promenljivim za nas je pogodno neuralno klasterovanje. Veštačke neuralne mreže su analitičke tehnike koje su formirane na osnovu pretpostavljenog procesa učenja u ljudskom mozgu. Kao što je ljudski mozak sposoban da posle procesa učenja izvlači pretpostavke na osnovu ranijih opažanja, tako su i neuralne mreže sposobne da nakon procesa učenja predvide promene i dešavanja u sistemu [1].

VI. ANALIZA PROBLEMA I NJEGOVOG REŠENJA

Na osnovu ovih ponuđenih podataka treba naći zavisnost prodaje od starosne strukture prodavaca. Kao pod-problem javlja se nalaženje formule ili definicije kvaliteta prodaje. Pitanje kvaliteta je diskutabilno i može se posmatrati iz više uglova. Problem je odrediti tačnu definiciju šta je kvalitetnija prodaja. U svakom slučaju, prodaja je bolja ako su dugoročna dugovanja manja, stalna imovina veća, trenutno neizvršene obaveze manje, ili neto zarada veća. Ono što je potrebno naći je razlika između prodavaca koji rade u sličnim firmama, na primer po veličini, i dati njihove razlike zavisno od godina.

Problem smo posmatrali sa različitih strana. Kao prvo ispitali smo koliko procentualno, prosečno prodaje pripada prodavcima iz različitih starosnih grupa. Izvršen je pregled proseka i po svim parametrima prodaje za svaku od starosnih grupa. Kako imamo uzorak u kome se nalaze nejednake firme, neke manje a neke veće, odlučili smo da izvršimo klasterovanje jednom po neto prihodu a drugi put po dugoročnim zaduženjima, a zatim da posmatramo prosečno učešće svake od starosnih grupa prodavaca po klasterima. Zaključili smo da su nam četiri klastera sasvim dovoljna i da sa više klastera dolazi do prevelikog usitnjavanja. Ovim klasterovanjem dobijeni su rezultati o zastupljenosti pripadnika pojedinih starosnih struktura po klasterima. Iz grafika klasterovanja koje je izvršeno po oba parametra zaključili smo da se mora ići dublje u analizu jer je porast neto dohotka po klasterima bio praćen porastom dugovanja čime nismo dobijali odgovarajuću sliku o kvalitetu prodaje. Ovim je izvršena podela firmi po veličini zato što se dobija da oni koji puno zarađuju istovremeno se puno zadužuju, i obrnuto. Iz tog razloga smo odlučili da izvršimo četiri nova klasterovanja po dva parametra zajedno (neto prihodu i dugoročnim dugovanjima) za svaki klaster klasterovanja izvršenog po neto prihodu sa početka istraživanja. Na ovaj način dobijamo šesnaest klastera, ali kontrolisano, i što će se

dalje videti u njima dolazi do izražaja kvalitet prodaje.

VII. REŠENJE PROBLEMA

A. Pregled prodaje po starosnim grupama na celom uzorku

Grupisanjem prodavaca po starosnoj grupi i određivanjem prosečnog procenta učešća u prodaji za svaku starosnu grupu dobijamo da najviše udela u prodaji na posmatranom uzorku imaju prodavci starosti od 21 do 30 godina pa se mogu smatrati radno najaktivnijom starosnom grupom. Ostatak prodavaca većim delom ima od 31 do 50 godina. I naravno prodavaca sa više od 61 godina ima najmanje.

Izračunat je prosek po parametrima kvaliteta prodaje (dugoročna dugovanja, stalna imovina, trenutno neizvršene obaveze i neto zarada) za svaku generaciju srazmerno njihovom udelu u firmi. Dobijeni rezultati ukazuju na veliku neto zaradu i stalnu imovinu kod prodavaca od 21 do 30 godina, ali takođe i velika zaduženja. Kod ostalih starosnih grupa prihod i rashod se srazmerno njihovom učešću smanjuju.

B. Primena metode klasterovanja na dati problem

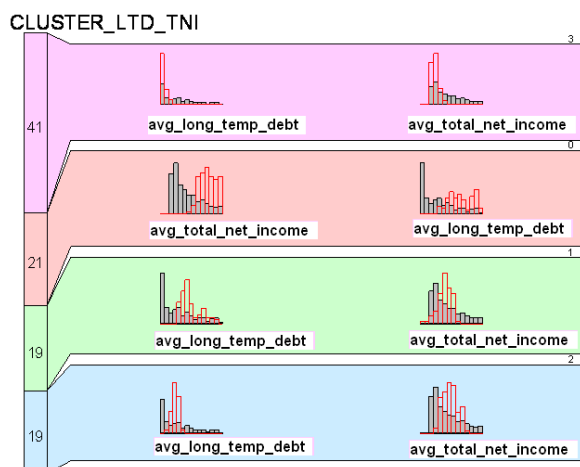
Prvo klasterovanje koje je izvršeno na celom uzorku po neto zaradi daje kao rezultat četiri klastera. Ako pogledamo starosnu strukturu ovih klastera možemo primetiti da se u klasteru čiji prodavci imaju veliku neto zaradu nalazi se najmanje pripadnika od 31 do 40 i od 51 do 60 u odnosu na druge klaster, a najviše ima pripadnika druge starosne grupe. Konkretno u ovom klasteru mladi ostvaruju veću neto zaradu. Dalje, iz ostalih klastera se može zaključiti da je pomeranje starosne granice na gore praćeno smanjenjem neto zarada. Međutim kako se radi o malim razlikama procentualne zastupljenosti određene starosne grupe prodavaca po klasterima ispitivanje je produženo sa ciljem nalaženja sigurnije veze.

Analogno prethodnom izvršeno je klasterovanje na celom uzorku po dugoročnim dugovanjima i izvršen uvid u starosnu strukturu klastera. Kao rezultat dobijeno je da se najmlađi najviše zadužuju a da sa porastom godina prodavaca dugovi su manji.

Kada je izvršeno klasterovanje po neto zaradi i dugoročnim dugovanjima došlo se do zaključka da su prethodnim klasterovanjima dobijeni isti podskupovi jer visokim neto zaradama odgovaraju visoka zaduženja i obrnuto. Sledeća slika prikazuje rezultat ovog klasterovanja.

Na slici su klasteri označeni različitim bojama roze, narandžasto, zeleno i plavo. Na levoj strani stoji broj koji označava procent uzorka koji je ušao u klaster (41, 21, 19 i 19). Na desnoj strani je redni broj klastera (3, 0, 1, 2). Za svaki klaster su predstavljena dva grafika jedan za prosečnu neto zaradu (avg_total_net_income) i drugi za prosečnu dugoročna dugovanja (avg_long_temp_debt). Na graficima se izdvajaju sivi stubovi koji predstavljaju na koji su način raspoređene vrednosti promenljive prosečna neto zarada (avg_total_net_income) odnosno prosečna dugoročna dugovanja (avg_long_temp_debt) na celom uzorku, a dodatnim crvenim stubovima je označeno kako su te vrednosti raspoređene na delu uzorka koji pripada

posmatranom klasteru [3]. Sa slike je vidljiva srazmernost neto zarade i dugoročnih dugovanja.



Sl. 1. Grafički prikaz klastera dobijenih klasterovanjem po neto zaradi i dugoročnim dugovanjima na celom uzorku.

Problem koji je nastao je usled nejednakih prometa kod firmi koje obrađujemo zato smo se na kraju odlučili da klastere dobijene po neto zaradi ponovo podelimo na nova četiri klastera ali sada i po zaradi i po dugovanjima i da uočimo da li među sličnima tačnije onima sa sličnim prometom ima razlika i ko se po čemu izdvaja.

Novo klasterovanje u svaka četiri klastera dalo je po dva specifična klastera. U jednom imamo visoku neto zaradu i niska dugovanja a u drugom nisku neto zaradu a visoka dugovanja. Ovo su situacije kada možemo reci da su prodavci dobro odradili posao ili u drugom slučaju da su ga loše odradili. Ostali klasteri predstavljaju prodavce sa srazmerno niskim ili visokim neto zaradama i dugoročnim dugovanjima pa za njih ne možemo reći ni da su uspešni niti da nisu uspešni u prodaji. Starosne strukture svaka dva klastera se razlikuju zavisno kojoj grupi po neto zaradi pripadaju. Ako pođemo od onih koji imaju najmanju neto zaradu starosnih struktura dva karakteristična klastera može se zaključiti da bolje posluju preduzeća čiji je prosečan broj godina radnika manji. Kako posmatramo dalje preduzeća sa većom zaradom zaključujemo da se starosna struktura menja tako što poziciju boljih prodavaca zauzimaju sve stariji da bi na kraju kod onih sa najvećim prometom se pokazalo da stariji ostvaruju najbolje poslovanje a mlađi najgore. Ovim smo došli do rešenja koje nije jedinstveno za ceo uzorak već definisano na četiri podskupa posebno.

VIII. ZAKLJUČAK

Data mining tehnikom došli smo do neočekivanog i neuniformnog rezultata što je bio slučaj sa ispitivanjima vršenim tradicionalnim tehnikama. I ako se za prodavce od 21 do 30 godina može reći da najviše rade, ali i da najviše troše, titula najboljih prodavaca nije uvek njihova. Kod velikih baza podataka sa puno različitih objekata ne sme se biti suviše grub u donošenju odluka. Zato i u data

mining-u bitnu ulogu igra opet čovek koji određuje koje će metode koristiti i na kom delu uzorka i na koji način će čitati dobijene rezultate. Ovim istraživanjem firme koje su njime obuhvaćene dobile su jasne smernice na koji način treba vršiti odabir radnika. Kao što smo videli nema uniformnog zaključka i treba prvo pronaći sebe u jednoj od posmatranih podgrupa, a onda prema njoj birati starosnu strukturu prodavaca. U firmama sa jako velikim prihodima i rashodima bolje se pokazuju stariji prodavci dok u firmama u kojima su niski prihodi bolje se pokazuju mlađi prodavci. Pa bi generalno za veći obrt novca trebalo angažovati osobe sa više iskustva dok za manje firme predlog bi bio da treba zaposliti mlađe energičnije prodavce. Kako je ispitivanje izvršeno na uzorku sa firmama koje se bave maloprodajom različitih artikala dalje istraživanje bi se moglo nastaviti u smeru određivanja zavisnosti kvaliteta prodaje od starosne strukture prodavaca i vrste artikala koji prodaju. Očekujemo da će se na primer kod sportske opreme i kompjuterskih delova bolje snaći mlađi prodavci, ali takođe očekujemo da primenom data mining-a dobijemo neočekivane rezultate, kao što smo ovim istraživanjem dobili da praksa zapošljavanja samo mladih radnika nije uvek i najefikasnija.

LITERATURA

- [1] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, ISBN 0-321-20448-4
- [2] IBM DB2 Intelligent Miner for Data, Using the Intelligent Miner for Data
- [3] IBM DB2 Intelligent Miner Visualization, Using the Intelligent Miner Visualizers

ABSTRACT

The purpose of this work is to analyze influence of age structure of sellers in small and medium-sized enterprises by using proper software package and data mining technique. The criteria while employing new employers in retail trade is defined based on the analysis. The analysis of the quality of sales dependency from age structure of sellers is carried out on the sample of 414 different retail trades whose businesses have been followed up through semi-annual financial report for the period of three years (2004-2006). The research is carried out by neural clustering, one of data mining methods. As a result we get cluster templates with the proper attributes which may describe good or bad sale and with proper age structure of the seller who is responsible for that.

APPLICATION OF THE DATA MINING IN IMPROVING THE EMPLOYMENT POLICY IN SMALL AND MEDIUM-SIZED ENTERPRISES

Ulfeta A. Marovac.

