Metode strojnog učenja u predviđanju profitabilnosti kupaca

Ćorić, Ivica

Doctoral thesis / Disertacija

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Split, Faculty of economics Split / Sveučilište u Splitu, Ekonomski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:124:612459

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

REFST - Repository of Economics faculty in Split

UNIVERSITY OF SPLIT

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U MOSTARU
EKONOMSKI FAKULTET
SVEUČILIŠTE U SPLITU
EKONOMSKI FAKULTET

Ivica Ćorić

METODE STROJNOG UČENJA U PREDVIĐANJU PROFITABILNOSTI KUPACA

DOKTORSKI RAD

Mentor: prof.dr.sc., Mirjana Pejić-Bach

Mostar, 2017

UNIVERSITY OF MOSTAR
FACULTY OF ECONOMICS
UNIVERSITY OF SPLIT
FACULTY OF ECONOMICS

Ivica Ćorić

MACHINE LEARNING METHODS IN CUSTOMER PROFITABILITY PREDICTION

DOCTORAL THESIS

Mentor: prof.dr.sc., Mirjana Pejić-Bach

Mostar, 2017

PODACI I INFORMACIJE O DOKTORANDU

Ime i prezime: Ivica Ćorić
Datum i mjesto rođenja: 29.1.1968., Mostar, BiH
Naziv završenog fakulteta i godina diplomiranja: Ekonomski fakultet, Sveučilište u Mostaru, 1992.

INFORMACIJE O DOKTORSKOJ DISERTACIJI

Naslov disertacije: **METODE STROJNOG UČENJA U PREDVIĐANJU PROFITABILNOSTI KUPACA**

Fakultet na kojem je disertacija obranjena: Ekonomski fakultet Sveučilišta u Mostaru

POVJERENSTVA, OCJENA I OBRANA DOKTORSKE DISERTACIJE

Datum prijave disertacije: 19.11.2013.

Povjerenstvo za ocjenu teme doktorske disertacije i predlaganje mentora/ocjenu i obranu teme:

1. prof. dr.sc. Vanja Bevanda, Fakultet ekonomije i turizma „dr.Mijo Mirković“ Sveučilišta u Puli – predsjednik,
2. prof. dr. sc. Mirjana Pejić Bach , Ekonomski fakultet Sveučilišta u Zagrebu, mentor i član
3. doc. dr. sc. Tina Vuko , Ekonomski fakultet Sveučilišta u Splitu, član

Povjerenstvo za ocjenu disertacije:

1. prof. dr.sc. Vanja Bevanda, Fakultet ekonomije i turizma „dr.Mijo Mirković“ Sveučilišta u Puli – predsjednik,
2. prof. dr. sc. Mirjana Pejić Bach , Ekonomski fakultet Sveučilišta u Zagrebu, mentor i član
3. doc. dr. sc. Tina Vuko , Ekonomski fakultet Sveučilišta u Splitu, član

Povjerenstvo za obranu doktorske disertacije:

1. prof. dr.sc. Vanja Bevanda, Fakultet ekonomije i turizma „dr.Mijo Mirković“ Sveučilišta u Puli – predsjednik,
2. prof. dr. sc. Mirjana Pejić Bach , Ekonomski fakultet Sveučilišta u Zagrebu, mentor i član,
3. doc. dr. sc. Tina Vuko , Ekonomski fakultet Sveučilišta u Splitu, član ,
4. prof. dr.sc. Brano Markić, Ekonomski fakultet Sveučilišta u Mostaru, zamjenski član,
5. mr.sc. Mirela Maibć, tajnik/zapisničar.

Datum obrane doktorske disertacije: 13.2.2017.

Ovaj rad posvećujem mojim roditeljima, majci Mili i pokojnom ocu Anti, koji su sa svojim skromnim obrazovanjem, ali ogromnom ljubavlju i ljudskom mudrošću usadili u mene bezgraničnu vjeru u obitelj, znanje, poštenje i čovjeka; te mojoj obitelji, supruzi Sandri i djeci, koji su mi vječita inspiracija i poticaj za svaki naredni korak u životu.

SAŽETAK

Suvremeni trendovi izučavanja profitabilnosti kupaca doveli su do saznanja da troškovi proizvoda čine samo jedan dio troškova poslovnog odnosa na relaciji poduzeće-kupac. Ovaj odnos je opterećen i različitim troškovima koje generiraju poslovne aktivnosti u potpori ove relacije. Alokacija ovih troškova dovodi do zaključka kako se kupci, sukladno svojim karakteristikama, dosta razlikuju u veličini pripadajućih im troškova, te da svaka novčana jedinica prihoda pojedinačnog kupca ne učestvuje jednako u profitu poduzeća.

Raznovrsnost spomenutih troškova i alociranje na pojedinačne kupce omogućeno je pojavom novih metoda obračuna troškova. Na osnovama ovih metoda, razvijen je model izračuna stvarnih pokazatelja profitabilnosti kupaca. Model identificira skup varijabli internog i eksternog okruženja bitnih za izračun prihoda, troškova i profita po pojedinom kupcu. Izlazni rezultati modela su mjere profitabilnosti pojedinačnog kupca. Empirijska vrijednost ovih pokazatelja je osnova za poduzimanje menadžerskih akcija usmjerenih na upravljanje kupcima i okruženjem na osnovu podataka iz prošlosti. Ipak, cilj menadžmenta nije samo reaktivno, već proaktivno djelovanje na bazu kupaca.

Proaktivno djelovanje zasniva se na rezultatima modela koji posjeduje i prediktivne sposobnosti. Osnova prediktivne sposobnosti modela pronađena je u primjeni tehnika i metoda multivarijacijske statističke analize i strojnog učenja. Ulazne varijable modela su i nefinancijske, pored onih čisto financijskih, a koje su po svojoj prirodi uglavnom nepogodne za obradu putem tradicionalnih statističkih metoda. Metode multivarijacijske statističke analize nadilaze neke od ovih nedostataka i mogu se djelomično upotrijebiti u analiziranju podataka o profitabilnosti kupaca. Metode strojnog učenja mogu otkriti skriveno znanje i obrasce ponašanja kupaca iz retrospektivnih analiza uzimajući u razmatranje, pored linearnih varijabli, nelinearne i nefinancijske elemente. Sposobnost da aproksimativno opišu bilo koju neprekinutu funkciju, čini ih dobrim izborom metoda za predviđanje profitabilnosti kupaca.

Cilj ovog rada jeste izgradnja modela za mjerenje tekuće i predviđanje buduće profitabilnosti pojedinačnog kupca uz upotrebu metoda strojnog učenja i multivarijacijske statističke analize. Pri tome, rad uspoređuje odabrane metode strojnog učenja i multivarijacijske analize s ciljem utvrđivanja pogodnijih metoda za namjeru predviđanja profitabilnosti kupaca.

ABSTRACT

Contemporary trends in customer profitability studies have led to the awareness that product costs are only one part of the business-customer relationship costs. This relationship is burdened with the various costs generated by business activities that support this relationship. Allocation of these costs leads to the conclusion that buyers, in accordance with their characteristics, differ greatly in the size of the associated costs and that each monetary unit of the individual customer's income does not participate equally in the company's profits.

The diversity of aforementioned costs and its allocation to individual customers is enabled by the introduction of new cost accounting methods. Based on these methods, a model of calculation of actual customer profitability indicators has been developed. The model identifies a set of variables for internal and external environment that are important for calculating revenue, costs, and profits per each customer. The outcomes of the model are profitability measures of the individual customer. The empirical value of these indicators is the basis for taking management actions focused on managing the customers based on past data. However, the management goal is not only reactive but proactive acting onto the customer database.

Proactive action is based on the results of the model that holds even predictive capabilities. The basis for predictive ability of the model was found in the application of techniques and methods of multivariate statistical analysis and machine learning. The input variables of the model are also non-financial, in addition to purely financial ones, which are by their nature mostly unsuitable for processing using traditional statistical methods. Methods of Multivariate Statistical Analysis exceed some of these deficiencies and may be partially used to analyze data on customer profitability. Machine learning methods can reveal hidden knowledge and patterns of customer behavior from retrospective analyzes by taking into consideration nonlinear and nonfinancial elements in addition to linear variables. The ability to approximately describe any uninterrupted function makes them a good choice of methods for predicting customer profitability.

The aim of this paper is to build a model for measuring current and predicting the future profitability of an individual customer by using machine learning methods and Multivariate Statistical Analysis. In that process, the paper compares selected methods of

machine learning and multivariate analysis with the purpose of identifying more suitable methods for predicting customer profitability.

Ključne riječi: strojno učenje, profitabilnost kupaca, neuronske mreže, genetički algoritam, multivarijantna statistička analiza, RFM, predviđanje.

Keywords: machine learning, customer profitability, neural network, genetic algorithm, multivariate statistical analysis, RFM, predicting.

SADRŽAJ

SADRŽAJ 1
1. UVOD 4
1.1 Motivacija za provedbu istraživanja *4*
1.2 Domena i ciljevi istraživanja *5*
1.3 Teorijske osnove i hipoteze istraživanja *7*
1.3.1 Analiza profitabilnosti kupaca 7
1.3.2 Metode strojnog učenja 9
1.3.3 Metode strojnog učenja u predviđanju profitabilnosti kupaca 11
1.3.4 Hipoteze rada 12
1.4 Struktura rada *14*
2. METODOLOGIJA STROJNOG UČENJA 17
2.1 Metode strojnog učenja *18*
2.2 Povijest razvoja neuronskih mreža *22*
2.3 Neuronska mreža *24*
2.3.1 Biološki neuron 25
2.3.2 Umjetni neuron 26
2.4 Arhitekture neuronskih mreža *29*
2.5 Učenje u neuronskim mrežama *31*
2.5.1 Učenje „s učiteljem“ 32
2.5.2 Učenje „bez učitelja“ 33
2.5.3 Algoritmi učenja 34
2.5.3.1 Učenje korekcijom pogreške 35
2.5.3.2 Učenje zasnovano na memoriranju 37
2.5.3.3 Hebbovo učenje 38
2.5.3.4 Kompetitivno učenje 39
2.5.3.5 Boltzmanovo učenje 41
2.5.4 Algoritam širenja unazad 42
2.6 Optimizacija neuronskih mreža *47*
2.6.1 Genetički algoritam 48
2.6.2 Genetički operatori 49
3. MULTIVARIJACIJSKA ANALIZA 52
3.1. Metode i tehnike multivarijacijske analize *55*
3.2. Metode regresija *59*
3.2.1 Višestruka linearna regresija 59
3.2.2 Logistička regresija 63
3.3 Ograničenja metoda multivarijacijske analize *65*
4. PROFITABILNOST KUPCA 67
4.1 Povijest analize profitabilnosti kupca *67*

4.2 Opći okvir profitabilnosti kupca *69*
4.2.1 Cjenovna politika 70
4.2.2 Troškovi 72
4.2.2.1 Raspored troškova 75
4.2.2.2 Distribucija profitabilnosti 78
4.2.3 Značajke kupca 82
4.2.4 Značajke poduzeća 83
4.2.5 Okolina 84
4.3 RFM model *86*
4.3.1 Komponente RFM modela 86
4.3.2 RFM model u analizi vrijednosti kupaca 87
4.3.2 Metode strojnog učenja i statističke analize u RFM analizi 89
4.4 Karakteristike industrije sušenog voća, orašastih proizvoda, sjemenki, mahunarki i žitarica *91*
4.4.1 Industrijski procesi 92
4.4.2 Proizvodi 93
4.4.3 Dobavljači i kupci 96
4.4.4 Opće karakteristike industrije 97
5. MODEL PREDVIĐANJA PROFITABILNOSTI KUPCA 103
5.1 Definiranje okvira za predviđanje profitabilnosti kupca *103*
5.1.1 Troškovi (L1) 105
5.1.2 Cjenovna politika (L2) 112
5.1.3 Značajke kupca (L3) 113
5.1.3.1 Veličina kupca 113
5.1.3.2 Kompleksnost kupca 114
5.1.3.3 Ponašanje kod plaćanja 115
5.2. Dimenzije profitabilnosti *116*
5.3 Normalizacija podataka *117*
5.4 Statističke metode za modeliranje profitabilnosti kupca *122*
5.4.1 Višestruka linearna regresija 122
5.4.2 Binarna logistička regresija 124
5.5 Modeliranje neuronske mreže za predviđanje profitabilnosti kupca *125*
5.5.1 Varijable modela 126
5.5.2 Arhitektura mreže 127
5.5.2.1 Topologija mreže za predviđanje R varijable 130
5.5.2.2 Topologija mreže za predviđanje F varijable 132
5.5.2.3 Topologija mreže za predviđanje M varijable 134
5.5.3 Optimizacija mreže 135
5.6 Metode usporedbe različitih metoda za predviđanje profitabilnosti kupaca *141*
6. REZULTATI EMPIRIJSKOG ISTRAŽIVANJA 144
6.1 Ciljevi empirijskog istraživanja *144*
6.2 Polazne hipoteze istraživanja *145*

6.3 Metodološki aspekti *147*
6.3.1 Definicija podataka 147
6.3.2 Instrumenti istraživanja 151
6.3.3 Preliminarni pregled podataka 153
6.4 Obrada podataka *159*
6.4.1 Regresija u predviđanju varijabli profitabilnosti 160
6.4.1.1 Predviđanje R varijable 160
6.4.1.2 Predviđanje F varijable 162
6.4.1.3 Predviđanje M varijable 164
6.4.2 Neuronska mreža u predviđanju varijabli profitabilnosti 164
6.4.3 Regresija u ocjeni poslovne politike na profitabilnost kupca 166
6.4.4 Regresija u ocjeni karakteristika kupca na profitabilnost kupca 168
6.5 Interpretacija rezultata *170*
7. ZAKLJUČAK 175
7.1 Provjera postavljenih hipoteza istraživanja *176*
7.2 Ograničenja provedenog istraživanja *179*
7.3 Preporuke za buduća istraživanja *180*
LITERATURA 181
TABLICE 189
SLIKE 191
ŽIVOTOPIS 192
POPIS OBJAVLJENIH RADOVA 194

1. UVOD

1.1 Motivacija za provedbu istraživanja

Suvremeni trendovi izučavanja profitabilnosti poslovanja evoluirali su iz promatranja profitabilnosti na razini proizvoda/usluge u postavljanje profita pojedinačnog kupca u centar pozornosti. Jedna od okolnosti koja je osigurala ovakve trendove bio je tzv. „informacijski bum“ obilježen generiranjem ogromnih količina podataka pohranjenih u bazama i skladištima podataka te intenzivnom ICT podrškom poslovnim subjektima. Ovakva podatkovna osnovica zaposlila je različite metode strojnog učenja (Ngai et al. 2009, Sonnenburg et al. 2010) uz prebacivanje fokusa analiza profitabilnosti tržišnih subjekata s razine proizvoda i usluga na pojedinačnog kupca (Ngai, 2005).

Profitabilnost kupca direktno je povezana s količinom i strukturom troškova. Promatranje troškova kupljenih proizvoda ili usluga samo je jedan dio ove strukture do kojega je relativno lako doći iz podataka prikupljenih u informacijskim sustavima poduzeća. Mnogo teže je doći do izračuna troškovnih opterećenja kupca generiranih aktivnostima poput interne logistike, administriranja narudžbi kupaca, prodaje, marketinga i post-prodajnog održavanja. Razvojem metoda koje su pomogle u alociranju ovih troškova, došlo se do zaključka kako svaka novčana jedinica prihoda pojedinačnog kupca ne sudjeluje jednako u profitu poduzeća. Pažnja se usmjerava na pojedinačnog kupca i mjerenje „vrijednosti“ novčane jedinice prihoda nastalog u poslovanju s njim.

Podrobnija empirijska analiza profitabilnosti kupaca unutar poduzeća, generirala je rezultate koje su ovo područje učinile interesantnim sa stanovišta analize i upravljanja bazom kupaca. Utvrđivanje razine troškova i disperzije profitabilnosti osnova su za utvrđivanje strategije upravljanja neprofitabilnim kupcima i uspostavljanje odgovarajuće infrastrukture za njeno provođenje. Jasan, precizan i relativno jednostavan model neophodan je za identifikaciju varijabli internog i eksternog okruženja bitnih za izračun prihoda, troškova i profita po pojedinom kupcu, ali i za definiranje mjera profitabilnosti kupaca kao izlaznih rezultata modela. Spomenute varijable modela, pored financijskih elemenata, uključuju i skup

nelinearnih i nefinancijskih elemenata koji opisuju model profitabilnosti kupca čineći ga pogodnim za primjenu metoda strojnog učenja.

Opisana retrospektivna analiza ne osigurava dovoljno pravovremenih informacija menadžmentu poduzeća, pa se u današnjem poslovanju preferira prospektivna analiza profitabilnosti kupaca, tj. pogled na buduću profitabilnost kupaca. Prospektivna analiza ekstrahira skriveno znanje i obrasce ponašanja kupaca iz retrospektivnih analiza, a služi se metodama strojnog učenja i multivarijantne statističke analize.

Ovo istraživanje nalazi se na tragu definiranih aktualnih potreba realnog sektora. Nužnost individualnog pristupa svakom kupcu, utvrđivanja disperzije profitabilnosti i generiranja neophodnih podataka za definiranje i provođenje strategije upravljanja kupcima uz nepostojanje odgovarajućih modela dovoljna su motivacija za provođenje ovog istraživanja.

1.2 Domena i ciljevi istraživanja

Globalni cilj istraživanja je izgradnja modela za mjerenje profitabilnosti kupaca u realnom vremenu na čijim osnovama će se istražiti sposobnosti metoda strojnog učenja u predviđanju buduće profitabilnosti pojedinačnog kupca. Model treba ponuditi relevantne nezavisne varijable koje će na najbolji način definirati svakog kupca i okruženje u kojem djeluje. Nezavisne varijable se u prvom redu odnose na karakteristike pojedinog kupca, ali i poslovnu politiku koju poduzeće vodi prema pojedinačnom kupcu. Pitanje elemenata profitabilnosti preuzeto je iz RFM modela (engl. *Recency, Frequency, Monetary*) profitabilnosti kupca. Na njegovoj definiciji definirane su zavisne varijable. Istraživanje će provjeriti metode logističke i višestruke linearne regresije u predviđanju RFM komponenti profitabilnosti te usporediti prediktivnu sposobnost metode neuronske mreže u odnosu na spomenute metode multivarijacijske statističke analize.

Znanstveni cilj istraživanja bi se stoga mogao definirati kroz razvoj općeg teorijskog modela, utemeljenog na postojećoj znanstvenoj literaturi, za mjerenje i predviđanje profitabilnosti kupaca koristeći podatke iz upravljačkih informacijskih sustava. Namjera

istraživanja je i da otkrije smjer i intenzitet veze između karakteristika kupca i profitabilnosti kupca te smjer i intenzitet veze između poslovne politike koju poduzeće provodi prema kupcu i profitabilnosti kupca.

Pragmatični cilj istraživanja odnosi se na stvaranje odgovarajućeg podatkovno-aplikacijskog modela za potrebe implementacije u postojeće upravljačke informacijske sustave proizvodno-distributerskih poduzeća sa svrhom predviđanja profitabilnosti postojeće osnove kupaca.

Svrha istraživanja definirana kroz davanje doprinosa u pronalasku odgovora na znanstveno pitanje kako mjeriti i kojim metodama predviđati profitabilnost pojedinog kupca, nametnula je i niz parcijalnih pitanja koja traže odgovore:

- Koji je opći okvir za definiranje profitabilnosti kupca?
- Kako nadoknaditi nedostatak povijesnih financijskih i transakcijskih podataka za procjenu ukupne vrijednosti kupca?
- Jesu li metode strojnog učenja i multivarijacijske statističke analize primjenjive u predviđanju profitabilnosti kupca?
- Koje metode koristiti u predviđanju profitabilnosti kupca?
- Koje metode su bolje sa stanovišta primjene i točnosti predviđanih mjera profitabilnosti?
- Koje mjere koristiti u izražavanju profitabilnosti kupaca?
- Jesu li odabrane mjere profitabilnosti stvaraju dobru osnovu za donošenje odluka menadžmenta u svrhu povećanja profitabilnosti kupaca?
- Kako poslovna politika prema pojedinom kupcu utječe na njegovu profitabilnost?
- Kako karakteristike pojedinog kupca utječu na njegovu profitabilnost?

1.3 Teorijske osnove i hipoteze istraživanja

1.3.1 Analiza profitabilnosti kupaca

Trenutni fokus analize profitabilnosti kupaca (engl. *Customer Profitability Analysis - CPA*) evoluirao je do izučavanja poslovanja s mjerama uspjeha izraženim na razini pojedinog kupca. 1950-ih godina se kupac promatrao kao svrha poslovanja (Drucker, 1954). 1980-ih zadovoljstvo kupca se stavlja u prvi plan, a 1990-ih fokusira se izgradnja odnosa s kupcem i njegovo održavanje (Storbacka, 1997; Peck, 1997). Ovakav položaj kupca zahtijevao je prikupljanje povećanog skupa informacija o svakom pojedinačnom kupcu. Ti podaci postaju esencijalni u smislu sposobnosti davanja odgovora na uvelike rastuće potrebe kupaca. Kao primarna informacija iz ovog skupa izdvaja se informacija o profitabilnosti pojedinačnog kupca (Cooper i Kaplan, 1991; Storbacka, 1997; Niraj et al. 2001) . Što je to što ovu informaciju izdvaja od ostalih? Na navedeno pitanje Gupta (2003) navodi odgovor kroz činjenicu kako svaka novčana jedinica prihoda pojedinog kupca ne učestvuje jednako u dobiti poduzeća.

Razlike u vrijednosti novčane jedinice prihoda pojedinog kupca nastaju zbog razlika u pripadajućim prihodima ili troškovima. Mjere prihoda ostvarenih od pojedinog kupca su uglavnom jasno definirane i prepoznate od strane poslovne organizacije. Mnogo je teže doći do informacija koliki su stvarni troškovi podrške procesima u kojima se ostvaruje taj prihod.

Troškovi proizvoda čine samo jedan dio troškova poslovnog odnosa na relaciji poduzeće-kupac (Kaplan i Anderson, 2007). Oni su inače bili u centru pozornosti izučavanja profitabilnosti na razini proizvoda. Suvremeno poslovanje, praćeno napretkom tehnologije, povećalo je diversifikaciju i kompleksnost proizvoda, usluga i kupaca. Indirektni troškovi preuzeli su primat u ukupnoj masi troškova u odnosu na direktne troškove (Cokins, 2001; Kaplan i Anderson, 2007). Ova raznovrsnost troškova, postavila je problem pred klasične računovodstvene metode rasporeda troškova. Istovremeno je i otvorila prostor razvoju alternativnih metoda obračuna troškova. Cooper, Kaplan i Anderson (Cooper i Kaplan,1991; Kaplan i Anderson, 2007) uvode koncepte ABC (engl. *Activity Based Costing*) i TDABC (engl. *Time Driven Activity Based Costing*) troškovnog upravljanja. Ovim konceptima našlo se rješenje kojim su poduzeća prevladala nedostatke tradicionalnih računovodstvenih sustava.

Nova saznanja otvorila su prostor za intenzivna istraživanja na polju retrospektivnih metoda izračuna profitabilnosti kupca poput Niraj modela (Niraj et al., 2001), ali i otvorila put prospektivnim modelima za predviđanje profitabilnosti (Campbell i Frei, 2004; Lee i Park, 2005; Malthouse i Blattberg, 2005; Ayres, 2007; Davenport i Harris, 2007; Donkers et al., 2007; Blattberg et al., 2008).

Retrospektivna koncepcija promatra kupca iz povijesne perspektive. Analizira apsolutnu i relativnu vrijednost profitabilnosti pojedinačnog kupca ili grupe kupaca u nekom proteklom vremenskom periodu. Empirijska istraživanja sljedbenika ove koncepcije dovela su do jako zanimljivih saznanja. Pokazala su da u bazi kupaca postoji određeni dio profitabilnih i određeni dio neprofitabilnih kupaca. Ovo saznanje samo po sebi nije otkriće. Otkriće se ogleda u stupnju diferencijacije ove dvije grupe kupaca i činjenici da taj stupanj zavisi od više činitelja. Zanimljiva vizualizacija spomenute diferencijacije predstavljena je Kanthal krivuljom (engl. *Kanthal curve*). Formirao ju je Kaplan (1989-1, 1989-2) izučavajući profitabilnost kupaca uporabom ABC metode obračuna troškova na kompaniji Kanthal. Ona grafički prezentira distribuciju kupaca i definira tzv. „Kanthal efekt" koji izražava svu ozbiljnost potrebe izračuna profitabilnosti pojedinačnog kupca. Kanthal efekt negira uvriježeno Paretovo načelo (20:80) po kome 20% kupaca stvara 80% profita. Kaplan i Cooper (1998) tako prikazuju primjer u kome 40% kupaca stvara 250% profita, 10% generira gubitak od gotovo 150% vrijednosti profita, a ostatak baze kupaca se nalazi oko točke pokrića. Storbacka (1997) prezentira istraživanje s ekstremnim slučajem Kanthal efekta u kome je više od pola baze kupaca neprofitabilno. Ova i slična istraživanja (van Raaij et al., 2003; Wang i Hong, 2006; Kumar et al., 2006) definirala su važnost promatranja pojedinačnog kupca i ukazala na mogućnosti CPA analize sa stanovišta upravljanja troškovima, prihodima i ciljanim marketinškim aktivnostima.

Prospektivna koncepcija promatra kupca kao „nematerijalnu imovinu" poduzeća i sukladno tome promatra mu „buduću vrijednost" u formi vjerojatnog profita koji će se ostvariti u poslovanju s njim u nekom definiranom budućem periodu (Gupta, 2003). Ona definira i pojam CLV-a (engl. *Customer Lifetime Value – CLV*) kao trenutnu vrijednost očekivanih zarada u predstojećem poslovanju s kupcem. Sukladno tome, izračunati pokazatelji CLV-a čine osnovu za strateško marketinško i prodajno djelovanje menadžmenta poduzeća nad bazom svojih kupaca. Jedna od brojnih klasifikacija u marketinškoj literaturi, Gupta et al. (2006), definira sljedeće pristupe modeliranju CLV-a:

- RFM modeli;
- Modeli vjerojatnosti (engl. *Probability Models*) bazirani na Pareto/NBD modelu i Markovljevom lancu;
- Ekonometrijski modeli, također bazirani na NBD (engl. *Negative Binomial Distibution*) modelu;
- Modeli dosljednosti (engl. Persistence Models).

Gupta s koautorima (2006) ističe RFM model kao najjednostavniji. Klasificira ga također pri tome i kao najmoćniji model za prognoziranje profitabilnosti kupaca. RFM metoda klasificira kupce na osnovi Recency/Frequency/Monetary pokazatelja iz proteklog poslovanja s kupcem. Pri tome se R (engl. *Recency*) pokazatelj odnosi na dužinu vremenskog perioda proteklog od zadnje kupnje, F (engl. *Frequency*) se odnosi na broj kupovina u definiranom vremenskom periodu analiziranja kupovnih navika kupca, te M (engl. *Monetary*) na novčanu vrijednost kupnje promatrane u spomenutom vremenskom periodu. RFM analitički model koji upotrebljava tri jednako važne dimenzije za segmentiranje baze kupaca, originalno je predstavio Hughes 1994. godine (Hughes, 1994). Promatrajući važnost pojedinih varijabli, nastaju prošireni ponderirani RFM (engl. *Weighted RFM*) modeli koji od strane različitih autora pojedinim dimenzijama daju različite težinske faktore, odnosno različit redoslijed važnosti dimenzija. Stone (1995) i Chuang i Shen (2008) smatraju kako se pojedine dimenzije ne mogu jednako tretirati u različitim industrijama te se kao pitanje postavlja raspored važnosti R-F-M komponenti modela. Novija istraživanja, sasvim obrnuto navedenom, postavljaju M kompontentu na prvo mjesto, a R komponentu na zadnje mjesto (Chuang i Shen, 2008). Nastaju i neki novi modeli koji tri navedene varijable dopunjuju dodatnim, poput RFMTC (engl. *Recency, Frequency, Monetary, Time since first purchase, and Churn probability*) modela (Cheng i Chen, 2009) koji uključuje vrijeme od prve kupovine i vjerojatnost zadržavanja kupca u bazi kupaca poduzeća.

1.3.2 Metode strojnog učenja

Metode strojnog učenja iskazale su se kao logičan izbor pri odabiru metoda s kojima će se predviđati profitabilnost kupaca. Što je to strojno učenje? *Polje strojnog učenja se bavi pitanjem kako konstruirati računalni program koji će se automatski unaprjeđivati s iskustvom*

(Mitchell, 1997). Navedenu definiciju Mitchell proširuje definirajući strojno učenje s aspekta računalnih znanosti. Hastie et al. (2008) promatraju ove metode sa statističke točke gledišta stavljajući podatke u centar pozornosti i tretirajući strojno učenje *„učenjem iz podataka"*. U osnovi se javljaju dva pristupa. Prvi, statistički, koji stavlja podatke i ekstrakciju znanja iz njih u centar pozornosti, te drugi, računalni, koji definira metode računalnim arhitekturama i algoritmima za optimalno prikupljanje, pohranjivanje i spajanje podataka u svrhu samoprogramiranja. Bishop (2008) promatra strojno učenje s inženjerske strane izučavajući prepoznavanje uzoraka (engl. *pattern recognition*) dodajući računalnoj komponenti i inženjersku definirajući „*dva lica istog polja*" kod metoda strojnog učenja. Ipak, strojno učenje je multidisciplinarna kombinacija (Marsland, 2009) u kojoj svaka od akademskih disciplina, poput statistike, matematike, računalnih znanosti i inženjerstva, igra pojedinačno značajnu ulogu.

Najveća razina generalizacije poznaje četiri tipa strojnog učenja: nadzirano (engl. *Supervised Learning*), nenadzirano (engl. *Unsupervised Learning*), pojačavajuće (engl. *Reinforcement Learning*) i evolucijsko (engl. *Evolutionary Learning*). Osnovnu komponentu metoda čine algoritmi učenja, a Brownlee (2014) ih klasificira na sljedeći način:

- Regresijske metode (engl. *Regression*)
- Instanca-bazirane metode (engl. *Instance-based Methods*)
- Regularizacijske metode (engl. *Regularization Methods*)
- Stabla odlučivanja (engl. *Decision Tree Learning*)
- Bayesian metode (engl. *Bayesian Methods*)
- Kernel metode (engl. *Kernel Methods*)
- Klaster metode (engl. *Clustering Methods*)
- Asocijativna pravila (engl. *Association Rule Learning*)
- Neuronske mreže (engl. *Artificial Neural Networks*)
- Metode redukcije dimenzija (engl. *Dimensionality Reduction*)
- Orkestrirane metode (engl. *Ensemble Methods*)

Metodu neuronske mreže (engl. *Neural Network*) Haykin (1999) definira kao *masivno paralelno distribuirani procesor sačinjen od jednostavnih procesnih jedinica (umjetnih neurona), koji ima prirodnu sposobnost pohranjivanja iskustvenog znanja*

raspoloživog za daljnju uporabu. Ova metoda je značajna sa stanovišta predmetnog istraživanja.

Neuronske mreže su svoju primjenu našle u mnogobrojnim aplikacijama problemskog tipa klasifikacije, prepoznavanja uzoraka, predviđanja, optimizacije i klasteriranja. Prediktivna sposobnost je područje na kojem su se tradicionalno koristile statističke metode (Razi i Athappilly, 2005, Kumar et al. 1995, Lee i Young, 2000, Zang et al., 2004,). Paliwal i Kumar (2009) tvrde kako je fokus primjene neuronskih mreža na problemima predviđanja i klasifikacije u različitim poljima te ih stavljaju u odnos sa statističkim tehnikama. Promatraju ih kao konkurentne metode upravo na polju predviđanja te je na ovim usporedbama postavljen i jedan od akcenata u ovom istraživanju.

1.3.3 Metode strojnog učenja u predviđanju profitabilnosti kupaca

U literaturi se susreću dva tipa modela za predviđanje profitabilnosti kupaca. Prvi model svoja predviđanja temelji na podacima dobivenim istraživanjima ponašanja kupca u poslovnom odnosu s konkurencijom i poduzećem koja je u centru istraživanja (Rust et al. 2004). Drugi model promatra samo poslovanje kupca s poduzećem koje predviđa njegovu vrijednost zanemarujući pri tome kupčevo poslovanje s konkurencijom (Niraj et al. 2001, Fader et al. 2005). U dosadašnjim istraživanjima predviđanja profitabilnosti po drugom modelu, a koji će koristiti i ovo istraživanje, ipak nisu postignuti izvrsni rezultati uporabom metoda strojnog učenja za razliku od primjena opisanih u prethodnom poglavlju.

Campbell i Frei (2004) koristili su tekuću profitabilnost za predviđanje buduće, te su istraživanjem utvrdili kako predviđanje pogrešno klasificira mnoge kupce. Malthouse i Blattberg (2005) u svom istraživanju koriste regresijski model i neuronske mreže. Svoje rezultate prezentirali su kovanicama 20-55 i 80-55. Dakle, od 20% visokoprofitabilnih kupaca pogrešno su klasificirali 55% njih u projekciji buduće profitabilnosti i osigurali im pogrešan marketinški tretman, dok su od 80% niskoprofitabilnih kupaca pogrešno klasificirali njih 15%. Donkers et al. (2007) promatraju CLV vrijednosti kupaca u industriji osiguranja. Grade nekoliko modela i ustvrđuju da jednostavni modeli daju najbolje rezultate (naivni model preuzimanja profita iz zadnjeg perioda promatranja i njegova neograničena projekcija na buduće periode za pojedinačnog kupca) te predviđaju bolje od regresijskih modela i

modela kojima su eksplicitno modelirali stopu zadržavanja kupaca i vjerojatnost ponovljene kupovine.

Postojeća istraživanja predviđanja profitabilnosti kupaca uz uporabu metoda strojnog učenja i na osnovi povijesnih podataka koja poduzeća posjeduju u svojim bazama podataka, nisu pokazala potpunu pouzdanost i veliki uspjeh. Neke jednostavnije metode klasične statističke analize uz korištenje smanjenog broja nezavisnih varijabli, pokazala su se jednako dobrim, ponekad i boljim od ovih složenijih. Na ovom tragu, Rust et al. (2011), razvijaju model čija je osnova BG/NBD (engl. *Beta-Geometric/Negative Binomial Distribution*) model razvijen od Fadera (Fader et al. 2005) i koji nadilazi jednostavne upravljačke heurističke modele predviđanja buduće profitabilnosti kupca. Sukladno navedenom, postoji mnogo prostora za istraživanje u cilju pronalaženja primjenjivog modela za izračun tekuće i predviđanje buduće profitabilnosti pojedinačnog kupca iz postojećih podataka, a što je i predmet ovog rada.

1.3.4 Hipoteze rada

Ovim istraživanjem nastoji se stvoriti model predviđanja profitabilnosti kupaca, utvrditi prednost specifične metode strojnog učenja, neuronske mreže, u odnosu na metode multivarijacijske statističke analize te dokazati utjecaj provođenja marketinških politika i karakteristika kupca na mjere profitabilnosti. Sukladno navedenom definirane su tri glavne hipoteze istraživanja i pet pomoćnih koje će se nastojati potvrditi istraživanjem.

H1: Metoda neuronske mreže može predvidjeti profitabilnost kupaca jednako dobro ili bolje od statističkih metoda logističke regresije i višestruke linearne regresije.

Neuronska mreža će se koristiti za predviđanje sve tri izlazne varijable modela, dok će se logit metoda koristiti za predviđanje **R** i **F** varijabli, a MLR metoda za predviđanje **M** varijable. Ova hipoteza stavlja u odnos odabrane metode strojnog učenja i multivarijacijske analize i nastoji dokazati dominaciju metode strojnog učenja u području predviđanja profitabilnosti kupaca.

H2: Politika prema kupcima utječe na profitabilnost kupaca

Istraživačka pretpostavka ove hipoteze u odnos stavlja politiku koju poduzeće provodi prema pojedinom kupcu i tvrdi da ona direktno utječe na veličinu mjera profitabilnosti kupca. Kako prema definiranom modelu istraživanja politiku prema kupcu čine cjenovna politika i veličina troškova nastalih poslovnim odnosom s kupcem, formirane su dvije pomoćne hipoteze koje će potpomognuti dokazivanje glavne hipoteze.

H2.1: Troškovna komponenta politike prema kupcu utječe na profitabilnost kupca

H2.2: Cjenovna politika prema kupcu utječe na profitabilnost kupca

H3: Karakteristike kupca utječu na profitabilnost kupaca

Istraživačka pretpostavka treće hipoteze stavlja u odnos heterogenost skupa kupaca te tvrdi da ona direktno utječe na veličinu mjera profitabilnosti kupca. Definiranim modelom istraživanja karakteristike kupca su grupirane u kategorije veličine, kompleksnosti i ponašanja pri plaćanju obveza. Veličinu kupca čine financijski elementi poslovnog odnosa iskazani veličinom prometa i ostvarene marže u poslovnom odnosu s kupcem. Osim ovih čisto financijskih elemenata, ovu kategoriju karakterizira i pripadnost grupi kupaca sukladno veličini (malo, srednje ili veliko poduzeće) i teritorijalnoj pripadnosti (domaće ili inozemno poduzeće). Kompleksnost kupca definira stupanj složenosti poslovnog odnosa s kupcem. Modelirana je kroz varijable koje kvantificiraju količinu narudžbi od kupca, različitost asortimana isporučenih proizvoda, broj mjesta isporuke te količinu i vrijednost transakcija povrata robe. Ponašanje pri plaćanju definira način ponašanja kupca u finalnom koraku poslovnog odnosa na relaciji promatranog poduzeća i pojedinačnog kupca. Varijable ove kategorije određuju iznose potraživanja naplaćene na vrijeme, vremensku strukturu naplate potraživanja van ugovorenog vremena plaćanja, nenaplaćena potraživanja i prosječno vrijeme kašnjenja u plaćanju. Sukladno navedenom, formirane su tri pomoćne hipoteze koje će potpomognuti dokazivanje navedene glavne hipoteze.

H3.1: Veličina kupca utječe na profitabilnost kupca

H3.2: Kompleksnost kupca utječe na profitabilnost kupca

H3.3: Ponašanje pri plaćanju kupca utječe na profitabilnost kupca

1.4 Struktura rada

Doktorska disertacija se sastoji od šest osnovnih poglavlja.

Uvodni dio predstavlja znanstveni problem kojim se bavi cjelokupno istraživanje. Posebno je predstavljen motivacijski moment koji je potaknuo autora na istraživanje. Poglavljem je definirana domena istraživanja te predstavljeni znanstveni i pragmatični ciljevi istog. Potom su dane teorijske osnove znanstvenog problema i definirane hipoteze istraživanja. Poglavlje je zaključeno prikazom strukture disertacije.

Metodologija strojnog učenja naslov je drugog poglavlja. Ono započinje teorijskom podlogom metoda, dajući poseban osvrt na podjele metoda sukladno procesu i ciljevima učenja. Izložene su potom osnove algoritamskih struktura, sukladno njihovoj funkciji i formi, po kojima se u osnovi i razlikuju pojedine metode strojnog učenja. Kako je za empirijsko istraživanje korištena metoda neuronske mreže, nastavak poglavlja daje povijesnu dimenziju njenog razvoja. Opisan je biološki neuron i na njemu nastala osnovna komponenta neuronskih mreža – umjetni neuron, uz definiranu matematičku komponentu njegove funkcionalnosti. Način povezivanja i međusobna interakcija osnovnih gradivnih komponenti mreže dana je u dijelu poglavlja koji opisuje arhitekture neuronskih mreža. Sastavni dio procesa djelovanja mreže je učenje, u kojem se slobodni parametri neuronske mreže adaptiraju kroz proces stimuliranja od strane okoline čiji je mreža dio. Iz ovog razloga, slijede teorijske osnove različitih metoda učenja i njihova matematička interpretacija. Posebna pozornost se posvećuje algoritmu „širenja unazad“ koji je najviše korišten algoritam praktične primjene neuronskih mreža. Na kraju poglavlja postavlja se pitanje optimiziranja modela generiranih metodom neuronske mreže i pokušaja njihovog optimiziranja. Tražeći prigodnu metodu za ovaj problem, rad se dotiče metode koja svoju analogiju, baš kao i metoda neuronske mreže, zasniva na prirodnim, biološkim procesima: genetički algoritam.

Treće poglavlje bavi se pregledom metoda **multivarijacijske analize**. Jedan od ciljeva istraživanja je poređenje odgovarajućih metoda ove statističke discipline s prethodno spomenutom metodom neuronske mreže. Predstavljeni su i opisani ciljevi znanstvenih istraživanja upotrebom metoda multivarijacijske analize. Izvršena je podjela metoda sukladno broju i podjeli varijabli te sklali njihovog mjerenja. Definiran je skup od deset osnovnih

tehnika multivarijacijske analize na kojima se zasniva djelovanje svih metoda ove analize i za svaku tehniku je dana teorijska osnovica primjene. Za potrebe empirijskog istraživanja disertacije korištene su dvije metode multivarijantne statističke analize: višestruka linearna regresija i logistička regresija. Obje metode su regresijske metode, a u nastavku rada su prikazane njihove osnovne značajke. Na kraju su prikazane karakteristike kompleksnosti i ograničenja opisivanih metoda.

U centru istraživanja je ekonomska kategorija **profitabilnosti kupca** kojoj je posvećeno četvrto poglavlje. Povijest analize profitabilnosti daje vremensku komponentu razvoja znanstvene misli ovog područja od početaka kada je proizvod bio u središtu pozornosti, do vremena kad je pojedinačni kupac osnovni predmet poslovnih analiza. Razmatranja definiraju nužnost proaktivnog djelovanja na bazu kupaca i definiraju opći okvir za promatranje tekuće i predviđanje buduće profitabilnosti kupaca. Posebno se razmatraju činitelji profitabilnosti iskazani kroz troškove, cjenovnu politiku, značajke poduzeća, značajke kupca i osobine poslovnog okruženja. Značajan dio ovih razmatranja posvećen je rasporedu troškova i na njima zasnovanoj distribuciji profitabilnosti. Prospektivna koncepcija analize profitabilnosti promatra kupca kao „nematerijalnu imovinu“ poduzeća i sukladno tome promatra mu „buduću vrijednost“. Za potrebe modeliranja pokazatelja buduće profitabilnosti, nakon analize postojećih modela, odabran je i u poglavlju opisan RFM model. Ukazalo se na njegove komponente, te dao povijesni prikaz korištenja modela u analizi vrijednosti kupca. Slijedi povijesni prikaz upotrebe metoda strojnog učenja i multivarijacijske statističke analize u RFM analizi. Na kraju poglavlja dane su karakteristike industrije iz koje su preuzeti podaci za empirijsko istraživanje.

U petom poglavlju definiran je **model predviđanja profitabilnosti kupaca**. Definiran je opći okvir za predviđanje profitabilnosti s četiri kategorije ulaznih, nezavisnih varijabli: troškovi, cjenovna politika, značajke kupca i ponašanje kod plaćanja. Definirane su nakon toga dimenzije profitabilnosti prema odabranom RFM modelu. Prikazani su algoritmi normalizacije podataka i njihove pripreme za primjenu metoda višestruke linearne regresije, binarne logističke regresije i neuronske mreže. U nastavku su prezentirane postavke regresijskih modela i uspostavljene veze projektiranih modela i pojedinačnih glavnih i pomoćnih hipoteza istraživanja. Predstavljene su potom arhitekture neuronskih mreža koje su korištene u procesu učenja i parametri koji su poslužili usporedbi postignutih rezultata pojedinih arhitektura. Dan je prikaz „pobjedničkih“ topologija u predviđanju svake od tri izlazne varijable modela, te su predstavljene i metode optimizacije mreže uz poseban osvrt na

metodu genetičkog algoritma. Kako model treba ponuditi usporedbu metoda strojnog učenja i metoda multivarijacijske analize, na kraju poglavlja su definirane pokazatelji usporedbe korišteni u istraživanju.

Šesto poglavlje prikazuje **rezultate empirijskog istraživanja**. Početak poglavlja definira ciljeve istraživanja i polazne hipoteze istraživanja. Metodološki aspekti prikazuju definiciju podataka, odnosno pojedinačne varijable s osnovnim deskriptivnim statističkim pokazateljima. Detaljno su definirane zavisne varijable modela i pojedinačne metode, kao instrumenti istraživanja, po pojedinim hipotezama istraživanja. Preliminarni pregled podataka istraživanja dan je kroz deskriptivnu analizu podataka o kupcima, proizvodima i prometu promatranog poduzeća čiji su podaci korišteni za istraživanje. Deskriptivna statistika profitabilnosti kupaca prikazuje profil ukupne profitabilnosti promatranog poduzeća po godinama. Za potrebe obrade podataka sve nekategorijalne vrijednosti podataka su normalizirane i dana je njihova deskriptivna statistika. Na kraju su prikazani rezultati regresijskih modela i modela neuronske mreže u predviđanju R, F i M varijabli modela, te regresijskih modela u određivanju utjecaja poslovne politike, karakteristika kupaca i ponašanja pri plaćanju obveza na profitabilnost kupaca. Sukladno rezultatima, na kraju poglavlja je dana interpretacija rezultata i ocjena pojedinačnih hipoteza istraživanja.

Z**aključak** je naslov zadnjeg poglavlja. U njemu su sintetizirani rezultati istraživanja. Izvršena je evaluacija teorijskih i empirijskih spoznaja nastalih provedenim istraživanjem i pisanjem disertacije. Predstavljeni su znanstveni i praktični doprinosi istraživanja, uz navođenje ograničenja istraživanja i predlaganje smjernica za naredna istraživanja ovog područja.

2. METODOLOGIJA STROJNOG UČENJA

Polje strojnog učenja se bavi pitanjem kako konstruirati računalni program koji će se automatski unaprjeđivati s iskustvom (Mitchell, 1997) . Ovom definicijom Mitchell otvara predgovor svog djela „*Machine Learning*“, dajući definiciju predmetnog područja s programerske strane i postavljajući pred struku izazov pisanja programa koji će sami sebe unaprjeđivati. U jednom od svojih narednih radova Mitchell (Mitchell, 2006) proširuje navedenu definiciju definirajući centralno pitanje znanstvenog polja koje se bavi strojnim učenjem: „*Kako se može izgraditi računalni sustav koji se unapređuje s iskustvom i koji su fundamentalni zakoni koji upravljaju svim procesima učenja?*“. Mitchell definira strojno učenje s aspekta računalnih znanosti.

Statističari isti problem promatraju sa svog stanovišta. Hastie et al. (2008) tako impliciraju da „*se velike količine podataka generiraju na mnogim poljima, a zadatak statističara je da izvuku smisao iz njih: da ekstrahiraju značajne predloške i trendove i da razumiju „što podatci govore“. Ovo se zove učenje iz podataka“*. Statističari fokus svojih napora stavljaju na izvlačenje zaključaka iz raspoloživih povijesnih podataka. Nasuprot ovome, strojno učenje pokušava definirati računalne arhitekture i algoritme za optimalno prikupljanje, pohranjivanje i spajanje podataka u svrhu samoprogramiranja na osnovi određenih inicijalnih struktura i iskustva iz tretiranih podataka.

Bishop (2008) promatra strojno učenje s inženjerske strane izučavajući prepoznavanje uzoraka (engl. *pattern recognition*). On tako konstatira da „*prepoznavanje uzoraka originalno potječe iz inženjerstva, dok strojno učenje nastaje iz računalnih znanosti*“, ali i da se ove aktivnosti mogu vidjeti kao „*dva lica istog polja*“.

Marsland (2009) naglašava multidisciplinarnu stranu strojnog učenja navodeći kako „*počiva na osnovama nekoliko akademskih disciplina“* (računalne znanosti, statistike, matematike i inženjerstva), ali da je „*najčešće izučavana u području umjetne inteligencije što je primarno svrstava u polje računalnih znanosti“*. Ipak, razumijevanje algoritama strojnog učenja zahtijeva detaljnije poznavanje statističkih i matematičkih metoda na kojima su isti zasnovani.

2.1 Metode strojnog učenja

Na najvećoj razini generaliziranja dva su tipa strojnog učenja: nadzirano i nenadzirano (Turban et al., 2005). Nadzirano učenje (engl. *supervised learning*) je proces učenja nad setom podataka za učenje pri čemu su poznati točni odgovori (izlazni podaci sustava). Produkt ovakvog procesa učenja je sustav koji generalizira odgovore na sve moguće ulazne podatke. Kod nenadziranog učenja (engl. u*nsupervised learning*) u setu podataka za učenje nisu poznati točni odgovori. Umjesto njih, algoritam nastoji identificirati sličnosti među ulaznim podacima tako da se ulazi koji imaju nešto zajedničko svrstaju u zajedničke kategorije.

Marsland (2009) pored ove dva tipa strojnog učenja definira još dva: pojačavajuće (engl. *reinforcement learning*) i evolucijsko (engl. *evolutionary learning*). Pojačavajuće učenje je između nadziranog i nenadziranog, a evolucijsko je zasnovano na procesima adaptiranja bioloških organizama na uvjete okoline kako bi povećali sposobnost opstanka (Guyon, 2007).

Izučavanje strojnog učenja predstavlja u osnovi izučavanje algoritama strojnog učenja (Langley, 1992). Brownlee (2014) prema funkciji i formi definira sljedeće metode odnosno grupe algoritama:

- Regresijske metode (engl. *Regression*)
- Instanca-bazirane metode (engl. *Instance-based Methods*)
- Regularizacijske metode (engl. *Regularization Methods*)
- Stabla odlučivanja (engl. *Decision Tree Learning*)
- Bayesian metode (engl. *Bayesian Methods*)
- Kernel metode (engl. *Kernel Methods*)
- Klaster metode (engl. *Clustering Methods*)
- Asocijativna pravila (engl. *Association Rule Learning*)
- Neuronske mreže (engl. *Artificial Neural Networks*)
- Metode redukcije dimenzija (engl. *Dimensionality Reduction*)
- Orkestrirane metode (engl. *Ensemble Methods*).

Regresijske metode modeliraju relacije među varijablama modela. One kvantificiraju kako su ciljne (zavisne) varijable u relaciji s objašnjavajućim (nezavisnim, prediktorskim) varijablama (Abraham i Ledolter, 2005).

Instanca-bazirane metode ne generaliziraju pravilo predviđanja na osnovi slučajeva iz povijesti. One, nasuprot ovome, prikupljaju bazu slučajeva i pripadajućih izlaznih rezultata. Nova instanca događaja čiji izlazni rezultat se predviđa, nastoji se povezati s odgovarajućim slučajem iz povijesne baze kako bi se potom odredila ciljna vrijednost upotrebom jednog od dva algoritma: k-najbližih susjeda (engl. *k-Nearest Neighbor learning*) ili lokalne težinske regresije (engl. *Locally Weighted Regression*). Ove metode često se nazivaju i „lijene" (engl. *lazy*) metode iz razloga što ne tvore generalno pravilo predviđanja u procesu prikupljanja i analize povijesnih podataka, već tek po pojavi nove instance podataka.

Regularizacijske metode operiraju nad modelima koji pretpostavljaju visok stupanj kompleksnosti i male količine podataka, a nastoje riješiti problem pretreniranosti (engl. *overfitting*) koji je karakterističan za skupove podataka s ovim osobinama. „*Pretreniranost je tendencija metode rudarenja podataka da skroji model za podatke iz skupa za učenje, na štetu generalizacije do tada neviđenih podataka"* (Provost i Fawcett, 2013). Pretreniranost prepoznaje šum u podacima, umjesto generiranja relacije sposobne za generalizaciju modela. Kako regularizacijske metode rješavaju ovaj problem? Pod pretpostavkom da je s **L(X,Y)** označena funkcija greške gdje je **X** vektor željenih izlaznih vrijednosti modela, a **Y** vektor dobivenih izlaznih vrijednosti projektiranog modela, a s **w** označen vektor težinskih faktora modela, vrši se korekcija funkcije gubitka multipliciranom $\mathbf{L_1}$ (Lasso) ili $\mathbf{L_2}$ (Ridge) normom vektora težinskih faktora:

$$L(X,Y)+\lambda N(w) \qquad (2.1)$$

U navedenoj jednadžbi **N** predstavlja L1, L2 ili bilo koju drugu normu. Faktor multiplikacije, označen s **λ**, predstavlja regularizacijski izraz do kojeg se dolazi empirijskim putem kroz proces učenja modela. Konačni rezultat procesa regularizacije je model s odgovarajućom vrijednosti faktora **λ** pri kojem je minimizirana funkcija greške.

Stabla odlučivanja čine grupu metoda strojnog učenja koja kao rezultat daju klasifikacijski model oblika stabla. Interni čvorovi stabla predstavljaju testove atributa promatranog modela, dok grane stabla predstavljaju rezultate pripadajućih testova. Krajnji elementi stabla, listovi, čine izlaz modela. Top čvor stabla naziva se korijenski čvor. Proces

učenja generira, na osnovu skupa podataka za učenje, niz putova koji vode od korijenskog čvora prema klasifikacijskom listu kao produktu modela. Slično kao kod pretreniranosti (engl. *overfitting*) modela, postoji opasnost da se u procesu učenja "skrije" šum testnih podataka u generirani model te da model izgubi generalizaciju. Metode "podrezivanja stabla" (engl. *tree pruning*) nastoje eliminirati grane stabla koje sadrže spomenuti šum s ciljem unapređenja klasifikacijske sposobnosti modela na neviđenim serijama podataka (Han i Kamber, 2001).

Bayesianske metode su statistički klasifikatori, a svoju osnovu nalaze u Bayesovom teoremu. Pripadaju grupi klasifikacijskih metoda, a mnoge komparativne analize utvrdile su pozitivne komparativne sposobnosti naivne Bayesianske metode klasifikacije (engl. *Naive Bayesian Classification*) u poređenju s klasifikacijskim algoritmima stabla odlučivanja i neuronske mreže. Bitna osobina ovih metoda je brzina rada i točnost u primjeni na velikim skupovima podataka. Spomenuti naivni Bayesianski klasifikator pretpostavlja nezavisnost pojedinačnih atributa u utjecaju na pripadnost klasi pojedine instance podataka. Ova pretpostavka pojednostavljuje procese primjene metode i u tom smislu je deklarira „naivnom" metodom. Za razliku od ove, Bayesianske mreže povjerenja (engl. *Bayesian belief networks*) pretpostavljaju vezu među skupovima atributa i grafičke su metode klasifikacije.

Kernel metode pripadaju grupi metoda koji svoju svrhu nalaze u problematici prepoznavanja uzoraka (engl. *pattern recognition*). Ime su dobile po kernel funkciji koja osigurava njihovu primjenu u višedimenzionalnom implicitnom prostoru svojstava promatranog problema računajući i poredeći unutarnji produkt između parova podataka promatranog prostora. Ovakav pristup zahtijeva manju računalnu snagu u procesu učenja od eksplicitnog računanja koordinata. Zbog ovakvog pristupa ove metode mogu se promatrati kao instanca bazirane metode, a neki od algoritama koji primjenjuju ovu metodu su kernel perceptron, SVM (engl. *Support Vector Machines*), PCA (engl. *Prncipal Component Analysis*) i kanonička korelacijska analiza (engl. *Canonical correlation analysis*).

Klaster metode karakterizira analiza pojedinačnih uzoraka u nastojanju da se otkrije međuovisnost pojedinih članova (Chiu i Tavela, 2008). Pri tome se pod članovima podrazumijevaju objekti promatrane studije, a uzorci predstavljaju kolekcije članova korištene za provođenje analize. Analiza podrazumijeva „mjerenje sličnosti" pojedinih članova uzorka računajući udaljenost među njima, pri čemu se koristi neka od metoda poput npr. Euklidske udaljenosti. Tehnike klasteriranja uključuju hijerarhijske i particijske metode.

Najpoznatija particijska metoda je K-Means klaster metoda, a od hijerarhijskih poznate su metode najbližih susjeda (engl. *Nearest neighbor*) i najdaljih susjeda (engl. *Furthest neighbor*), Wardova greška sume kvadrata (engl. *Ward's error of sum of squares*), CART (engl. *Classification and Regression Tree*), CHAID (engl. *Chi-Square Automatic Interaction Detection*).

Metode *asocijativnih pravila* čine grupu dobro istraženih metoda za otkrivanje relacija među varijablama u velikim setovima podataka. Namjera im je identificirati čvrsta pravila među podacima na osnovi korištenja različitih mjera značajnosti i zanimljivosti (Frawley et al, 1992). Najpoznatija primjena ove metode svakako je tzv. košarica proizvoda, odnosno definiranje pravila koje indicira da osoba koja kupuje proizvod A ima tendenciju da kupi proizvod B.

Sa stanovišta ovog rada zanimljiva je metoda neuronske mreže (engl. *Neural Network*). Primjenom ove metode generira se arhitektura sustava koja simulira rad ljudskog mozga pri obavljanju određenog zadatka ili neke funkcije. Svoju su primjenu pronašle u mnogobrojnim aplikacijama problemskog tipa klasifikacije, prepoznavanja uzoraka, predviđanja, optimizacije i klasteriranja. Prediktivna sposobnost mreže otvorila im je područje na kome su našle veliku primjenu. Područje je to na kojem su se tradicionalno koristile statističke metode, kako klasične tako i multivarijacijske (Kumar et al. 1995, Lee i Young, 2000, Zang et al., 2004, Razi i Athappilly, 2005). Paliwal i Kumar (2009) tvrde kako je fokus primjene neuronskih mreža na problemima predviđanja i klasifikacije u različitim poljima, te ih stavljaju u odnos sa statističkim tehnikama i promatraju kao konkurentne alate za izgradnju predviđajućih sustava. Ovo razmatranje i usporedba različitih tipova metoda i tehnika je zanimljivo sa stanovišta praktične primjene te je na istraživanje tog problema postavljen jedan od akcenata i u ovom istraživanju.

Metode redukcije dimenzija operiraju nad višedimenzijskim setovima podataka. Pitanje koje se nastoje riješiti u procesu stjecanja znanja je koje su to varijable koje su značajne i važne u izučavanom fenomenu. Iako mnoge suvremene metode strojnog učenja konstruiraju prediktivne modele visokog stupnja točnosti uz višedimenzijske modele, svakako da se nastoji u mnogim primjenama smanjiti broj dimenzija originalnog modela (Fodor, 2002). Dva su osnovna tipa redukcijskih modela: linearni i nelinearni. Najpoznatiji linearni redukcijski metoda je PCA (engl. *Principal Component Analysis*) analiza (Fodor, 2002; Jackson, 1991), a slijede je faktorska analiza, PPR (engl. *Projection Pursuit*

Regression) metoda i ICA (engl. *Independent Component Analysis*). Nelinearna ICA metoda predstavnik je nelinearnih metoda redukcije dimenzija, a tu su još i metode vektorske kvantifikacije, genetičkog i evolucijskog algortima te poznate metode redukcije u regresijskoj analizi.

Orkestrirane metode su algoritmi učenja konstruirani od skupa klasifikatora čije se klasifikacije temelje na težinskom ili prostom glasovanju nad njihovim pojedinačnim predikcijama (Dietterich, 2000). Počivaju na pretpostavci da kombinacija više klasifikatora u predviđanju nezavisne varijable modela, češće daje točne rezultate nego pojedinačna od kombiniranih metoda klasifikacije. Hansen i Salamon (1990) definiraju neophodan i dovoljan uvjet da orkestar klasifikatora bude točniji od pojedinačnih članova takvim da su pojedinačni klasifikatori *točni* i *različiti*. Točan klasifikator pri tome je onaj koji ima bolje predviđanje od slučajnog predviđanja nezavisne varijable modela, a dva klasifikatora su različita ako prave različite pogreške na novim slučajevima klasifikacije. Neki od osnovnih tipova orkestriranja različitih metoda klasifikacije su Bayesov optimalni klasifikator (engl. *Bayes Optimal Classifier*), bootstrap agregiranje (engl. *Bagging*), poticanje (engl. *Boosting*), Bayesovo uprosječivanje parametara (engl. *Bayesian parametar averaging*), stekiranje.

2.2 Povijest razvoja neuronskih mreža

Moderna povijest znanosti o neuronskim mrežama započinje 1943. godine pionirskim radom Warrena McCulloha i Waltera Pitta pod naslovom „*A Logical Calculus of the Ideas Immanent in Nervous Activity*“ (McCulloch, Pitts, 1943). Rad je rezultat njihovih istraživanja provedenih u tridesetim i četrdesetim godinama prošlog stoljeća, a definira model kojim objašnjavaju način rada bioloških neurona čovjekovog živčanog sustava i mozga. Iako je nastao u vremenu nepostojanja digitalnog računala i bavio se razumjevanjem anatomije mozga, generirao je ideju na čijim osnovama je razvijeno polje umjetne inteligencije unutar kompjuterske znanosti. U periodu 1943.-1946. godine von Neumann koristi elemente ovog rada u projektiranju prvog elektronskog računala opće primjene (ENIAC). Donald Hebb (1949) definira poznati *postulat učenja* izučavajući

generalnu teoriju ponašanja u psihologiji. Iako čista neuropsihološka teorija, Hebbov rad postaje bitna teoretska osnova razvoja adaptivnih sustava i modela strojnog učenja.

1950-ih postaju dostupna i prva elektronska računala. Rochester et al. (1956) po prvi puta koriste računalo u simuliranju Hebovog postulata učenja. Formiraju neuronsku mrežu sastavljenu od 69 neurona pri čemu je svaki od njih povezan s 10 ostalih, a mreža u procesu simulacije modificira svoje težinske faktore. Nakon smrti von Neumanna, objavljeni su njegovi rukopisi u djelu pod nazivom *The Computer and the Brain*. Neuronsku mrežu von Neumann definira u kontekstu *pouzdanih strojeva za računanje sastavljenih od nepouzdanih elemenata* (1958), te postaje svjestan dubokih razlika između kompjutera i ljudskog mozga.

Iste godine Rosenblatt (1958) predlaže *perceptron* kao prvi model za nadzirano učenje. Perceptron čini najjednostavniju formu neuronske mreže. Sastoji se od jednog neurona s ulazima koji posjeduju prilagodljive sinpatičke težine i jednog pristranog ulaznog parametra (engl. *bias*). Algoritam za učenje, koji iterativno modificira slobodne težinske faktore mreže, definirao je Rosenblatt (1958, 1962). Njegovom uporabom stroj je bio sposoban klasificirati uzorke koji su linearno razdvojeni. 1960-e godine obilježene su perceptronom i mnogim radovima koji su dokazivali njegovu primjenjivost i Rosenblattovim *konvergencijskim teoremom*. Ipak, Minsky i Papert (1969) matematički dokazuju ozbiljna ograničenja mogućnosti perceptrona, točnije neuronske mreže s jednim slojem (engl. *Single-layer perceptron*). Istovremeno pribjegava se konceptu višeslojnog perceptrona u nadi da on može riješiti navedena ograničenja modela mreže s jednim slojem neurona.

Ovakva situacija dovela je do gotovo cijele dekade, u 1970-im i 1980-im godinama, u kojoj su se zahuktala istraživanja usporila skoro do zaustavljanja. Nedostatak računalne snage jako skupih računala i smanjenje ulaganja u ova istraživanja zbog izraženih sumnji u vrijednosti modela bili su neki od razloga zastoju. Stvar se ponovo pokreće 1982. godine kada John Hopfield s Kalifornijskog Tehnološkog Instituta (engl *California Institute of Technology*) definira propagiranje unatrag kao moguću metodu učenja neuronske mreže. Rumelhart, Hinton i Williams definiraju (Rumelhart et al. 1986) algoritam propagiranja unatrag (engl. *Back Propagation Algorithm*). Naknadno je otkriveno da je algoritam opisan u doktorskoj tezi Werbosa s Harvradskog sveučilišta iz 1974. godine, te u dijelu „*Applied Optimal Control"* od Brysona i Hoa iz 1969.godine.

Nakon toga kreće preporod ove grane strojnog učenja izražen naročito kroz aplikativnu primjenu. ANN metoda počinje se primjenjivati u rješavanju različitih problema,

ali posebno zapaženo mjesto zauzimaju primjene poput otkrivanja prijevara u upotrebi kreditnih kartica, prepoznavanja iznosa na rukom pisanim čekovima ili predviđanja cijena dionica (Zhang et al, 2004). Razvijajući nadalje polje neuronskih mreža i strojno učenje, višeslojni perceptron dobiva niz alternativnih metoda poput npr. RBF (engl. *Radial Basis Functions*) (Duda i Hart, 1973) i SVM (engl. *Support Vector Machines*) (Vapnik, 1992, 1995, 1998) metoda. U ovom vremenu su i regresijske statističke metode doživjele renesansu u korištenju, te se uvidjelo da su ove metode komplementarne u mnogim primjenama i da se mogu kombinirati u cilju dobivanja najboljih rezultata.

Zasluga za ovakav procvat pripada, među ostalim, računalima koja postaju procesorski mnogo jača, ali i financijski pristupačna široj istraživačkoj i poslovnoj zajednici. Bitan faktor uspjeha čini i automatizacija procesa u poslovnoj sferi djelovanja, te strukturirana upotreba računala i baza podataka s podacima nad kojima su neuronske mreže počele dokazivati svoju vrijednost. Sa znanstvenog stajališta Haykin (1999) pridaje radu Hopfielda iz 1982. godine i knjizi Rumelharta i McLelanda iz 1986. godine, definirajući neuronske mreže *interdisciplinarnim područjem s dubokim korijenima u neuroznanosti, psihologiji, matematici, fizici i inženjerstvu*.

2.3 Neuronska mreža

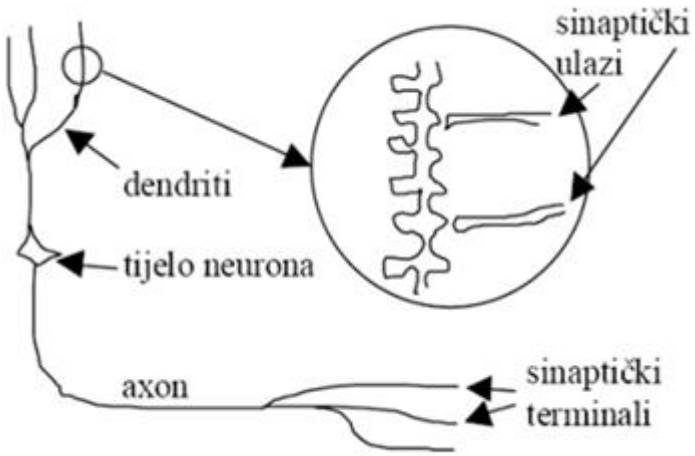
Neuronska mreža je masovno paralelno-distribuiran procesor s prirodnom sposobnošću memoriranja iskustvenog znanja (Fausett, 1994) i njegovog korištenja u naknadnim primjenama nad sustavom koji ga je modelirao svojim iskustvom. Haykin (1999) nudi definiciju mreže kao adaptivnog stroja i definira ANN (engl. Artificial Neural Network) kao *masivno paralelno distribuirani procesor sačinjen od jednostavnih procesnih jedinica (umjetnih neurona), koji ima prirodnu sposobnost pohranjivanja iskustvenog znanja raspoloživog za daljnju uporabu*. ANNs su, prema definiciji, distribuirani sustavi za procesiranje informacija, sastavljeni od mnogo jednostavnih i međusobno povezanih čvorova inspiriranih anatomijom ljudskog mozga i nervnog sustava čovjeka. Čvor ANN-a je analogija pojedinačne stanice živčanog sustava neurona. Strukturu ljudskog mozga čini u prosjeku između 50 i 150 milijardi neurona pri čemu svi nisu iste vrste. Ima ih oko 100 različitih vrsta, povezani su u skupine od nekoliko tisuća, te svojim međusobnim vezama čine pojedinačne neuronske mreže. Mozak dakle čini ogroman broj međusobno povezanih neuronskih mreža.

Osnovnu značajku ovako prikazanog seta mreža čini sposobnost učenja i odgovaranja na poticaje iz okoline. Uspoređujući biološke neuronske mreže i ANN, moglo bi se reći da je ANN stroj koji modelira način na koji ljudski mozak, kao set bioloških neuronskih mreža, odrađuje određene zadatke ili funkcije. Pri tome, ANN usvaja znanje iz okoline procesom učenja, a jačina interneuronskih veza, poznata kao sinaptička težina, koristi se za pohranjivanje znanja.

2.3.1 Biološki neuron

Biološki živčani sustav predstavlja mrežu specijaliziranih stanica koje šalju, prenose ili primaju informacije iz samog biološkog organizma ili iz njegove okoline. Na osnovi ovih informacija, živčani sustav upravlja njegovim ponašanjem. Ova arhitektura činila je polazište za produciranje sustava koji oponašaju procese rezoniranja svojstvene čovjeku. Središte živčanog sustava čini mozak. Ljudski mozak je kompleksno, nelinearno računalo sa sposobnošću paralelnog procesiranja velikog broja informacija. Osnovna gradivna stanica živčanog sustava je neuron. Neuron je specijalizirana ćelija koja propagira elektrokemijske signale. Slika 2.1 prikazuje osnovnu građu neurona koju čine dendriti kao ulazne jedinice, tijelo neurona, i axon kao izlazna jedinica.

Slika 2.1 Struktura biološkog neurona



Brzina procesiranja neurona mjeri se u milisekundama, dok se, za usporedbu, brzina procesiranja digitalnih jedinica računala mjeri u nanosekundama. Ova „sporost“ u

procesiranju nadoknađuje se ogromnim brojem neurona i spomenutom sposobnošću paralelnog procesiranja informacija. Interkonekcija se ostvaruje putem sinapsi koje povezuju axone jednog neurona s dendritima drugog neurona. Aktivnost neurona ogleda se u generiranju elektrokemijskog signala i njegovom prijenosu putem axona do dendrita drugog neurona. Bitno je istaknuti da ovo generiranje i prijenos elektrokemijskog signala, nastaje uz podražaj signala primljenog od drugog neurona i to tek ako jakost tog signala prelazi određenu razinu koja se naziva aktivacijski prag (engl. *firing threshold*).

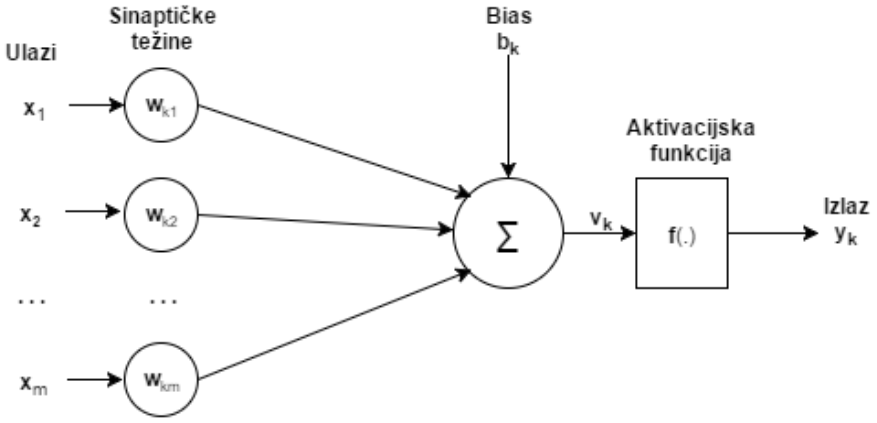
Spomenuta jačina signala koju prima neuron zavisi od djelotvornosti sinapsi. Svaka sinapsa sadrži prostor ispunjen neurotransmiterima koji prenose signal. Hebbova (1949) istraživanja neurološkog sustava definirala su postulat da je učenje u principu sastavljeno od alteriranja „snage“ sinaptičkih veza. Mozak na temelju rada velikog broja relativno jednostavnih procesnih jedinica uspijeva upravljati ekstremno kompleksnim procesima. Svaka od ovih procesnih jedinica (neurona), prima i zbraja niz ulaznih signala te ako njihova veličina prelazi određenu razinu, generira slijedeći signal i prosljeđuje ga slijedećoj procesnoj jedinici. Ovo je pojednostavljen model djelovanja na kome su izgrađene umjetne neuronske mreže.

2.3.2 Umjetni neuron

Na osnovama biološkog neurona izgrađen je model umjetnog neurona koji čini elementarnu gradivnu jedinicu umjetne neuronske mreže. Tri su osnovna elementa koja ga čine: skup sinapsi ili ulaznih veza, procesna jedinica i aktivacijska funkcija.

Ulazne veze su kanali kojima ulazni signal dolazi do procesne jedinice neurona. Svaku od ulaznih veza $\mathbf{x_j}$ karakterizira sinaptička težina $\mathbf{w_{kj}}$. Sinaptička težina umjetnog neurona može imati pozitivnu i negativnu vrijednost, za razliku od sinaptičke težine prirodnog neurona koja ne može posjedovati negativnu vrijednost. Indeks ***j*** kod oznake ulaznog signala označava sinapsu ulaznog signala, dok indeks ***kj*** kod oznake sinaptičke težine definira neuron (k) i pripadajuću sinapsu (j).

Slika 2.2 Umjetni neuron



Procesna jedinica zbraja ulazne signale multiplicirane vrijednošću pripadajuće sinaptičke težine i generira ulaz u aktivacijsku funkciju neurona. *Aktivacijska funkcija* transformira dobiveni ulaz u izlazni signal neurona. Njen zadatak je normalizirati amplitudu izlaznog signala, uobičajeno na interval [0,1] ili [-1,1].

Matematička notacija opisuje model umjetnog neurona sljedećim jednadžbama.

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \qquad (2.2)$$

$$v_k = u_k + b_k \qquad (2.3)$$

$$y_k = f(v_k) \qquad (2.4)$$

pri čemu su:

u_k ulaz procesne jedinice neurona *k,*

w_{kj} sinaptičke težine sinapse *j* neurona *k,*

$x_1, x_2, \ldots, x_m$ ulazni podaci sinapsi *1,2,…,m* ,

v_k aktivacijski potencijal neurona k,

b_k bias faktor neurona *k,*

$\mathbf{y_k}$ izlazni podataka neurona *k*,

f(.) aktivacijska funkcija neurona.

U praksi, ovisno o problemu koji se modelira, koristi se mnogo tipova aktivacijskih funkcija, ali tri su osnovna tipa: funkcija praga, linearna funkcija po odsječcima i sigmoidna funkcija.

Matematička notifikacija *funkcije praga* definira da je vrijednost funkcije jednaka 1 za aktivacijski potencijal neurona veći ili jednak od 0, te jednaka 0 za ostale vrijednosti.

$$f(\mathbf{v}) = \begin{cases} 0, v < 0 \\ 1, v \geq 0 \end{cases} \qquad (2.5)$$

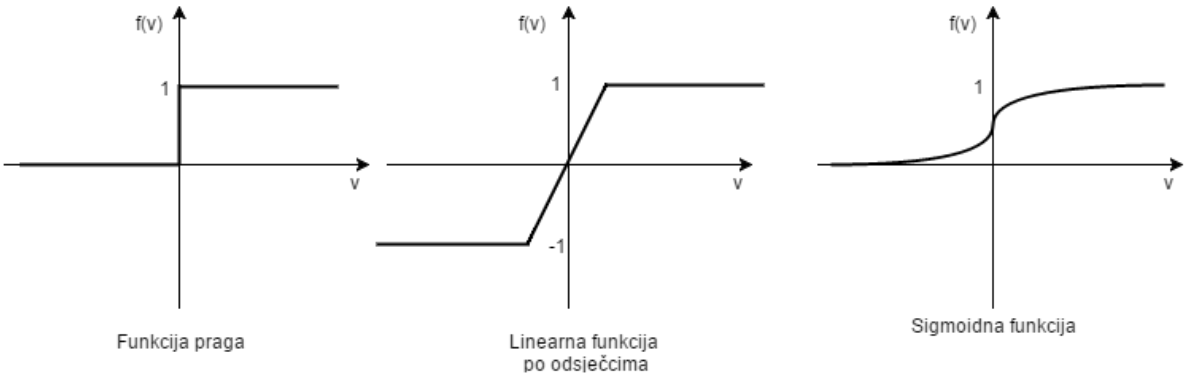
Linearna funkcija po odsječcima vraća vrijednost samog argumenta v za vrijednosti v koje su između -1 i 1. Funkcija vraća vrijednost -1, za aktivacijski potencijal manji od -1, odnosno 1 za aktivacijski potencijal veći od 1.

$$f(v) = \begin{cases} 1, & v \geq 1 \\ v, -1 \leq v \leq 1 \\ -1, v \leq -1 \end{cases} \qquad (2.6)$$

Sigmoidna funkcija je najčešći tip funkcije koji se koristi kao aktivacijska funkcija neurona. U navedenoj definiciji (jednadžba 2.6), α je parametar koji definira izgled S krivulje. Funkcija poprima vrijednosti u rasponu od 0 do 1, tj. $0 \leq f(v) \geq 1$.

$$f(v) = \frac{1}{1+e^{-\alpha v}} \qquad (2.7)$$

Slika 2.3 Grafovi osnovnih aktivacijskih funkcija



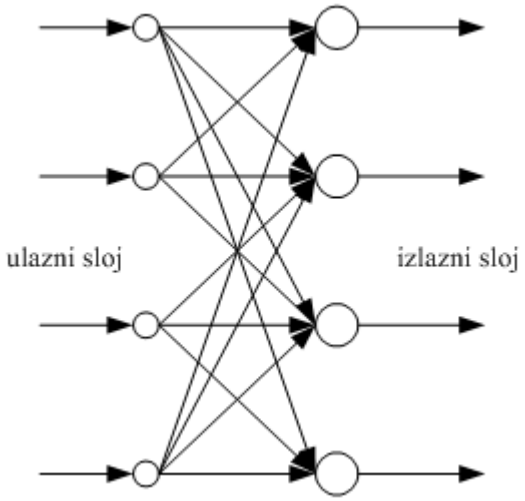
2.4 Arhitekture neuronskih mreža

U strukturi umjetne neuronske mreže neuroni su organizirani u slojeve. Na osnovu broja slojeva i smjera veze između neurona, definirane su tri osnove klase arhitekture neuronskih mreža:

- Jednoslojne mreže bez povratnih veza (engl. *Single-Layer Feedforward Networks*),
- Višeslojne mreže bez povratnih veza (engl. *Multilayer Feedforward Networks*),
- Mreže s povratnim vezama (engl. *Recurrent Networks*)

Jednoslojna mreža je najjednostavnija struktura mreže.

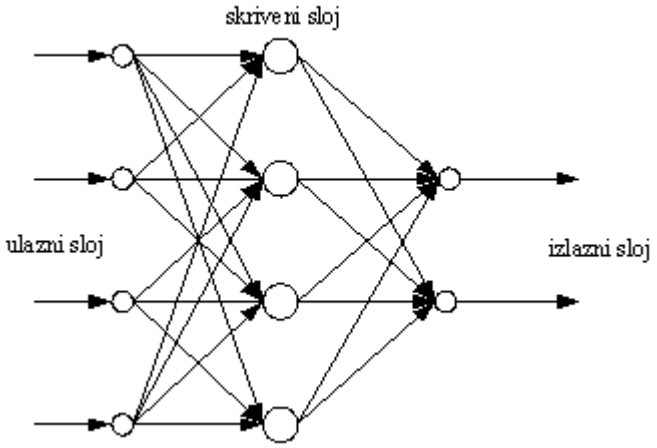
Slika 2.4 Jednoslojna mreža bez povratnih veza



Ulazni sloj mreže čine čvorovi koji propagiraju ulazne podatke sustava na neurone izlaznog sloja mreže. Na ulaznom sloju se ne vrše računanja i obrade podataka, te se ovaj sloj ne računa u slojeve mreže. Pojedini neuroni izlaznog sloja mreže vrše računanja i transformiraju skup ulaznih podataka u izlazni signal. Stoga ovaj sloj čini jedini sloj mreže prema čemu je mreža i dobila naziv. Ne postoji povratna veza koja bi izlazni signal mreže na bilo koji način učinila ulaznim podatkom, te je ova mreža deklarirana kao mreža *bez povratnih veza*.

Višeslojna mreža bez povratnih veza osim ulaznog i izlaznog sloja sadrži i jedan ili više skrivenih slojeva neurona. Primjer ovakve strukture sa slike x pokazuje strukturu koja se osim ulaznog sloja (koji se u biti i ne sastoji od neurona već od ulaznih jedinica) i izlaznog sloja (koji sadrži dva neurona) sastoji od još jednog skrivenog sloja (koji sadrži 4 neurona).

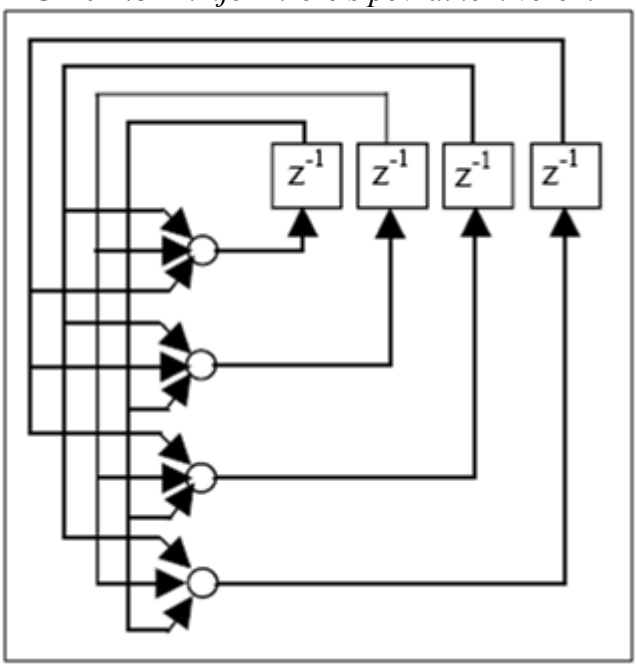
Slika 2.5 Višeslojna mreža bez povratnih veza



Skriveni slojevi donose novi skup sinaptičkih veza čime povećavaju računalnu moć mreže osobito u slučajevima velikog broja ulaznih podataka mreže. Vektor ulaznih podataka propagira se na neurone sljedećeg sloja. Nakon transformacije ulaznog vektora podataka, izlazi pojedinih neurona promatranog sloja čine ulaze sljedećeg sloja i tako sve do izlaznog sloja mreže. Izlazi zadnjeg sloja mreže čine odgovor mreže na poticaj ulaznog vektora podataka. Kao i kod prethodne arhitekture, veze su jednosmjerne i idu isključivo od ulaza prema izlazu, odnosno ne postoje povratne veze. Za mrežu sa slike x može se reći da je *potpuno povezana* mreža, jer je svaki čvor mreže jednog sloja u potpunosti povezan s čvorovima mreže sljedećeg sloja. Nedostatak neke od ovih veza, mrežu bi činile *parcijalno povezanom* mrežom.

Mreže s povratnom vezom imaju barem jednu povratnu vezu. Primjer sa slike x prikazuje jednu takvu arhitekturu. U ovom primjeru može se uočiti da ulaz bilo kojeg neurona čine izlazi svih ostalih neurona osim samog sebe. Povratna veza daje dodatnu kvalitetu mreži, ali je istovremeno i jako usložnjava.

Slika 2.6 Primjer mreže s povratnom vezom



2.5 Učenje u neuronskim mrežama

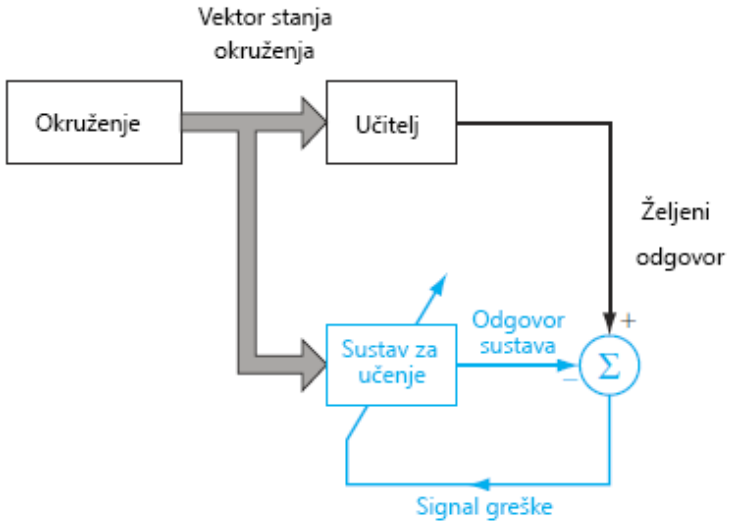
Haykin (2009) definira učenje kao proces u kojem se slobodni parametri neuronske mreže adaptiraju kroz proces stimuliranja od strane okoline u koju je mreža uklopljena, a tip učenja je određen načinom na koji se dešavaju promjene parametara. Definicija ističe glavno svojstvo mreže, ali i svih drugih metoda strojnog učenja, a to je sposobnost učenja od okruženja i unapređenja performansi kroz proces učenja. Sposobnost učenja svojstvena čovjeku bila je osnova za konstruiranje strojeva ili računalnih programa sa sposobnošću učenja iz okruženja. Po načinu učenja, analogno učenju ljudskih bića, dijele se i mreže na one koje uče s učiteljem i bez učitelja. Učiteljem se naravno ne smatra čovjek ili stroj koji uči mrežu, već postojanje rezultata učenja. Zbog toga se učenje uz prisutnost rezultata učenja naziva nadziranim, a ono bez njih nenadziranim učenjem. Postoji još i dodatna vrsta učenja bez učitelja, tzv. podržano učenje.

2.5.1 Učenje „s učiteljem“

Učenje s učiteljem je paradigma koja pretpostavlja postojanje znanja o okruženju iz kojeg mreža uči. Elementarno znanje o okruženju, u ovom kontekstu, može se definirati skupom nezavisnih (ulaznih) varijabli i skupom zavisnih (izlaznih) varijabli čija vrijednost ovisi o nezavisnim. Ovo znanje je iskustveno znanje, nastalo na osnovu poznatog ponašanja iz prošlosti, te se njime nadzire ili kontrolira proces učenja. Iz ovog je proizašao termin nadziranog učenja.

Vektor ulaznih varijabli, koje opisuju stanje sustava koji modelira neuronska mreža, djeluje na modelirani sustav za učenje i na učitelja. Sustav za učenje odgovara na poticaj iz okoline i postavlja skup sinaptičkih težina na odgovarajuće vrijednosti. Generirani skup težinskih faktora zajedno s vektorom ulaznih podataka proizvodi izlaznu vrijednost. Kako je učitelju poznata točna izlazna vrijednost okruženja, on je prezentira te se generira signal greške koji činu razliku dvije izlazne vrijednosti. Povratna veza ga prenosi sustavu za učenje, a na osnovu njega sustav prilagođava interne težinske faktore u nastojanju da svoje ponašanje prilagodi okruženju i generira željeni izlaz. Ovo prilagođavanje se odigrava kroz niz ponavljanja opisanih koraka.

Slika 2.7 Blok dijagram učenja „s učiteljem“



Preuzeto iz (Haykin, 1999)

Stupanj prilagođenosti sustava okruženju koje se modelira, iskazuje se statističkim pokazateljima signala greške. Kada ovi pokazatelji, a obično se radi o srednjoj vrijednosti

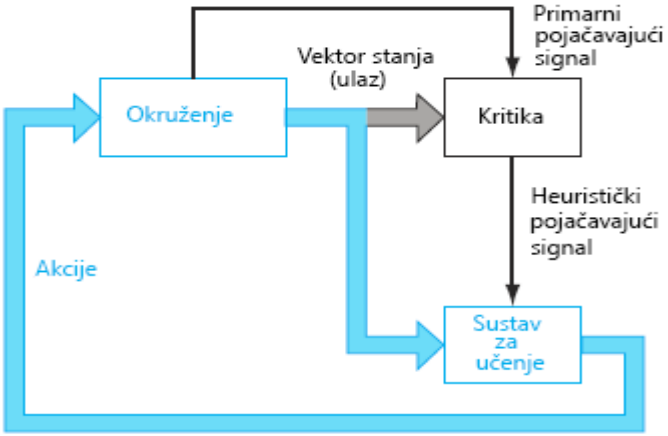
greške ili sumi kvadrata greške, dostignu zadovoljavajuće vrijednosti sa statističkog gledišta, mreža može samostalno djelovati i bez učitelja. Znanje stečeno procesom učenja, pohranjeno je u težinskim faktorima sustava. Ovo učenje predstavlja osnovu učenja korekcijom pogreške. Učenje se odvija u zatvorenoj povratnoj petlji, a vanjski sustav koji se modelira van je ove petlje.

2.5.2 Učenje „bez učitelja“

Nenadzirano učenje ne pretpostavlja poznato ponašanje okruženja koje se modelira. Praktično govoreći, ne postoje vektori podataka pojedinačnih stanja promatranog sustava s točno određenim ulaznim i izlaznim veličinama. Ovi sustavi učenja razlikuju dvije kategorije: podržano i nenadzirano učenje.

Blok dijagram procesa *podržanog učenja* prikazan je na slici 2.8. Proces učenja se i ovdje odvija u zatvorenoj povratnoj petlji, ali za razliku od prethodno opisanog nadziranog učenja, vanjska okolina je dio petlje. Glavnina računskog dijela procesa učenja odvija se oko tzv. kritičara (engl. *Critic*). On zaprima primarni pojačavajući signal iz okruženja i varijable stanja okruženja, te generira heuristički pojačavajući signal.

Slika 2.8 Blok dijagram podržanog učenja



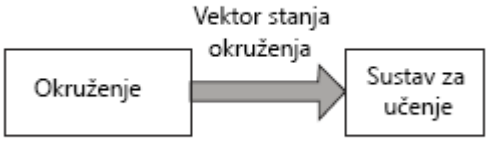
Preuzeto iz (Haykin, 1999)

U kombinaciji vektora ulaznih veličina okruženja i heurističkog pojačavajućeg signala, sustav za učenje generira izlaznu veličinu sustava učenja koju ponovo šalje u

povratnu petlju procesa. Cilj podržanog učenja je minimiziranje *cost-to-go funkcije*, definirane kao očekivanje kumulativnog *troška akcija* obrađenih slučajeva, kroz sekvencu koraka učenja, umjesto jednostavnog pojedinačnog troška slučaja (Haykin, 2009). Stoga se funkcionalnost ovog sustava učenja ogleda u prepoznavanju ovih slučajeva i njihovom vraćanju u sustav povratnom vezom. Prepoznavanje slučajeva očigledno uključuje vremensko promatranje slučajeva i pamćenje njihovih pojačavajućih signala kao indikatora poželjnih stanja sustava. Ovakav stroj za učenje mora biti u stanju vrednovati svaki od slučajeva u vremenskom slijedu koraka koji vodi do konačnog ishoda učenja, te čini jednu klasu sustava za učenje veoma interesantnu sa stanovišta mogućnosti interakcije s okruženjem.

Nenadzirano učenje odvija se i bez poznatog znanja i bez *kritičara*. Slučajevi koji se promatraju nisu označeni ni na koji način. Sustav uči iz skupa ulaznih podataka prepoznajući pravilnosti i uzorke u njima, dolazeći do određenih spoznaja koje rezultiraju grupiranjem podataka. Pred sustav se postavlja prag tolerancije, neovisna mjera kvalitete koju sustav treba postići. Sukladno njoj, sustav optimizira svoje ponašanje, te u težnji da dostigne ovu mjeru razvija sposobnost dekodiranja vektora ulaznih podataka i njihovog smještanja u odgovarajuće klase podataka.

Slika 2.9 Blok dijagram nenadziranog učenja



Preuzeto iz (Haykin, 1999)

2.5.3 Algoritmi učenja

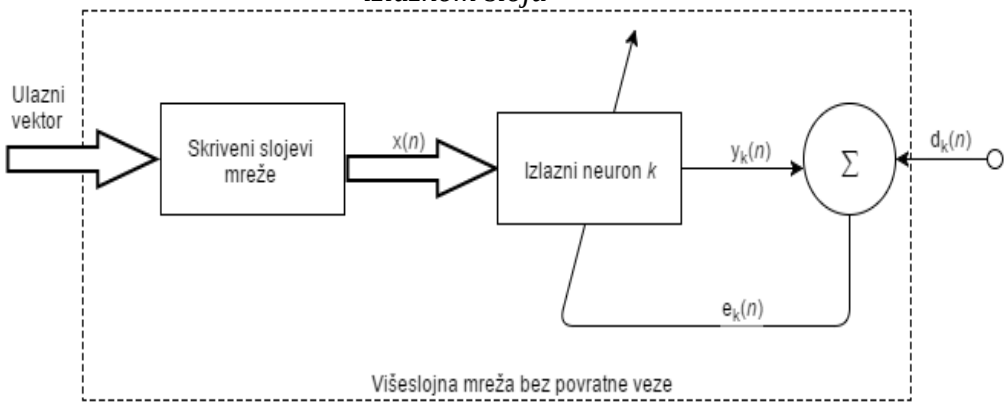
Proces učenja odvija se kroz jasno definiran slijed koraka. Mreža se aktivira impulsima iz okoline. Na osnovu stimulansa mijenja se vrijednost težinskih faktora. Na nove stimulanse mreža odgovara prilagođavanjem svoje interne strukture kako bi adekvatno odgovorila novim uvjetima okruženja. Opisani slijed koraka čini algoritam učenja. Pet je osnovnih algoritama učenja u neuronskim mrežama:

1. Učenje korekcijom pogreške (engl. *Error correction learning*),
2. Učenje zasnovano na memoriranju (engl. *Memory-based learning*),
3. Hebbovo učenje (engl. *Hebbian learning*),
4. Kompetitivno učenje (engl. *Competitive learning*),
5. Boltzmanovo učenje (engl. *Boltzman learning*).

2.5.3.1 Učenje korekcijom pogreške

Algoritam učenja korekcijom pogreške klasičan je predstavnik učenja s postojanjem znanja modelirane okoline. Za objašnjenje algoritma koristiti će se se višeslojni perceptron s jednim neuronom u izlaznom sloju. Blok dijagram višeslojne mreže prikazan je na slici 2.10. Mreža može sadržavati više neurona u izlaznom sloju, ali zbog jednostavnosti prikazan je samo jedan neuron i označen sa *k*. Vektor *x(n)* potiče rad neurona *k*, a dobiven je procesiranjem ulaznog vektora kroz prethodne skrivene slojeve mreže. Argument *n* predstavlja vremensku notaciju, odnosno vremenski korak u ponavljajućem procesu učenja mreže. Izlaz promatranog neurona označen je s $y_k(n)$.

Slika 2.10 Blok dijagram višeslojne mreža bez povratne veze s jednim neuronom u izlaznom sloju



Izlazni signal $y_k(n)$ usporedi se s željenim izlaznim signalom $d_k(n)$ generirajući *signal greške* $e_k(n)$.

$$e_k(n) = d_k(n) - y_k(n) \qquad (2.8)$$

Generirani signal greške čini korektivni mehanizam ovog algoritma. On je osnova za korekciju težinskih faktora neurona *k* s ciljem smanjenja pogreške koju će generirati mreža obrađujući naredni skup veličina ulaznog vektora na ovom neuronu. Algoritam učenja teži minimizirati troškovnu funkciju *C(n)* definiranu signalom pogreške pri čemu se iteracija zaustavlja, a mreža dostiže stabilno stanje. Matematička notacija troškovne funkcije je:

$$C(n) = \frac{1}{2} e_k^2(n) \qquad (2.9)$$

Na osnovama minimiziranja troškovne funkcije nastalo je delta pravilo koje se još naziva i Widrow-Hoff pravilo (Widrow i Hoff, 1960) prema njegovim tvorcima. Matematička notacija ovog pravila prikazana je slijedećom jednadžbom:

$$\Delta w_{kj}(n) = \mu e_k(n) x_j(n) \qquad (2.10)$$

U jednadžbi koja definira spomenuto pravilo, $w_{kj}(n)$ predstavlja težinski faktor neurona *k* koji je potaknut djelovanjem elementa $x_j(n)$ ulaznog vektora *x(n)* u iteracijskom koraku *n*. Δw_{kj} predstavlja prilagodbu težinskog faktora w_{kj} u n-toj iteraciji i upravo je proporcionalno tzv. parametru stope učenja μ. Izvorna definicija Widrow-Hoff pravila glasi:

> *Korekcija sinaptičke težine neurona proporcionalna je produktu signala greške i ulaznog signala odnosne sinapse.*

Delta pravilo egzistira na mogućnosti direktnog mjerenja signala pogreške i u osnovi je sustav zatvorene povratne veze. Stabilnost ovakve veze determinirana je parametrima koji je direktno određuju. Jedan od njih je parametar učenja μ. Odabir veličine ovog parametra ima odlučujuću ulogu u preciznosti i pouzdanosti sustava nastalog procesom učenja.

2.5.3.2 Učenje zasnovano na memoriranju

Znanje potrebno za primjenu ovog algoritma pohranjeno je u fizičku memoriju sustava. Matematički bi ovo znanje mogli predstaviti skupom uređenih parova. Prvi član para je vektor ulazne veličine $\mathbf{x}_i$, a drugi član je izlazna veličina sustava $\mathbf{d}_i$. Ako je u memoriju sustava za učenje ovim algoritmom pohranjeno **N** slučajeva, matematička notifikacija memorije bila bi:

$$\{(x_i, d_i)\}_{i=1}^{N} \qquad (2.11)$$

Algoritam učenja prihvata vektor ulaznih veličina $\mathbf{x}_j$ i zadatak mu je da klasificira promatrani slučaj i odredi ciljnu veličinu $\mathbf{d}_j$. Kako ovaj algoritam radi? Postoje dvije varijante algoritma: *metoda najbližih susjeda* i metoda *k-najbližih susjednih klasifikacija*. Razlikuju se po kriteriju specificiranja najbližih susjeda i pravilu učenja koje se primjenjuje na njih.

Prva metoda definira najbližeg susjeda na način da pronalazi vektor iz skupa pohranjenog u memoriji sustava koji ima minimalnu Euklidovu distancu s vektorom ulaznih veličina promatranog slučaja koji se želi klasificirati. Označe li se s $\mathbf{x}_t$ ulazni podatci slučaja koji se testira i s $\mathbf{x}_A$ najbliži vektor susjed, tada najbliži susjed zadovoljava slijedeću jednadžbu:

$$\min_i d(x_i, x_t) = d(x_A, x_t), \qquad (2.12)$$

u kojoj $d(x_i,x_t)$ predstavlja Euklidsku udaljenost vektora $\mathbf{x}_i$ i $\mathbf{x}_t$. Promatranom slučaju dodijeliti će se izlazna veličina na način da se ista preuzme od određenog najbližeg susjeda.

Metoda k-najbližih susjeda funkcionira na sličan način. Prethodna metoda je određivala jednog susjeda (najbližeg) i preuzimala klasifikaciju od njega, a ova određuje *k* najbližih susjeda i klasificira promatrani slučaj na osnovu njihove frekventnije klasifikacije.

2.5.3.3 Hebbovo učenje

Hebb (1949) definira pravilo učenja s neurobiološkog stanovišta stavljajući u odnos sinaptičku aktivnost povezanih neurona. Prema ovom postulatu, ako akson stanice A potakne stanicu B i učestvuje permanentno u ovom poticanju, tada se u stanici A odigrava određeni metabolički proces na način da se povećava njena aktivnost. Prevedeno na jezik učenja u umjetnoj neuronskoj mreži, pravilo definira da se povećava težinski faktor sinapse na čijim krajevima se nalaze neuroni koji se sinkrono aktiviraju. Matematička notacija generalnog pravila Hebbovog učenja može se predstaviti izrazom:

$$\Delta w_{kj}(n) = F\left(y_k(n), x_j(n)\right) \qquad (2.13)$$

Promjena sinaptičke težine neurona k je funkcija dva argumenta: presinaptičkog signala (vektor ulaznih veličina x_j) i postsinaptičkog signala y_j (izlazna veličina neurona k) u vremenskom koraku n. Najjednostavnija forma Hebbovog pravila

$$\Delta w_{kj}(n) = \mu y_k(n) x_j(n) \qquad (2.14)$$

definira stopu učenja μ. Stopa učenja je pozitivna konstanta i definira odnos između promjene sinaptičke težine neurona i produkta ulaznog i izlaznog signal neurona.

Sejnowski (1977) modificira Hebbovu hipotezu definirajući tzv. hipotezu kovarijance. On mijenja u Hebbovoj hipotezi ulazni i izlazni signal s promjenom njihove razine u odnosu na srednje vrijednosti ovih signala za određeni vremenski period.

$$\Delta w_{kj}(n) = \mu\left(x_j - \bar{x}\right)\left(y_j - \bar{y}\right) \qquad (2.15)$$

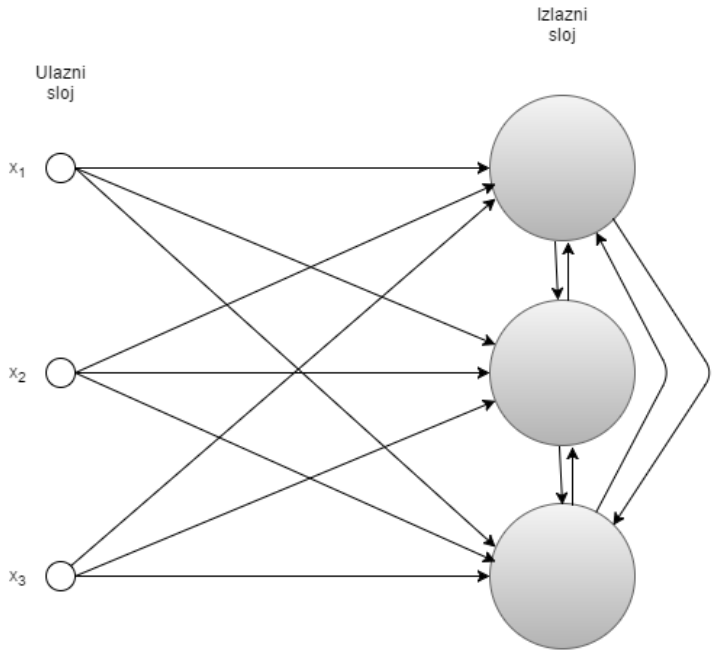
Iz jednadžbe se može zaključiti da do povećanja sinaptičke težine dolazi uslijed postojanja dovoljno velike razine presinaptičke i postsinaptičke aktivnosti. Velika razina pretpostavlja vrijednosti ovih veličina (ulazi i izlaz neurona *k*) većih od prosječne vrijednosti, odnosno takve da je $x_j > \bar{x}$ i $y_j > \bar{y}$. Do pada sinaptičke težine neurona *k* dolazi uslijed smanjenja jedne od sinaptičkih aktivnosti uz postojanje druge na odgovarajućoj razini.

2.5.3.4 Kompetitivno učenje

Algoritam kompetitivnog učenja počiva na koncepciji mreže u kojoj samo jedan od izlaznih neurona može biti aktivan u bilo koje vrijeme, dok su svi ostali izlazni neuroni neaktivni. Očigledna je svrha mreže koja uči prema ovom algoritmu, a to je klasifikacija predložaka ulaznog seta podataka. Mehanizam učenja osigurava izlaznim neuronima da se međusobno *natječu* tako da samo jedan od njih *pobjeđuje*.

Kako bi se osiguralo *natjecanje* izlaznih neurona mreže, neophodno je postojanje veze među njima na način da svaki izlazni neuron pruža ostalim izlaznim neuronima informaciju o vrijednosti vlastitog lokalnog induciranog polja v_k. Prosta arhitektura mreže koja uči po ovom algoritmu sadrži jedan ulazni sloj s tri čvora i samo jedan, izlazni, sloj mreže.

Slika 2.11 Jednostavna arhitektura mreže kompetitivnog algoritma učenja



Matematička notifikacija pravila učenja definira da izlazni neuron s najvećom vrijednošću lokalno induciranog polja pobjeđuje, odnosno ima izlaznu vrijednost 1, dok svi ostali imaju vrijednost 0:

$$y_k = \begin{cases} 1 & v_k > v_j, j \neq k \\ 0 \; za \; ost. slučajeve \end{cases} \quad (2.16)$$

Inducirano polje v_k predstavlja vrijednost kombinacije aktivnosti ulaznih signala neurona k koji dolaze s ulaznih čvorova mreže i povratnom vezom od generiranih vrijednosti neurona izlaznog sloja.

Proces učenja, sa stanovišta algoritma, definiran je prilagodbom težinskih faktora mreže. Inicijalno se skupu sinapsi jednog neurona dodjeljuje ukupna vrijednost svih težina u iznosu od 1. Pretpostavka je također, da je svaka pojedinačna vrijednost pozitivna. Označivši s w_{kj} sinaptičku težinu sinapse koja povezuje ulazni čvor mreže *j* s neuronom *k,* na osnovu definiranih pretpostavki, može se za svaki neuron mreže napisati

$$\sum_j w_{kj} = 1 \quad (2.17)$$

U nastavku procesa učenja odvija se *natjecanje* među neuronima. U prvom koraku, pojedinačni neuroni vrše preraspodjelu sinaptičke težine neaktivnih ulaznih čvorova na aktivne ulazne čvorove. Zatim se očitava aktivnost neurona uz ovako preraspodijeljene težinske faktore i odgovarajući vektor ulaznih podataka. Neuroni mogu reagirati ili ne reagirati, a ovisno o tome nastavljaju se naredne akcije procesa učenja. U situaciji da neuron ne reagira, on se proglašava neaktivnim i ispada iz daljeg procesa učenja. Ukoliko pojedini neuron pobjedi u natjecanju, svaka ulazna sinapsa tog neurona odriče se proporcionalnog dijela svog težinskog faktora, a taj dio se alocira ravnomjerno na ostale aktivne čvorove. Kao i kod prethodnih algoritama s μ je označena stopa učenja, te se promjena težinskih faktora w_{kj} matematički može definirati na slijedeći način:

$$\Delta w_{kj} = \begin{cases} \mu(x_j - w_{kj}) & neuron\ k\ pobjeđuje\ u\ natjecanju \\ 0 & neuron\ k\ ne\ pobjeđuje\ u\ natjecanju \end{cases} \quad (2.18)$$

Algoritam kompetitivnog učenja prikazuje sposobnost neuronskih mreža da provode klaster analizu. Svakako, preduvjet uspješnosti ovih analiza je postojanje predložaka ulaznih vektora podataka koji gravitiraju dovoljno jasno distribuiranim klasterima. U suprotnom, mreža se ponaša nestabilno i ne može jasno odgovarati odgovarajućim signalima izlaznih neurona na poticaje vektora ulaznih signala.

2.5.3.5 Boltzmanovo učenje

Algoritam Boltzmanovog učenja osnovu nalazi u principima termodinamike i statističke mehanike. Neuronsku mrežu zasnovanu na ovom algoritmu, Boltzmanov stroj, čine neuroni raspoređeni u vidljive i skrivene slojeve mreže s povratnim vezama. Pretpostavka mreže je da neuroni ne mogu imati povratnu vezu na samoga sebe i da djeluju binarno, odnosno da imaju dva moguća stanja: 1 i -1.

Stanje mreže opisuje tzv. energetska funkcija određena stanjem svih neurona i pripadajućih težinskih faktora. Ako se energetska funkcija označi s *E*, stanje neurona j s x_j, stanje neurona k s x_k, te sinaptička težina veze između neurona j i k s w_{jk}, matematička notacija funkcije *E* može se predstaviti jednadžbom:

$$E = -\frac{1}{2}\sum_j \sum_j w_{kj} x_k x_j \text{ , pri čemu je } j \neq k \quad (2.19)$$

Algoritam učenja, slučajnim odabirom, bira pojedinačni neuron mreže i mijenja mu stanje iz x_k u $-x_k$ u uvjetima postojanja određene razine *temperature T*. T je pseudotemperatura stroja, a ne nikakva stvarna fizička temperatura. Izrazi *temperatura* i *energija* preuzeti su iz termodinamičke teorije na čijim osnovama je zasnovan algoritam.

Konstantnim ponavljanjem ovog procesa, mreža mijenja razinu svoje energetske funkcije i dostiže tzv. termalni ekvilibrijum, odnosno stanje u kome s određenom vjerojatnošću ispravno odgovara na poticaje iz okoline. Vjerojatnost da će baš neuron *k* biti slučajno odabran i da će mu se promijeniti stanje iskazana je jednadžbom u kojoj je s ΔE_k izražena promjena energije mreže u promatranom koraku učenja:

$$P(x_k \rightarrow -x_k) = \frac{1}{1+exp(-\Delta E_k/T)} \qquad (2.20)$$

Neuroni Boltzmanovog stroja iz vidljivog sloja djeluju pod utjecajem okoline, dok oni iz skrivenih slojeva djeluju slobodno. Promjena težinskih faktora mreže, prema Boltzmanovom algoritmu, upravo je proporcionalna parametru stope učenja mreže i razlike korelacija stanja neurona j i k pri slobodnom i djelovanju uvjetovanom okolinom.

$$\Delta w_{kj} = \mu\left(\rho_{kj}^{+} - \rho_{kj}^{-}\right), j \neq k \qquad (2.21)$$

U jednadžbi koja opisuje osnovno pravilo učenja Boltzmanovog algoritma, Δwkj je promjena težinskog faktora veze između neurona *k* i neurona *j*, μ je stopa učenja, a ρ_{kj} predstavlja korelaciju između stanja neurona *j* i *k* u uvjetima slobodnog djelovanja (+) ili djelovanja pod utjecajem okoline (-).

2.5.4 Algoritam širenja unazad

Algoritam „širenja unazad“ (engl. *Back Propagation Algorithm*) je najviše korišten algoritam praktične primjene neuronskih mreža. Algoritam su predložili Widrow i Hoff (1960), a doprinos popularnosti dali su mu Rumelhart, Hinton i Williams (Rumelhart et al., 1986). Zanimljivo je da su do ovog algoritma svojim nezavisnim istraživanjima došli i mnogi drugi autori.

Algoritam se sastoji od tri koraka:

1. označavanje ulaznih čvorova, punjenje ulaznim varijablama u odgovarajućem redu i propagiranje unaprijed kroz slojeve mreže ka izlaznom sloju,
2. produciranje i zapisivanje vrijednosti izlaznih čvorova i računanje greške svakog izlaznog čvora,
3. propagiranje greške unazad počevši od izlaznog sloja uz modificiranje sinaptičkih težina kako bi se minimizirala očekivana greška.

Osnova za izgradnju algoritma je Delta pravilo, pa ga se još naziva i generalizirano delta pravilo. Delta pravilo pojačava vezu između dva neurona koristeći vrijednost izlaza prvog neurona O_j i potencijal drugog neurona za redukciju greške,a relativno s željenom ciljnom vrijednošću.

Algoritam se koristi za prilagodbu sinaptičkih težina tijekom faze testiranja mreže. U radu se oslanja na pravilo gradijentnog opadanja, optimizacijsku tehniku koja koristi pravac maksimalne promjene (gradijent) dužinom višedimenzijske nelinearne funkcije kako bi se inkrementalno došlo do globalnog optimuma. Spomenuta inkrementalna veličina bira se od strane korisnika. Pravilo može biti prvog reda (najstrmije opadanje) ili drugog reda (konjugirani gradijent). U neuronskim mrežama uporaba ovog pravila se ogleda u konstantnom smanjivanju izlazne greške korištenjem malih prilagodbi težinskih faktora mreže. Gradijent definiran izrazom

$$\frac{\partial E}{\partial w_{ij}} \quad (2.22)$$

daje relativnu promjenu promatrane greške E za izvršenu promjenu težinskog faktora w_{ij}. Dakle, cilj je kretanje najstrmijim pravcem u prostoru težinskih faktora izvodeći simultane izmjene svih ulaznih težinskih faktora za fiksnu izlaznu grešku pojedinog neurona. Koristeći testni set podataka postupak se ponavlja i trebao bi rezultirati sve boljom i boljom predikcijom izlaznih veličina. Vrijeme potrebno za potpuno učenje relacije između ulaznih i izlaznih veličina mreže, naziva se vrijeme učenja ili vrijeme konvergencije. Preveliko ili

premalo vrijeme učenja indicira neadekvatnost topologije mreže ili pak grešku u odabiru odgovarajućeg algoritma učenja, aktivacijske funkcije, transformacije podataka i sl. Slijedeći problem može izazvati pre-trenirana mreža (engl. *Over-trained network*) u fazi korištenja. Potrebno je naime i sam proces učenja zaustaviti u određenoj fazi. Indeks performansi mreže može se koristiti za zaustavljanje učenja na odgovarajućoj razini utreniranosti mreže.

Postoji mnogo varijanti algoritma širenja unazad. Uz standardnu proceduru algoritma, mogu se spomenuti neke poput:

- algoritam s promjenjivom veličinom koraka (engl. *Backpropagation with Variable Step Size* - BPVS)
- online adaptivni BP (engl. *Online Adaptive Backpropagation* – OABP)
- koeficijent učenja i momentum adaptivni algoritam (engl. *learning rate and momentum adaptive BP*)
- elastični algoritam (engl. *Resilient BP*)
- delta-bar-delta algoritam (DBD)
- ALECO (engl. *Algorithm for Learning Efficiently with Constrained Optimisation*)
- RLS BP (engl. *Recursive Least Square BP*).

Svi navedeni algoritmi ekstenzija su standardnog algoritma čiju se matematička prezentacija uporabom diferencijalnog računa (Chattamwelli, 2009) navodi u nastavku rada.

Pretpostavit će se prvo jednostavna dvoslojna mreža. Ako se izlaz j-og neurona mreže označi s O_j , a stvarna (očekivana) vrijednost izlaza istog neurona sa d_j, kvadrat greške se može označiti izrazom E_j i izrazom

$$E_j = (O_j - d_j)^2 \qquad (2.23)$$

Ako se s α označi konstanta koja predstavlja koeficijent učenja, tada se prilagodba težinskih faktora odvija uporabom algoritma gradijentnog opadanja

$$\Delta w_{ji} = -\alpha \frac{\partial \mathrm{E}}{\partial w_{ji}} \qquad (2.24)$$

Ulazni težinski faktori mreže transformirani su aktivacijskom funkcijom, a potom prolaze i kroz prag funkciju kako bi dali izlaznu veličinu. Greška je stoga funkcija od funkcije težinskog faktora. Diferencijalni račun koristi pravilo ulančavanja po kome je

$$\frac{\partial}{\partial x} f(g(x)) = \left(\frac{\partial}{\partial g(x)} f(g(x)) * \left(\frac{\partial}{\partial x} g(x) \right) \right) \qquad (2.25)$$

Derivacija greške po O_j će dati:

$$\frac{\partial E}{\partial O_j} = 2(O_j - d_j) \qquad (2.26)$$

Stopa promjene izlaza u odnosu na ulaz prag funkcije, ako je prag funkcija sigmoidna funkcija, biti će:

$$\frac{\partial O_j}{\partial T_j} = O_j(1 - O_j) \qquad (2.27)$$

Ako je aktivacijska funkcija linearna funkcija (ponderirana suma funkcija), stupanj promjene u odnosu na težinske faktore je konstanta, odnosno vrijednost ulazne veličine x_j:

$$\frac{\partial T_j}{\partial w_j} = x_j \qquad (2.28)$$

Uz primjenu spomenutog pravila ulančavanja diferencijalnog računa dobije se:

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial O_j} * \frac{\partial O_j}{\partial T_j} * \frac{\partial T_j}{\partial w_j} = 2(O_j - d_j) * O_j(1 - O_j) * x_j \qquad (2.29)$$

Koristeći izraz (4.2), izvršavaju se prilagodbe svih težinskih faktora mreže uporabom izraza:

$$\Delta w_{ji} = 2\alpha(O_j - d_j) * O_j(1 - O_j) * x_j \qquad (2.30)$$

Kako standardna sigmoidna funkcija varira između 0 i 1, izraz $(1 - O_j)$ je uvijek pozitivan. Stoga će izraz $(O_j - d_j) * O_j(1 - O_j)$ biti jednak 0 kada O_j poprimi vrijednost 0,1 ili d_j. Ako je vrijednost O_j blizu 0 ili blizu 1, promatrani neuron je poprilično pobuđen i to može biti jedan od kriterija za zaustavljanje algoritma.

Uvođenjem skrivenih slojeva, mreža se usložnjava. Skriveni slojevi mreže nemaju očekivane vrijednosti izlaza stoga se prilagodba težinskih faktora na ovim slojevima odvija koristeći izračunatu grešku na narednim slojevima čvorova prolazeći dva koraka:

1. za svaki neuron skrivenog sloja pronalazi se greška

$$Err(j) = \left(\sum_k Err(k) w_{kj}\right) f'(x_j) \quad (2.31)$$

pri čemu je *Err(j)* izračunata greška na neuronu *k* i w_{kj} su težinski faktori iz neurona *k* prema neuronu *j* iz narednog sloja.

2. težinski faktori se modificiraju korištenjem izraza

$$w_{kj}(novi) = w_{kj}(stari) + \Delta w_{kj} \quad (2.32)$$

Postupak se ponavlja dok se ne ispuni kriterij za zaustavljanje algoritma.

2.6 Optimizacija neuronskih mreža

Optimizacija, uopćeno promatrajući, može se opisati kao proces u kojem se nastoji riješiti određeni problem na način da se, iz skupa mogućih rješenja, pronađe ono rješenje koje generira maksimalnu korisnost. Neuronska mreža koja se modelira i projektira za rješavanje određenog problema prolazi kroz iteracije u procesu učenja. Svaka iteracija proizvodi skup težinskih faktora mreže generirajući jedno od mogućih spomenutih rješenja. Koje od tih rješenja je odgovarajuće, odnosno koje generira maksimalnu sposobnost mreže da odgovori na postavljeni problem? Izlazne varijable neuronske mreže definiraju odgovor na ovo pitanje. Proces optimizacije neuronske mreže završava određivanjem skupa sinaptičkih težina koje će djelovanjem mreže generirati najtočniji izlazni signal.

Algoritmi učenja, prethodno opisani, u osnovi imaju zadatak da, sukladno svojim mogućnostima, dovedu do rješenja koje je optimalno s njihovog gledišta. Ipak, uvijek se postavlja pitanje da li je odgovarajući algoritam dovoljan da dovede do optimalne arhitekture mreže ili postoji dodatni prostor za optimizaciju. Vrlo često se s neuronskim mrežama kombiniraju druge metode strojnog učenja u nastojanju da se dobije odgovor na postavljeno pitanje. Istaknuto mjesto među tim metodama svakako zauzima metoda koja svoju analogiju, baš kao i metoda neuronske mreže, zasniva na prirodnim, biološkim procesima: genetički algoritam. Evolucija i prirodna selekcija, biološki su procesi koji milijunima godina rezultiraju, iz generacije u generaciju, vrstama koje su se adaptirale na svoje okruženje. Optimizacija neuronske mreže zahtjeva istu stvar, tj. odgovarajuću adaptaciju na uvjete definirane okolinom, iskazanu kroz optimalnu arhitekturu mreže.

2.6.1 Genetički algoritam

Prva istraživanja vezana za genetičke algoritme (GA) provedena su početkom druge polovice prošlog stoljeća u suradnji biologa i računalnih znanstvenika u nastojanju da se simulira proces evolucije na prvim računalima. Njemački znanstvenik Rechenberg (1965) uvodi pojam "evolucijske strategije" (njem. *Evolutionsstrategie*), kao naziv metode koju koristi u optimiziranju parametara uređaja poput zrakoplovnih krila. Polje evolucijske strategije ostalo je aktivno polje istraživanja nezavisno od istraživanja genetičkog algoritma (Back et al., 1991.). Fogel, Owens i Walsh (1966) razvijaju "evolucijsko programiranje" koje u svojim principima sadrži slučajno mutiranje i uporabu funkcije cilja.

Tih istih 1960-ih, profesor John Holland je izumio genetički algoritam. Značajan rad na njegovom razvoju uložili su i studenti i kolege profesora Hollanda sa Sveučilišta u Michiganu (Mitchell, 1999). Dok su *evolucijske strategije* i *evolucijsko programiranje* bili usmjereni na rješavanje specifičnih problema, cilj razvijenih GA bio je uopćeniji. Točnije, namjera Hollanda je bila da računalno modelira fenomen adaptacije koji je prisutan u prirodi. Konačni rezultat inventivnih istraživanja na ovom polju prezentiran je 1975. godine (Holland, 1975) kroz GA kao apstrakciju biološke evolucije.

Rad metode GA sastoji se u razvoju sukcesivnih generacija genoma koji postaju progresivno prilagođeniji okolini, odnosno funkciji cilja, dajući sve bolja i bolja rješenja modeliranog problema. Funkcija cilja u prirodi jasno je definirana sposobnošću organizma da preživi i da se reproducira, a ljudski genom je pri tome nositelj informacija. Računalno promatrajući, evolucija se simulira kroz sljedeće korake (Goldberg, 1989):

1. Identificiranje genoma i funkcije cilja.
2. Kreiranje inicijalne generacije genoma.
3. Modificiranje generacije genoma primjenom genetičkih operatora selekcije, križanja i mutacije.
4. Ponavljanje 3. koraka do momenta kada se ne postižu unapređenja novom generacijom genoma.

Genom u terminologiji GA predstavlja niz definirane veličine. Pojedinačni član ovog niza može poprimiti vrijednost 0 ili 1. Jedna generacija čini točno određeni broj genoma. Funkcija cilja je funkcija čiji ulazni parametar čine pojedinačni genomi generacije. Proizvod funkcije uz apliciranje pojedinačnog genoma je izlazna veličina funkcije. Zadatak GA je produkcija takve generacije genoma koja će u prosjeku dati minimalnu ili maksimalnu vrijednost funkcij cilja, ovisno o prirodi promatranog problema. Navedeni genetički operatori pomoću kojih se produciraju nove generacije genoma predmet su sljedećeg poglavlja.

2.6.2 Genetički operatori

Selekcija je proces analogan prirodnoj selekciji u kojoj najjače jedinke opstaju u prirodi prenoseći genetički materijal s generacije na generaciju. Broj jedinki generacije, za razliku od modeliranog prirodnog procesa, je konstantna veličina. Vjerojatnost da će jedan gen preživjeti proces selekcije upravo je proporcionalna vrijednosti funkcije cilja, odnosno bolja prilagođenost čini većom vjerojatnost postojanja u narednoj generaciji genoma. Vjerojatnost je definirana relativnim udjelom u ukupnoj vrijednosti funkcije cilja. Prosti prikaz jedne generacije genoma, funkcije cilja i vjerovatnoće pojavljivanja određenog genoma nakon procesa selekcije prikazan je u tablici xx, a preuzet je iz primjera prezentiranog od strane Berry i Linoff-a (2004). Generacija se sastoji od 4 genoma 10110, 00011, 00010 i 1101, koji predstavljaju binarnu prezentaciju brojeva 22, 3, 2 i 25. Funkcija cilja je parabola predstavljena izrazom:

$$f(p) = 31p - p^2, 0 \leq p \leq 31 \qquad (2.33)$$

Tablica 2.1 Izračun parametara rada genetičkog algoritma

Genom	Funkcija cilja	% očekivane funkcije cilja	Broj kopija genoma u narednoj generaciji	Rasponi za generiranje
1	2	3	4	5
10110	198	40,41%	1,62	0 - 0,4041
00011	84	17,14%	0,69	0,4041 - 0,5755
00010	58	11,84%	0,47	0,5755 - 0,6839
11001	150	30,61%	1,22	0,6839 - 1

Prilagođena iz (Berry, Linoff, 2004)

Relativni udio (kolona 2) određuje broj kopija svakog genoma (kolona 3) koji se očekuje u narednoj generaciji proizvedenoj procesom selekcije. Na koji način? Generator slučajnih brojeva između 0 i 1 generira onoliko brojeva koliko generacija ima genoma. Generirani broj se smješta u jedan od promatranih raspona (kolona 5) čime se određuje genom koji nastavlja svoj životni put u narednoj generaciji. Proces selekcije ovako definiran osigurava „preživljavanje“ i onih genoma koji nisu potpuno prilagođeni okruženju.

Križanje je druga vrsta genetičke operacije. Njome se kombiniraju dva genoma na način da zamijene dijelove genoma tvoreći dva potpuno nova. Postupak započinje generiranjem slučajne pozicije križanja, a potom prvi dio jednog genoma mijenja sadržaj s istim dijelom drugog genoma. Nakon križanja ponovo se računa funkcija cilja za svaki pojedinačni genom i izračunava njena zbrojna vrijednost kako bi se utvrdilo da li je ovaj korak proizvео unapređenje modela. Simulacija procesa križanja prikazanog primjera odrađena je na genomima 11001 i 00010. Pozicija križanja definirana slučajnim odabirom je između trećeg i četvrtog gena, pa se nakon procesa križanja dobiju novi genomi 11010 i 00001.

$$110 : 01 \rightarrow \mathbf{000}01 \qquad (2.34)$$

$$000 : 10 \rightarrow \mathbf{110}10 \qquad (2.35)$$

Mutacija je konačna operacija procesa GA. Mutacijom se mijenja vrijednost jednog gena unutar genoma u njemu suprotnu vrijednost (iz 0 u 1 i obrnuto). Njena je pojava dosta rijetka u prirodi i javlja se uglavnom kao pogreška u procesu prijenosa genetičkog materijala između populacija. Nije izvjesno kako će rezultirati te se stoga dosta rijetko i koristi u provedbi ovog algoritma, uglavnom se radi o maksimalno jednoj mutaciji po generaciji. Simulacija procesa mutacije prikazaće se na genomu 00001 pri čemu će mutirati treći gen.

$$00\mathbf{0}01 \rightarrow 00\mathbf{1}01 \qquad (2.36)$$

Baš kao i kod prethodne operacije, nakon ove se također vrši ponovni proces računanja funkcije cilja i valorizacije ukupnog modela.

3. MULTIVARIJACIJSKA ANALIZA

Znanstvena istraživanja društvenih i fizičkih fenomena u osnovi su iterativan proces. Pretpostavke istraživanja dokazuju se prikupljanjem i analizom podataka povezanih s promatranim fenomenom. Tijekom ovog procesa često se postojeće varijable isključuju iz generiranih modela, a nove se dodaju. Kompleksnost modela kojima se prezentira realni svijet je osobina koja zahtijeva prikupljanje i obradu velikog broja varijabli. Multivarijacijska analiza je metodološka osnovica koja uključuje istovremenu analizu više varijabli (Johnson i Wicher, 2014), a skup statističkih metoda koji služi u ovu svrhu nazvan je multivarijacijska statistička analiza.

Većina metoda multivarijacijske statističke analize u svojoj osnovi ima model vjerovatnoće poznat kao multivarijantna normalna distribucija (Rencher i Christensen, 2012).

Kompleksnost metoda nužno zahtijeva njihovu implementaciju isključivo upotrebom računala i za ovu svrhu izgrađenih specijaliziranih softverskih alatki. Primjer ovakvih softverskih alata su i programski paketi SPSS i Weka koji su korišteni u ovom istraživanju.

Johnson i Wichern (2014) definiraju sljedeće ciljeve znanstvenih istraživanja upotrebom metoda multivarijacijske analize:

1. Redukcija podataka i strukturno simplificiranje
2. Sortiranje i grupiranje
3. Istraživanje zavisnosti među varijablama
4. Predviđanje
5. Konstrukcija hipoteza i njihovo testiranje.

Redukcijom podataka nastoje se modeli istraživanja maksimalno pojednostaviti ne izostavljajući značajne informacije modela, a u cilju lakše interpretacije modela. *Sortiranjem i grupiranjem* formiraju se grupe sličnih objekata na osnovama sličnosti podatka bez postavljanja uvjeta grupiranja ili uz prethodno postavljanje nekih specifičnih uvjeta. *Međuovisnost varijabli* značajna je sa stanovišta promatranih modela, a pokazuje smjer i intenzitet ovih relacija bilo da se radi o ovisnosti svih varijabli uzajamno ili pojedinačnih

međusobno. Kroz *predviđanje* se pronalaze i definiraju relacije među grupama varijabli na način da se formira model kojim se determinira vrijednost zavisnih varijabli na osnovama vrijednosti zavisnih varijabli promatranog fenomena. Na kraju, parametri multivarijacijskog seta podataka čine osnovu za *postavljanje statističkih istraživačkih hipoteza* te za njihovo dokazivanje ili opovrgavanje.

Sva dosadašnja razmatranja jasno su ukazala na činjenicu da multivarijantna analiza uključuje promatranje i analizu fenomena koji su definirani s brojem varijabli koji je veći od jedan. Veliki broj varijabli, te promatranje fenomena koji uključuju veliku populaciju nužno su nametnule standardizaciju u organizaciji i notifikaciji podataka. Uobičajeno je označavanje da mjeru *k*-te varijable u j-oj iteraciji podataka definiramo notacijom $\mathbf{x_{jk}}$. Kompletan skup podataka definiran s *p* varijabli i *n* serija podataka predstavljamo matricom **X** (3.1), u kojoj x_{jk} predstavlja numeričku vrijednost varijable *k* u seriji podataka *j*.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \quad (3.1)$$

Matrična prezentacija pojednostavila je prikaz velikog skupa podataka. Ipak, izvlačenje zaključaka o pojedinim varijablama iz ovako velikih skupova podataka izloženo je velikim poteškoćama. Stoga se u multivarijacijskim analizama pribjegava kalkuliranju određenih izvedenih veličina za svaku od varijabli promatranog modela. Ovakve varijable sadrže dosta informacija i daju mnogo jasniju sliku o sveukupnom skupu vrijednosti pojedine od promatranih varijabli. Ove varijable nazivaju se deskriptivnom statistikom modela, odnosno deskriptivnim statističkim varijablama. Uobičajeno su to varijable koje mjere lokaciju, varijaciju i linearnu asocijaciju.

Mjera lokacije je *aritmetička sredina* (engl. *Mean*) pojedine varijable promatranog modela. Ukoliko se promatra nepotpun skup vrijednosti za pojedinačnu varijablu i računa aritmetička sredina tog skupa, ovakvu mjeru nazivamo aritmetičkom sredinom uzorka (engl. *Sample Mean*). Generalno promatrajući, sredinu možemo računati iz *n* mjera za svih *p* varijabli modela upotrebom sljedeće jednadžbe:

$$\bar{x}_k = \frac{1}{n}\sum_{j=1}^{n} x_{jk} \quad k = 1,2,\dots,p \quad (3.2)$$

Mjera varijacije predstavlja srednje odstupanje vrijednosti promatrane varijable od njene srednje vrijednosti. Naziva se *varijancom* i računa jednadžbom (3.3).

$$s_{kk} = \frac{1}{n}\sum_{j=1}^{n}(x_{jk} - \bar{x}_k)^2 \quad k = 1,2,\dots,p \quad (3.3)$$

Kvadratni korijen vrijednosti varijance je mjera koja se naziva *standardna devijacija* i izražava se u istim jedinicama mjere kao i promatrane varijable.

Sljedeća zanimljiva mjera na promatranom modelu jeste mjera koja definira linearnu povezanost biilo koje dvije varijable modela. Mjera se naziva kovarijanca i uz pretpostavku da promatramo varijable x_i i $x_{k,}$ predstavlja srednju mjeru produkta odstupanja vrijednosti varijabla od njihovih srednjih vrijednosti, a definira je sljedeća jednadža:

$$s_{ik} = \frac{1}{n}\sum_{j=1}^{n}\left(x_{ji} - \bar{x}_i\right)\left(x_{jk} - \bar{x}_k\right) \quad i = 1,2,\dots,p \quad k = 1,2,\dots,p \ (3.4)$$

Na kraju ovog razmatranja mjera multivarijacijske analize, svakako je potrebno definirati i *koeficijent korelacije r* koji opisuje linearnu vezu između bilo koje dvije varijable promatranog modela. Mjera ne zavisi od jedinica mjere promatranih varijabli i pod pretpostavkom da promatramo odnos između varijabli *i* i *k* definirana je jednadžbom (3.5):

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} \quad i = 1,2,\dots,p \quad k = 1,2,\dots,p \quad (3.5)$$

Vrijednost korelacijskog koeficijenat je u granicima između -1 i 1. Ukoliko je vrijednost koeficijenta $r_{ik}=0$, ne postoji linearna veza između varijabli x_i i x_k. U suprotnom postoji linearna veza između ovih varijabli, a predznak koeficijenta i njegova veličina određuju smjer i intenzitet ove veze.

3.1. Metode i tehnike multivarijacijske analize

Sve tehnike multivarijacijske analize podataka nastale su, uglavnom, iz njima srodnih tehnika univarijantne i bivarijantne statistčke analize. U nastojanju da klasificiraju ove tehnike i na njima zasnovane metode, Hair et al. (2009) se služe specificiranjem ciljeva istraživanja i tipom varijabli. Specificiranje cilja istraživanja podrazumijeva odgovor na pitanje da li promatrani problem uključuje relaciju zavisnosti, odnosno jasnu podjelu varijabli modela na nezavisne i zavisne ili se radi o problemu kod kojeg su sve varijable nezavisne i promatra se njihova međusobna relacija. Hair et al. postavljaju tri pitanja i sukladno dobivenim odgovorima (tablica 3.1) dobivaju sljedeću podjelu metoda:

- Modeliranje strukturnih jednadžbi SEM (eng. *Structured equations modeling*)
- Kanonička korelacijska analiza (engl. *Cannonical correlation analysis*)
- Kanonička korelacijska analiza s *dummy* varijablama (engl. *Cannonical correlation analysis with dummy variables*)
- Multivarijacijska analiza varijance (engl. *Multivariate analysis of variance*)
- Višestruka regresija (eng. *Multiple regression*)
- Conjoint analiza (engl. *Conjoint analysis*)
- Višestruka diskriminantna analiza (engl. *Multiple discriminant analysis*)
- Linearni modeli vjerojatnosti (engl. *Linear probability models*)
- Faktorska analiza (engl. *Factor analysis*)
- Klasterska analiza (engl. *Cluster analysis*)
- Multidimenzijsko skaliranje (engl. *Multidimensional scaling*)

- Korespodentna analiza (engl. *Correspodence analysis*).

Tablica 3.1 Metode multivarijantne analize sukladno tipu, broju varijabli i mjernoj skali

	Podjela na zavisne i nezavisne varijable?	Broj zavisnih varijabli?	Skala mjerenja varijabli?
Modeliranje strukturnih jednadžbi	Da	>1	
Kanonička korelacijska analiza	Da	>1	metrička
Multivarijacijska analiza varijance	Da	>1	nemetrička
Kanonička korelacijska analiza s dummy varijablama	Da	>1	nemetrička
Višestruka regresija	Da	1	metrička
Conjoint analiza	Da	1	metrička
Višestruka diskriminantna analiza	Da	1	nemetrička
Linearni modeli vjerojatnoće	Da	1	nemetrička
Faktorska analiza	Ne	0	
Konfirmatorna faktorska analiza	Ne	0	
Klaster analiza	Ne	0	
Multidimenzijsko skaliranje	Ne	0	(ne)metrička
Korespodentna analiza	Ne	0	nemetrička

Hair et al. (2009) definiraju skup od deset tehnika multivarijacijske analize na kojima se zasniva djelovanje svih metoda ove analize. Sve promatraju više, međusobno povezanih, varijabli pri čemu je svaka varijabla podjednako bitna, te zahtijevaju korištenje prikladnih softvera. Budući da kroz istraživanje dolazi do iterativnog procesa učenja tijekom kojeg se varijable dodaju ili isključuju iz analize, ovisno o dobivenim rezultatima, analiza je dobila atribut multivarijantna. Ove tehnike su:

i. Faktorska analiza
ii. Višestruka regresija
iii. Višestruka diskriminacijska analiza i logistička regresija
iv. Kanonička korelacija
v. Višestruka analiza varijance i kovarijance
vi. Conjoint analiza
vii. Klaster analiza
viii. Perceptualno mapiranje
ix. Korespodentska analiza
x. Modeliranje strukturnih varijabli i konfirmatorna faktorska analiza

Faktorska analiza uključuje analizu glavnih komponenti i prostu faktorsku analizu. Prilagođena je analiziranju međuovisnosti velikog broja opaženih varijabli. Osnovni cilj joj je reducirati veliki broj opaženih varijabli i pojednostaviti objašnjavanje promatranog fenomena. Postupak reduciranja generira manji broj latentnih varijabli, tzv.faktora. Faktorska analiza izvodi se iz korelacijske matrice opaženih varijabli, a iz dobivenih faktora moguće je generirati originalnu korelacijsku matricu opaženih varijabli (Rozga, 2010).

Višestruka regresija je tehnika koja uključuje jednu metričku zavisnu varijablu i više od jedne nezavisne metričke varijable. Cilj tehnike ogleda se u predviđanju promjena zavisne varijable kao reakciju na promjene u nezavisnim varijablama. Primjenu pronalazi u aplikacijama predviđanja zavisne varijable, baš kao što je i slučaj s ovim istraživanjem.

Višestruka diskriminacijska analiza operira s jednom, ne metričkom zavisnom varijablom koja može imati dvije (dihotomna) ili više od dvije moguće vrijednosti. Nezavisne varijable su metričke. Metoda dijeli, sukladno vrijednosti zavisne varijable, cjelokupan skup podatka na nekoliko grupa. Cilj je razumjeti razlike među njima procjenjujući vjerojatnost da pojedina instanca nezavisnih varijabli pripada određenoj grupi definiranoj ciljnom zavisnom varijablom modela.

Logistička regresija je tehnika koja uključuje elemente prethodno spomenutih tehnika višestruke regresije i diskriminacijske analize. Operira nad modelima koji uključuju više nezavisnih varijabli koji definiraju samo jednu zavisnu varijablu. Zavisna varijabla je ne metrička i to je razlikuje od višestruke regresije, dok su nezavisne varijable i metričke i ne metričke čime se razlikuje od diskriminacijske analize. Logistička regresija je primjerena fenomenima kod kojih je zavisna varijabla dihotomna, dok je u ostalim slučajevima primjerenija diskriminacijska analiza.

Kanonička korelacijska analiza je tehnika koja pokušava definirati relacije u fenomenima koji uključuju isključivo metričke varijable pri čemu je moguće imati više od jedne zavisne varijable. Na određeni način, ona proširuje tehniku višestruke regresije s ciljem simultanog koreliranja grupe od nekoliko nezavisnih i zavisnih varijabli. Pri tome se nastoji ostvariti maksimalna korelacija između seta nezavisnih i seta zavisnih varijabli.

Multivarijacijska analiza varijance promatra modele s nekoliko kategoričkih nezavisnih varijabli i dvije ili više metričke zavisne varijable, nastojeći simultano istražiti

relaciju među njima. *Multivarijacijska analiza kovarijance* se uglavnom koristi na rezultatima MANOVA-e kako bi uklonila učinke bilo koje nekontrolirane metričke nezavisne varijable na zavisne varijable modela.

Conjoint analiza je nastala na praksi istraživanja tržišta. Cilje tehnike je evaluacija instance objekta (npr. proizvod ili servis) na osnovi različitih vrijednosti njegovih atributa. Rezultat analize uobičajeno je kombinacija limitiranog broja atributa koja najviše utječe na donošenje odluke na tržištu. Na osnovu ovih rezultata kreiraju se marketinški modeli kojima se predviđaju pokazatelji tržišnog udjela i prihoda instanci promatranog objekta.

Klasterska analiza svrstava objekte u manji broj međusobno isključivih grupa (klastera) na osnovu sličnih obilježja promatranih entiteta. Prije analize nije poznata pripadnost pojedinog entiteta klasteru, baš kao ni broj grupa. Uključuje uobičajeno tri koraka. U prvom se mjeri sličnost među entitetima kako bi se utvrdio mogući broj klastera. Drugi korak uključuje podjelu entiteta u klastere, dok treći uključuje profilaciju varijabli. Treći korak vrlo često zapošljava metodu diskriminacijske analize za postizanje svog cilja.

Perceptualno mapiranje je tehnika koja prevodi preferencije potrošača u udaljenosti predstavljene u višedimenzijskom prostoru. Iz ovog razloga se tehnika često naziva i tehnikom višedimenzijskog skaliranja. Udaljenosti među entitetima koji imaju iste preferencije su manje nego među entitetima s različitim preferencijama. Na ovaj način se u višedimenzionalnom prostoru stvaraju skupovi entiteta. Njihove relativne pozicije podložne su daljnjim analizama kako bi se opisale relacije među atributima preferencija i stvorila podloga za predviđanje odgovarajuće pozicije.

Baš kao i prethodna tehnika, *korespodentska analiza* je tehnika koja uključuje perceptualno mapiranje objekata ali na skupu nenumeričkih podataka. Nenumerički podaci odnose se uglavnom na kvalitativne atribute. Kontigencijskom tablicom se vrijednosti ovih varijabli kvantificiraju, a potom se na njih primjenjuju metode dimenzijske redukcije i perceptualnog mapiranja. Tehnika osigurava moćan alat za promatranje fenomena definiranih nelinearnim odnosima među kvantitativnim i kvalitativnim varijablama i analizu njihove međuovisnosti.

Metoda strukturnih varijabli je tehnika predviđanja koja promatra modele definirane skupom jednadžbi višestruke regresije. Osigurava način na koji se definira koje nezavisne

varijable određuju pojedinu zavisnu varijablu. Isto tako moguće je povezati više nezavisnih varijabli prema srodnosti u jednu, novu latentnu varijablu.

3.2. Metode regresija

Za potrebe empirijskog istraživanja ove disertacije korištene su dvije metode multivarijantne statističke analize: višestruka linearna regresija i logistička regresija. Obje metode su regresijske metode, a u nastavku će se prikazati njihove osnovne značajke.

3.2.1 Višestruka linearna regresija

Regresijski modeli koji zapošljavaju više od jedne nezavisne varijable se nazivaju višestruki regresijski modeli (Bowerman, O'Connell, Murphree, 2014).

Opći oblik modela višestruke regresije može biti prikazan (Bahovec, Erjavec, 2009) jednim od izraza:

$$y = f(x_1, x_2, \dots, x_i, \dots, x_k) + \varepsilon \quad (3.6)$$

Ili

$$y = f(x_1, x_2, \dots, x_i, \dots, x_k) * \varepsilon \quad (3.7)$$

U jednadžbama (3.6) i (3.7) *y* predstavlja zavisnu varijablu modeliranog fenomena, dok x_1, ..., x_k predstavljaju nezavisne varijable pomoću kojih se izražavaju varijacije zavisne varijable *y*. Ɛ predstavlja „nepoznata odstupanja od funkcionalnog odnosa“.

Ukoliko pretpostavimo linearnu relaciju između zavisne i nezavisnih varijabli, model se naziva modelom višestruke linearne regresije i izražava se sljedećom jednadžbom:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j x_j + \cdots \beta_k x_k + \varepsilon \quad (3.8)$$

Varijabla y u jednadžbi (3.8) je zavisna varijabla, a naziva se još regresand, endogena ili output varijabla. Varijable x_1, x_2, ..., x_k nazivaju se nezavisnim, egzogenim ili input varijablama. $\beta_0, \beta_1, \ldots, \beta_k$ predstavljaju nepoznate parametre dok je Ɛ slučajna varijabla ili tzv. greška relacije.

Bahovec i Erjavec (2009) *procijenjeni regresijski model* definiraju jednadžbom (3.9):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_j x_j + \cdots \hat{\beta}_k x_k \quad (3.9)$$

U jednadžbi (3.9) $\hat{\beta}_j$ je procijenjeni parametar definiran parcijalnom derivacijom $\frac{\delta \hat{y}}{\delta x j}$. Ovaj parametar se stoga interpretira kao promjena regresijske vrijednosti zavisne varijable y za jedinični porast nezavisne varijable x_j uz pretpostavku da su ostale nezavisne varijable nepromijenjene.

Ukoliko se procijenjene regresijske vrijednosti definiraju izrazom (3.10)

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_j x_j + \cdots \hat{\beta}_k x_k \quad i = 1,2, \ldots, n \quad (3.10),$$

tada se *rezidualna odstupanja* $\hat{e}_i$ definiraju kao razlika između empirijskih vrijednosti i procijenjenih vrijednosti zavisne varijable, što se može predstaviti jednadžbom (3.11).

$$\hat{e}_i = y_i - \hat{y}_i \quad (3.11)$$

Prema Bahovec i Erjavec (2009), ukupna suma kvadrata odstupanja ST dijeli se na dio sume kvadrata odstupanja regresijskih vrijednosti od prosjeka i na dio sume kvadrata odstupanja regresijskih od opaženih vrijednosti. Prva suma kvadrata je protumačena suma kvadrata (SP), a druga je rezidualna suma kvadrata (SR), pa se jednadžba analize varijance može definirati sljedećom jednadžbom:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = (\hat{y}_i - \bar{y})^2 + (y_i - \hat{y}_i)^2 \quad (3.12)$$

Uobičajena je *analiza varijance* (ANOVA) u modelima višestruke linearne regresije. Osnovni oblik ANOVA tablice (tablica 3.2) sadrži podatke o stupnjevima slobode, sumama i prosjecima kvadrata odstupanja, te F-omjeru.

Tablica 3.2 Tablica ANOVA

Izvor varijacije	**Stupnjevi slobode DF**	**Sume kvadrata SS**	**Sredine kvadrata MS**	**F-omjer**	**PROB>F**
Protumačen modelom	k	SP	$SP/_k$	$\frac{SP/_k}{SR/_{(n-(k+1))}}$	
Ne protumačena odstupanja	n-(k+1)	SR	$\sigma^2 = SR/_{(n-(k+1))}$		
Ukupno	n-1	ST			

Preuzeto iz (Bahovec, Erjavec, 2009)

Procijenjena varijanca regresije $\boldsymbol{\sigma^2}$ je količnik rezidualne sume kvadrata odstupanja i broja stupnjeva slobode za ne protumačena odstupanja. Ukoliko se korjenuje ovaj izraz te uzme njegova pozitivna vrijednost dobije se procjena standardne devijacije regresije. To je vrijednost koja iskazuje prosječno odstupanje empirijskih od regresijskih vrijednosti. Kod procjene generiranog modela regresije, vrijednu informaciju daje koeficijent determinacije (3.13). On poprima vrijednosti iz segmenta [0,1]. Vrijednosti bliže jedinici ukazuju na regresijski model koji dobro opisuje modelirani fenomen i obratno.

$$R^2 = \frac{SP}{ST} = 1 - \frac{SR}{ST} = 1 - \frac{\lfloor n - (k+1) \rfloor \hat{\sigma}^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (3.13)$$

Postupke *tesitranja hipoteza* kojima se utvrđuje postojanje linearne zaviosnoti između zavisne varijable *y* i nezavisnih varijabli x_k, moguće je svrstati u pojedinačne i skupne testove (Bahovec, Erjavec, 2009):

a. Test o značajnosti jedne regresorske varijable (pojedinačni test)
b. Test o značajnosti svih regresijskih varijabli (skupni test)
c. Test o značajnosti podskupa regresorskih varijabli (skupni test).

Test o značajnosti jedne regresorske varijable najčešće je provođen test u regresijskoj analizi. Hipoteze ovog testa mogu se predstaviti na sljedeći način:

$$H_0: \beta_j = 0 \quad H_0: \beta_j = 0$$
$$(3.14)$$
$$H_1: \beta_j > 0 \quad H_1: \beta_j < 0$$

Uz pretpostavljenu razinu signifikantnosti α hipoteza H_0 će se odbaciti pod uvjetom da je $t_j > t_\alpha$ kod testiranja na gornju granicu. Isti zaključak vrijedi ako je $t_j < -t_\alpha$ pri testiranju na donju granicu. Pokazuju li testovi da podaci nisu u skladu s odabranim modelom, model je neprihvatljiv i treba ga zamijeniti novim modelom.

Test o značajnosti svih regresorskih varijabli definiran je nultom hipotezom H_0, koja pretpostavlja da ne postoji utjecaja niti jedne od regresorskih varijabli na promjenu zavisne varijable. Odbacivanje nulte hipoteze znači da je bar jedna od nezavisnih varijabli značajna, odnosno da značajno doprinosi modelu. Hipoteze ovog testa prikazane su matematičkim iskazom (3.15):

$$H_0: \beta_j = \beta_2 = \cdots = \beta_k = 0$$
$$(3.15)$$
$$H_1: \exists \beta_j \neq 0, \quad j = 1,2, \ldots, k$$

3.2.2 Logistička regresija

Višestruka linearna regresija se koristi u promatranju realnih modela opisanih varijablama čije su vrijednosti kontinualne ili metričke. Česta je pojava dihotomne zavisne varijable, npr. ženski ili muški spol, kupnja proizvoda X ili proizvoda Y, student je izašao na ispit ili ne, student je položio ispit ili ne.

Logistička regresija je korisna prilikom rješavanja problema gdje su u pitanju demografske varijable budući da su one većinom kategoričke. Posebno je uspješna ukoliko je „*kategorička varijabla kao zavisna promjenljiva sa jakom asimetrijom ili ako ima nelinearnu relaciju sa ostalim varijablama*“. Moguće rješenje problema ove vrste uobičajeno se rješavaju bilježenjem vrijednosti ovih varijabli brojevima 0 i 1 čime se dobiva regresijski model, međutim logit regresija je bolja alternativa.

Primjerice, želja je da se zna pripadnost svakog studenta pojedinoj, unaprijed određenoj, grupi. Korištenjem višestruke linearne regresije dobiti će se rješenje gdje zavisna promjeljiva varijabla imati vrijednost između 0 i 1 koja će se ogledati kao vjerojatnost da student pripada jednoj ili drugoj grupi.

Na primjer, ako je sa „0“ označen slučaj pripadnosti studenta grupi „nije položio“, a sa „1“ grupi „položio“, a vrijednost zavisne promjenljive (grupe) iznosi 0,80 za danog studenta, ispada da je veća vjerojatnost da će student pripasti u grupu „položio“ jer je vrijednost bliža jedinici. Problem koji se javlja je rezultat koji je manji od 0 ili pak veći od 1, iz čega se zaključuje kako model višestruke linearne regresije ne predstavlja sretno rješenje, stoga se nameće logistička regresija.

Često su socio-ekonomske varijable kategoričke, a ne nalaze se u nekim intervalnim skalama. Ukoliko se radi o određenom istraživanju gdje zavisna varijabla može poprimiti vrijednosti „uspješno“ / „neuspješno“, istraživača može zanimati kako je spomenuta varijabla povezana sa spolom, godinama, i sl. Višestruka linearna agresija ne bi bila podoban model za rješavanje ovog upita nego bi se problem riješio korištenjem logističke regresije.

Nezavisne varijable modela logističke regresije mogu biti dihotomne (npr. spol), nominalne s više modaliteta (npr. školska sprema), ali i numeričke. Numeričke se najčešće svode na kategorijalne varijable formiranjem kategorija temeljenih na pojedinim rasponima vrijednosti numeričke varijable.

Osnovni model logističke regresije izgleda kako slijedi (Rozga, 2010):

$$\ln\left(\frac{P(A)}{1-P(A)}\right) = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + ... + \beta_k \cdot X_k \quad (3.16)$$

Vrijednost s lijeve strane jednadžbe, uobičajeno nazivana logit vrijednost po kojoj je metoda i dobila ime, predstavlja prirodni logaritam omjera šansi (engl. Odds ratio). U brojniku količnika je vjerojatnost da će zavisna varijabla poprimiti jednu (od dvije dihotomne vrijednosti zavisne varijable) vrijednost, a u nazivniku je vjerojatnost suprotnog događaja. Ukoliko se moguće vrijednosti, a što je uobičajeno, kodiraju kao 1 i 0, onda se vjerojatnost iz brojnika odnosi na vrijednost prediktorske varijable predstavljene s 1, a one iz nazivnika na vrijednost prediktorske varijable predstavljene vrijednošću 0. Vrijednost se logita interpretira kao indikator koliko je puta veća vjerojatnoća da će se dogoditi događaj uzrokovan vrijednošću prediktorske varijable 1 umjesto istog događaja uzrokovanog prediktorskom varijablom s vrijednošću 0. Ovo je dakako bio primjer interpretacije s nezavisnom varijablom koja ima dihotomnu vrijednost. Ukoliko se radi o kategorijalnoj varijabli s više od dvije vrijednosti, tada se jedna vrijednost varijable uzima kao referencijalna, te se svi omjeri šansi iskazuju u odnosu na nju.

Logistička regresija, na osnovu navedenog, proizvodi regresijski model koji za rezultat ima vjerojatnost da pojedina n-torka ulaznog skupa nezavisnih varijabli pripada jednoj od dvije grupe događaja definiranih vrijednostima dihotomne zavisne varijable. Ovo svojstvo svrstava logit metodu u grupu klasifikacijskih modela kojima se jedinice iz uzorka svrstavaju u grupe.

Uobičajeno je kod primjene binarne logističke regresije, pored uobičajenih mjera validnosti generiranog modela, analizirati ROC krivulju. ROC krivulja ili točnije rečeno površina ispod ROC krivulje, predstavlja mjeru diskriminacije promatranog modela. Ona mjeri sposobnost modela da ispravno klasificira instance podataka na moguće klasifikacije. ROC vrijednost poprima vrijednost između 0 i 1. U literaturi se ovako ocjenjuju postotak površine ispod ROC krivulje:

- 0.90 – 1.00 izvrstan,

- 0.80 – 0.90 vrlo dobar,
- 0.70 – 0.80 dobar,
- 0.60 – 0.70 dovoljan,
- 0.50 – 0.60 nedovoljan,
- 0.00 – 0.50 beskoristan model.

Površina ispod ROC krivulje ne može nezavisno tumačiti model. Svakako je jako bitno obratiti pozornost na značajnost pojedinačnih varijabli logističke regresije, te i na sve druge pokazatelje statističke dijagnostike (Rozga, 2010).

3.3 Ograničenja metoda multivarijacijske analize

Tehnike multivarijacijske statističke analize su kompleksne. Uključuju uglavnom upotrebu složenih matematičkih postupaka, a samim tim i neophodnost korištenja statističkih softverskih paketa za analizu podataka. Iako je na tržištu čitav niz besplatnih, tzv. *open source* softverskih rješenja, za potpunu analizu vrlo često su neophodni komercijalni paketi koji su uglavnom skupi. Rad s ovim paketima je dosta jednostavan, ali interpretacija rezultata zahtijeva dobro poznavanje statistike i statističkih metoda.

Rezultati dobivenih analiza nisu uvijek jednostavni za interpretaciju i zasnivaju se na pretpostavkama koje je teško procijeniti. Za generiranje modela koji daju značajne rezultate potreban je veliki uzorak podataka što često nije moguće ostvariti kao preduvjet istraživačkom radu. S manjim setom podataka rezultat modela podložan je većoj standardnoj greški, a što vodi manjem povjerenju u model u odnosu na modele generirane s većim setovima podatka.

Kompleksnost u primjeni i nužnost dobrog poznavanja teorije statističke znanosti pri interpretaciji rezultata, stvorilo je dobru podlogu za razvoj novih tehnika kojima su se izvlačili zaključci iz analiziranih setova podataka. Česte su bile rasprave o valjanosti pristupa i ciljevima ovih metoda. Breiman (2001) ide toliko daleko da definira postojanje dvije kulture u statističkom modeliranju podataka. Jedna pretpostavlja da su podaci generirani danim stohastičkim modelima podatka, dok druga koristi algoritamske metode i pretpostavlja da su mehanizmi podatka nepoznati. Prvom pristupu je gotovo ekskluzivno posvećena statistička

zajednica, a takvim pristupom zajednicu su doveli do upitnih zaključaka i ograničenja rada na mnogim aktualnim problemima današnjice koji se nastoje objasniti različitim metodama analize podataka.

S druge strane, jedan dio statističke zajednice i zajednica znanstvenika s područja računalnih znanosti, razvijaju algoritamske metode u teoriji i praksi profilirajući novo polje znanosti izvan statistike. Strojno učenje može jednako dobro raditi na velikim i malim setovima podataka. Osnovni cilj novih metoda i tehnika je korištenje podataka u svrhu rješavanja i analize fenomena koji su predmet izučavanja. Metode statističke analize, zasnovane na modeliranju podataka i izučavanju njihove međuovisnosti, nisu suprotstavljene metodama strojnog učenja zasnovanim na algoritamskom modeliranju problema. Postoji prostor za obje grupe metoda, a namjera ovog istraživanja je da ih suprotstavi i utvrdi koja od metoda je bolja za korištenje u konkretnom problemu predviđanja profitabilnosti kupaca.

4. PROFITABILNOST KUPCA

4.1 Povijest analize profitabilnosti kupca

Analiza profitabilnosti kupaca , CPA (engl. Customer Profitability Analysis), podrazumijeva suvremene trendove izučavanja profitabilnosti poslovanja mjerene na razini kupca. Ovaj pristup je proizašao iz promatranja profitabilnosti na razini proizvoda/usluge (troškovi proizvoda i veličina prodaje), u periodu nakon 1980-ih kada je zadovoljstvo kupca bilo u prvom planu. Protekle godine od početka stoljeća obilježene su, u poslovnom okruženju, generiranjem ogromnih količina podataka pohranjenih u bazama i skladištima podataka, intenzivnom ICT podrškom poslovnim subjektima, te povećanom razinom primjene metoda strojnog učenja nad tim podacima (Ngai et al. 2009, Sonnenburg et al. 2010). Analiza profitabilnosti tržišnih subjekata prebacuje fokus s razine proizvoda i usluga na pojedinačnog kupca (Ngai, 2005). Analizom poslovanja uočeno je da troškovi proizvoda čine samo jedan dio troškova poslovnog odnosa na relaciji poduzeće-kupac. On je opterećen i različitim troškovima koje generiraju poslovne aktivnosti poput interne logistike, administriranja narudžbi kupaca, prodaje, marketinga i post-prodajnog održavanja. Kupci se, sukladno svojim karakteristikama, dosta razlikuju u alociranju ovih troškova. Uočena spoznaja svoje prve ozbiljnije teoretske analize doživljava u radovima oko 1990. godine (Shapiro et al., 1987; Howell i Soucy, 1990; Foster et al., 1996). Glavni zaključak ovih radova svodi se na činjenicu da svaka novčana jedinica prihoda ne učestvuje jednako u profitu poduzeća. Pažnja se usmjerava na pojedinačnog kupca i mjerenje „vrijednosti“ novčane jedinice prihoda nastalog u poslovanju s njim.

Spomenuta vrijednost ogleda se u razlici između prihoda i troškova nastalih poslovnim odnosom s kupcem. Dok je prihode jednostavno iskazati, troškovi predstavljaju kompleksan problem. Kalkuliranje profitabilnosti započinje oduzimanjem troškova proizvoda[1], a nastavlja se prepoznavanjem i oduzimanjem ostalih vrsta troškova koji opterećuju ovu poslovnu relaciju. Van Triest (2005) ih tako dijeli na troškove u svezi fizičke isporuke proizvoda, ciljane troškove kupca i troškove koji se ne mogu alocirati na

[1] Pod proizvodom podrazumijevamo i materijalni proizvod i nematerijalnu uslugu koju poduzeće prodaje kupcu u kupoprodajnom odnosu.

individualnog kupca. Shapiro et al. (1987) ih dijele na prije prodajne, produkcijske, distribucijske i troškove poslije prodajnog servisa.

Raznovrsnost troškova i namjera da se alociraju na pojedinačne kupce, postavlja upitnim mogućnosti klasičnih računovodstvenih metoda obračuna troškova. Pojava ABC (engl. Activity Based Costing) metode uvodi mnogo jasniju sliku u definiranju profitabilnosti kupaca identificiranjem poslovnih aktivnosti, alociranjem odgovarajućih troškova resursa, te pridruživanjem ovih aktivnosti kupcima, proizvodima ili procesima (Cookins, 2001). Uočene nedostatke i probleme u primjeni, Kaplan i Anderson (2007) nadilaze definiranjem TD-ABC metode (engl. Time Driven Activity Based Costing). Ova metoda je dugotrajne, relativno neprecizne i proračunski komplicirane procjene korištenja resursa na osnovi subjektivnih procjena iz ABC metode nadomjestila procjenom vremena korištenja kapaciteta za pojedine akcije uporabom vremenske jednadžbe (engl. Time Equation).

Izračun profitabilnosti pojedinačnih kupaca, ili pak grupa kupaca, daje osnovu za podrobniju analizu distribucije profitabilnosti kupaca unutar poduzeća. Neka empirijska istraživanja (Cooper i Kaplan, 1991; Kumar et al., 1995; Foster et al., 1996; Storbacka, 1997; Lejune, 2001; Krakhmal 2006, Kumar et al., 2006; Kujamaki, 2007) daju rezultate koji ovo područje čine veoma interesantnim sa stanovišta analize i upravljanja distribucijom kupaca. Upravljanje neprofitabilnim, ali i profitabilnim kupcima, čini prirodan slijed ovih istraživanja. Točno i precizno utvrđivanje troškova, utvrđivanje disperzije profitabilnosti čine osnovu za utvrđivanje strategije upravljanja neprofitabilnim kupcima i uspostavljanje odgovarajuće infrastrukture za njeno provođenje. Ponovno procjenjivanje odnosa s kupcima, educiranje kupaca, ponovno ugovaranje uvjeta kupovine, migracija kupaca te kao krajnja mjera terminiranje poslovnog odnosa s kupcem, samo su dio aktivnosti zahtjevnog procesa menadžmenta kupaca (Mittal et al., 2008).

Iznalaženje stvarnih pokazatelja profitabilnosti kupaca zahtijeva dizajn modela koji će to omogućiti. Model bi trebao identificirati varijable internog i eksternog okruženja bitne za izračun prihoda, troškova i profita po pojedinom kupcu. Izlazni rezultati modela bi trebale biti definirane mjere profitabilnosti kupaca. Empirijska vrijednost ovih pokazatelja je osnova za poduzimanje menadžerskih akcija usmjerenih na upravljanje kupcima i okruženjem (Ryals, 2003; Lee et al., 2009).

Opći okvir za definiranje profitabilnosti kupaca, pored čisto financijskih elemenata, uključuje i čitav niz nelinearnih i nefinancijskih elemenata koji se ne mogu obrađivati

tradicionalnim statističkim metodama. Ove pretpostavke, zajedno s osnovnim karakteristikama naprednih metoda strojnog učenja, nameću neke od njih kao metode koja će pružiti holistički pogled na profitabilnost kupaca. U osnovi, ovaj pristup je na tragu „the balanced scorecard“ principa koji ekonomsku korist postavlja kao najvažniji kriterij, ali u atraktivnost kupca uključuje i čitav niz nefinancijskih elemenata (Kaplan i Norton, 1996; Kaplan i Norton, 2005; Kaplan i Norton, 2007).

Ovakva analiza profitabilnosti, kako je do sada opisana, ipak je analiza proteklih događaja, tj. retrospektivna analiza. Menadžment poduzeća u njoj nalazi veliku vrijednost i osnovu za donošenje odluka. Jedan drugačiji pogled na profitabilnost, pogled na buduću profitabilnost kupaca, donio bi nova saznanja značajna za procese odlučivanja. Prospektivna CPA (van Raaij et al., 2003) predviđa neto sadašnju vrijednost očekivanih troškova i prihoda kupca za vrijeme njegovog budućeg životnog vijeka (kao kupca poduzeća), a osnovu za predviđanje nalazi u retrospektivnoj analizi.

Upravo metode strojnog učenja mogu ekstrahirati skriveno znanje i obrasce ponašanja kupaca iz retrospektivnih analiza. Za razliku od klasičnih statističkih metoda, koje u svoja razmatranja uzimaju linearne varijable, u obzir mogu jednako dobro uzeti nelinearne i nefinancijske elemente okruženja koji imaju utjecaj na rezultate profitabilnosti. Svojom sposobnošću da aproksimativno opišu bilo koju neprekinutu funkciju, čine dobar izbor pri odabiru metoda za predviđanje profitabilnosti kupaca.

4.2 Opći okvir profitabilnosti kupca

Za potrebe razmatranja profitabilnosti, nužno je osigurati empirijske pokazatelje poslovnog odnosa promatranog poduzeća sa svakim pojedinačnim kupcem ili grupom kupaca. Istraživanja ovog područja s kraja 1990-ih i prve dekade 2000-ih generirala su niz modela koji su se naslanjali na marketinške ili na analize upravljačkog računovodstva (Berger i Nasr, 1998; Mulhern, 1999; Niraj et al.; 2001, van Raaij et al., 2003; Krakhmal, 2006). Svakako najčešće navođen u literaturi koja se bavi ovim područjem, jeste model koji su izgradili Niraj et al. (2001). Na osnovama ABC metode obračuna troškova izgradili su model profitabilnosti za veleprodajno/distributersko poduzeće. Autori su istaknuli ograničenja

modela u linearnosti rasporeda troškova, ignoriranju varijacija u aktivnostima i njihovim pokazateljima kroz vrijeme za pojedinačnog kupca, te postojanju faktora potražnje koji se ne uzimaju u razmatranje. Ipak, od velikog je značenja primjenjivost modela na podacima koje posjeduju poduzeća ovog tipa.

Zajedničko za sve modele jeste definiranje troškova koji ulaze u izračun profitabilnosti, način njihovog rasporeda, te pregled ostalih činitelja, internog i eksternog okruženja, koji utječu na profitabilnost. Sa stanovišta ovog rada, pored troškova, u grupu ostalih činitelja pobrojani su:

- Cjenovna politika,
- Značajke poduzeća,
- Značajke kupca,
- Značajke poslovne okoline.

4.2.1 Cjenovna politika

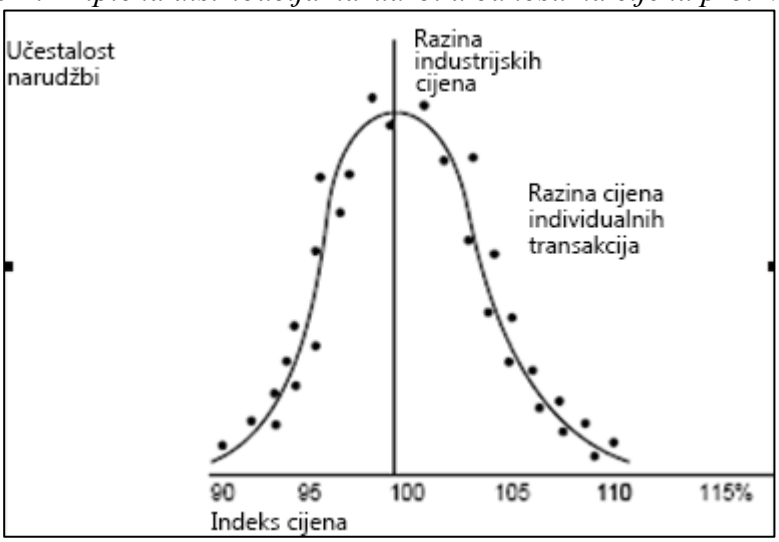
Cijena proizvoda, te s njom povezana bruto marža, čini bitan faktor profitabilnosti kupaca. Umnožak cijene proizvoda i količine prodanih proizvoda daje prihod. U izračunu profitabilnosti kupaca od ostvarenog prihoda po kupcu, oduzimaju se troškovi i dobiva se ostvareni bruto profit. Pri tome, cijena nije fiksna kategorija već se razlikuje od kupca do kupca. Razlozi su različiti (Niraj et al., 2001):

- Veće kupljene količine obično prate količinski rabati.
- Većim cijenama ponuda nastoji kompenzirati veće troškove karakteristične za pojedine kupce.
- Razlika u cijenama može biti uzrokovana i tržišnim konkurentskim razlozima pri čemu se nude niže cijene radi perspektiva rasta prodaje na duži rok ili pak veće cijene radi destimuliranja kupnje određenih skupina kupaca koji čine neprofitabilnu grupu kupaca.

Iz navedenih razlika u cijenama može se zaključiti da veća količina prodaje ne znači uvijek i veći profit. Naprotiv, mnoštvo faktora utječe na konačni profit. Iako veliki broj istraživanja pažnju posvećuje troškovnoj strani, pravilan odabir cijena jako je bitan. Ross

(1984) ističe da mnoge kompanije određuju cijene instinktivno na osnovi jednostavnih kriterija poput pokrića troškova, zadržavanja ili osvajanja tržišnog udjela ili pak usklađivanjem s konkurencijom. Krivulja raspodjele cijena varira od industrije do industrije, ali se oblikom približava izgledu krivulje normalne distribucije i ukazuje na razlike od oko 10 postotnih poena s obje strane.

Slika 4.1 Tipična distribucija narudžbi u odnosu na cijenu proizvoda



Preuzeto iz (Ross, 1984)

Ovakva razlika u cijenama rezultat je i faktora nesavršenosti tržišta sa obje strane: ponude i potražnje. Sa strane potražnje ogleda se u inertnosti kupaca, povećanim troškovima prijelaza na drugog dobavljača, relativnoj važnosti cijene ili pak navikama u kupovini. Ponuda pak ima svoje razloge u kvaliteti i osobinama proizvoda, troškovima servisiranja, troškovima distribucije, uvjetima prodaje, efikasnosti i sl.

Odabir odgovarajuće cjenovne strategije, koja nije reaktivna, već naprotiv proaktivna ima potencijal ostvarenja većeg profita. U tu svrhu, Ross (1984) definira kvalitetan i uspješan cjenovni sustav kao sustav kojeg odlikuju sljedeće karakteristike:

1. Osigurava sakupljanje točnih i pravovremenih bitnih informacija o tržištu, kupcima, sposobnostima i akcijama konkurencije i internim mogućnostima i troškovima.
2. Sakupljanje i analiza podataka započinje u fazi razvoja proizvoda i nastavlja se tijekom cijelog životnog puta proizvoda.

3. Poduzeće se organizira na način da efektivno iskoristi sve podatke neophodne za donošenje cjenovnih odluka.
4. Sustav za prikupljanje i korištenje cjenovnih podataka treba biti fleksibilan i na raspolaganju svima kojima su ovi podaci potrebni.
5. Na funkcioniranju sustava za prikupljanje i analizu cjenovnih podataka rade najkvalitetniji kadrovi.
6. Efektivna je kontrola i postojanje povratnih veza u praćenju i realizaciji donesenih cjenovnih odluka.

4.2.2 Troškovi

Troškovi predstavljaju u novcu izraženu vrijednost utrošaka resursa u svrhu ostvarivanja određenih učinaka. Postoji više klasifikacija troškova, a one proizlaze iz različitih ciljeva i potreba koji su u njima iskazani. Neke od njih su:

- Prema funkcijama
    - Proizvodni
    - Neproizvodni
- Prema mogućnosti obuhvata po nositeljima
    - Direktni
    - Indirektni
- Prema dinamičnosti
    - Fiksni
    - Varijabilni
    - Mješoviti
- Troškovi prema pojavnom obliku
    - Troškovi trajne imovine
    - Troškovi potrošne imovine
    - Troškovi rada
    - Troškovi usluga
    - Ostali troškovi

Interesantna podjela sa stanovišta CPA (Shapiro et al., 1987), jeste ona koja troškove dijeli na:

1. Prije-prodajne troškove
2. Produkcijske troškove
3. Distribucijske troškove
4. Troškove poslije-prodajnog servisa.

Prije-prodajni troškovi nastaju u primarnoj fazi odnosa s kupcima, onoj koja prethodi prodaji. Period je to koji poduzeće kontaktira kupca putem svog prodajnog osoblja, prezentira mu svoj asortiman proizvoda, te dogovara način i uvjete prodaje. Troškovi ovih procesa se dosta razlikuju od kupca do kupca. Njihovu veličinu i različitost uvjetuju faktori poput prostorne alokacije kupaca, zahtijevane količine pažnje, različitih načina komuniciranja, iskazane želje kupca za kontaktiranjem s visokostručnim osobljem itd. Manjak evidencija uspostavljenih kontakata i resursa utrošenih u tim procesima, često je uzrok nemogućnosti utvrđivanja realnih troškova po kupcu.

Produkcijski troškovi nastaju u fazi realiziranja narudžbi kupaca. Sva različitost narudžbi pojedinačnih kupaca utječe na količine resursa potrebnih za njihovu realizaciju. Narudžbe se prije svega razlikuju količinski. Prodavatelj preferira veće narudžbe zbog toga što efikasno koristi jedinstvene resurse za njihovu proizvodnju i distibuciju. Ipak, u uvjetima kada proizvođač ne drži zalihe na skladištu već proizvodi prema potrebama narudžbi, periodi veće potražnje će zahtijevati dodatne angažmane resursa i stvoriti veće troškove. Kupci često imaju dodatnih zahtjeva u svezi dizajna ili pakiranja proizvoda. Različitost potreba može zahtijevati i neke specijalne karakteristike proizvoda. Kako se tada radi o nestandardiziranim karakteristikama, ove narudžbe će prouzrokovati dodatne produkcijske troškove. Vrijeme isporuke za kupca može predstavljati bitan element u kupoprodajnoj transakciji. Kraće vrijeme isporuke od uobičajenog, uglavnom je izvor povećanja troškova isporučitelja. Ovi razlozi, među ostalima, mogu činiti izvor razlikovanja veličine resursa koje proizvođač troši po pojedinačnoj narudžbi, odnosno po pojedinačnom kupcu.

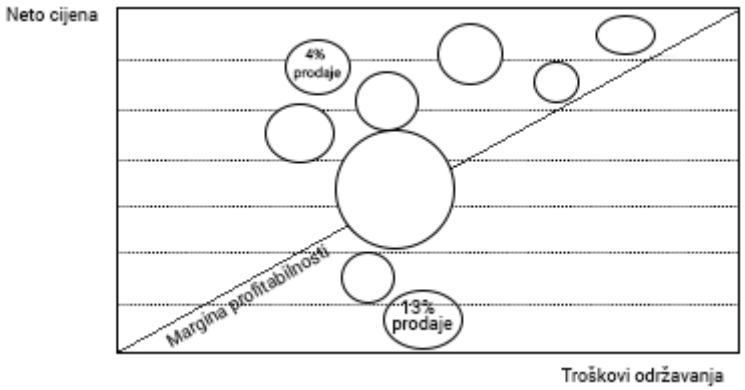
Nakon ugovaranja i proizvodnje slijedi distribucija proizvoda kupcima. *Distribucijski troškovi* variraju s zemljopisnom disperzijom kupaca. Distribuciju prate i različiti zahtjevi za brzinom isporuke. Brze isporuke tako povećavaju troškove jer su uz njih vezani drugačiji načini transporta u odnosu na uobičajene isporuke. Broj lokacija isporuke varira te tako i

pripadajući im troškovi. U slučajevima isporuka specijalnih proizvoda, kupac zahtijeva i odgovarajuće logističke mjere, poput ekipa specijaliziranih stručnjaka za instaliranje i puštanje u rad proizvoda ili pak isporuke infrastrukture neophodne za uporabu proizvoda.

Troškovi poslije-prodajnog servisa ovise o ugovorenim uvjetima isporuke i održavanja prodanih proizvoda. Nisu zanemarive situacije u kojima pojedini kupci zahtijevaju veliki angažman resursa prodavatelja na ovim procesima. Uglavnom se ovi procesi odnose na poslije prodajne instalacije, popravke, dorade, tehničku potporu, odnosno održavanje proizvoda. Ukoliko su ovi procesi uključeni u cijenu proizvoda, a nisu procijenjeni na pravi način, mogu predstavljati izvor dodatnih troškovnih opterećenja za prodavatelja. Istina i kupci se međusobno razlikuju s obzirom na stupanj zahtijevanih radnji u poslije-prodajnom procesu.

Zbroj svih navedenih troškova na razini kupca daje ukupan trošak po kupcu. Realno je očekivati da prodajna cijena proizvoda bude iznad ove razine troškova, te da poduzeće prodajom ostvaruje profit. Cijena nije jedinstvena za svakog kupca. Od kupca do kupca cijene znaju varirati u značajnom rasponu. Najčešće ove cijene nisu u direktnoj relaciji s iznosom troškova. Prema slici preuzetoj od Shapiro et al. (1987), koja je rađena na konkretnom primjeru, a za transakcije napravljene u vremenskom periodu od mjesec dana, jasno se uočava da određene narudžbe i određeni kupci generiraju gubitke, a drugi pak profit. Na apscisi se nalaze troškovi usluge korisnika, a na ordinati cijena proizvoda. Dijagonalna linija predstavlja razinu cijene koja je jednaka trošku, pa je tako područje ispod te linije ono u kome se ostvaruje gubitak, dok je iznad nje područje dobitka. U oba područja nalaze se grupe narudžbi/kupaca koje ostvaruju veće/manje dobitke/gubitke.

Slika 4.2 Odnos cijena i ukupnih troškova

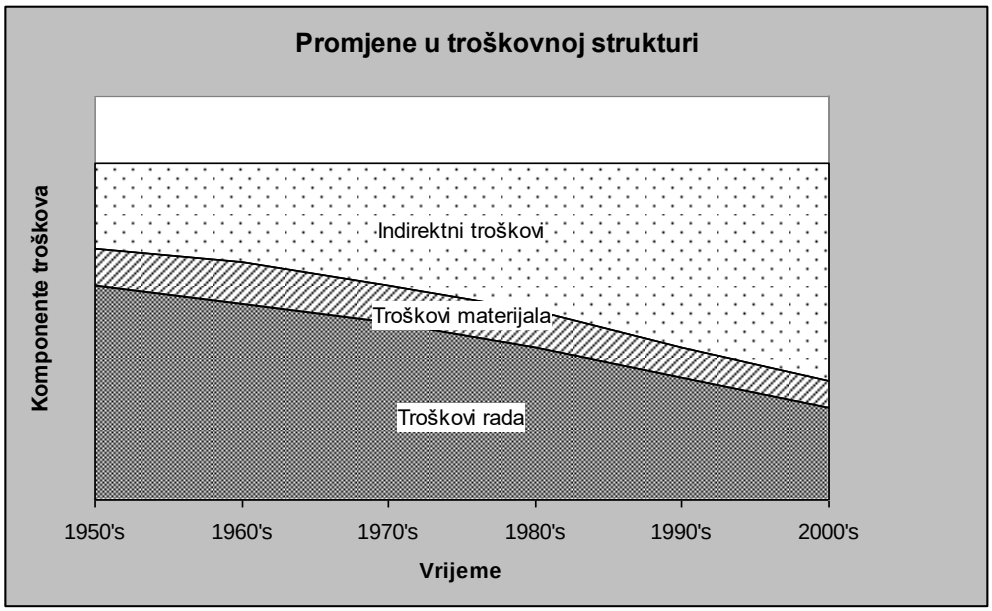


Preuzeto iz (Shapiro et al., 1987)

4.2.2.1 Raspored troškova

CPA analiza se provodi utvrđivanjem troškova na razini kupca. Raznovrsnost troškova, iskazana u prethodnom odjeljku, postavlja zid pred klasične računovodstvene metode rasporeda troškova. One raspoređuju troškove po mjestima nastanka troškova i prema načinu rasporeda na nositelje troškova.

Slika 4.3 Promjene u troškovnoj strukturi



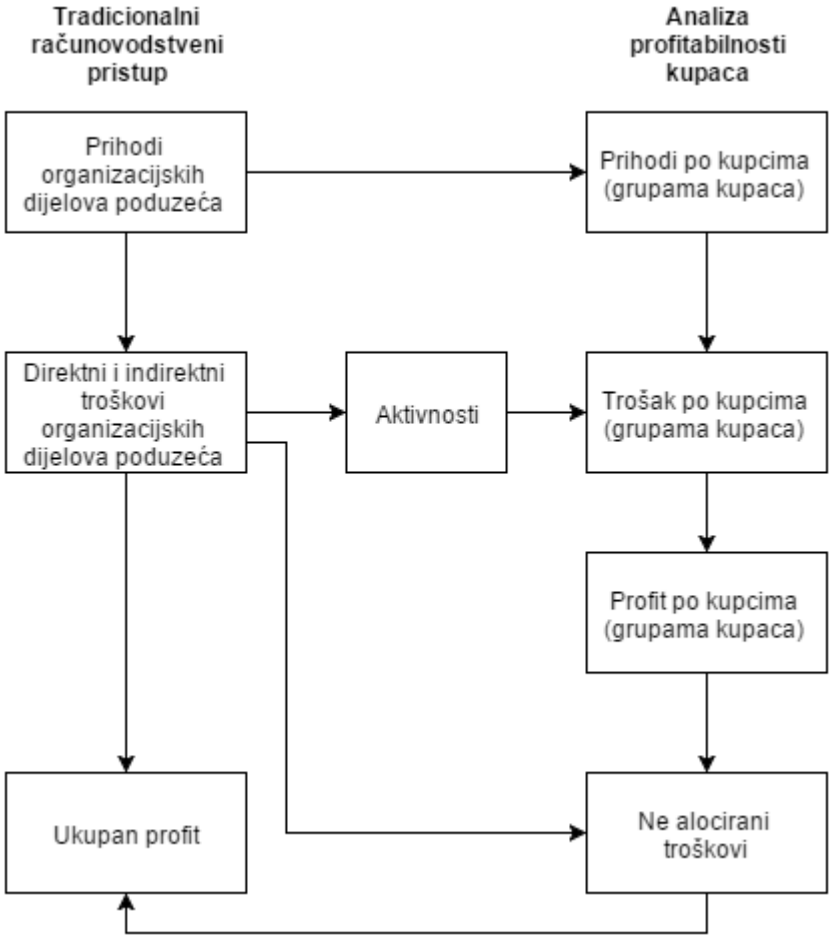
Prilagođeno iz (Cokins, 2001)

Zanimanje za profitabilnost kupaca jedan je od uzroka pojave zanimanja za istraživanje alternativnih metoda obračuna troškova. Ovo zanimanje ima i svoju povijesnu dimenziju (slika 4.3). Ogleda se u povećanoj diversifikaciji i kompleksnosti proizvoda, usluga i kupaca, čime indirektni troškovi preuzimaju primat u ukupnoj masi troškova u odnosu na direktne troškove (Cokins, 2001). Razlozi nove strukture troškova, tj. promjene odnosa indirektnih i direktnih troškova u ukupnoj masi, prvenstveno se očituju u kompleksnosti proizvoda i različitosti baze kupaca. U novim uvjetima poslovanja, klasične računovodstvene metode obračuna troškova ne mogu ispuniti zadatke koji se pred njih postavljaju. Za potrebe CPA analiza, potrebno ih je reorganizirati na način da se kupci

prepoznaju kao odgovarajuće jedinice analize. Cooper i Kaplan (1991) ističu primjenjivost koncepta ABC (engl. Activity-Based Costing) troškovnog upravljanja kojim poduzeća nadilaze nedostatke tradicionalnih računovodstvenih sustava.

Krakhmal (2006) prikazuje (slika 4.4) usporedne informacijske tijekove tradicionalnog računovodstvenog modela i CPA modela.

Slika 4.4 Informacijski tijek u analizi profitabilnosti kupca



Preuzeto i adaptirano iz (Krakhmal,2006)

Bitno je naglasiti da ABC metoda ne mijenja postojeći računovodstveni sustav troškova. Ona prevodi troškove na troškovne objekte, te ih tako čini dobrom osnovicom za donošenje točnijih i produktivnijih odluka na različitim razinama odlučivanja. Provodi se kroz sljedeća dva koraka:

- Identificiranje aktivnosti koje provodi poduzeće u svom poslovanju i njima pridruženih troškova resursa,
- Pridruživanje ovih aktivnosti objektima kao što su kupci, proizvodi ili procesi.

Spomenute aktivnosti su primjerice: naručivanje sirovine od dobavljača, zaprimanje sirovine, skladištenje proizvoda, prodaja proizvoda i sl. Resursi koji se koriste u tim aktivnostima su rad zaposlenika, oprema, tehnologija, potrošni materijal, novac i troškovi povezani uz njih. Kupci, proizvodi i procesi troše ove resurse i stoga im se pridružuju njihovi troškovi. Pridruživanje troškova se vrši na osnovu učešća u konzumiranju pojedinačnih aktivnosti, a učešće se procjenjuje intervjuiranjem i anketiranjem osoba uključenih u procese. Iako je doživjela široku primjenu, konvencionalna ABC metoda je u implementaciji naišla na niz problema (Kaplan i Anderson, 2007):

- Intervjuiranje i anketiranje su skupi i dugotrajni procesi,
- Podaci za ABC model su subjektivni,
- Pohranjivanje podataka, procesiranje i izvještavanje su skupi procesi,
- Većina ABC modela su lokalnog karaktera i ne nude integrirani pogled na profitabilnost cjelokupnog poduzeća,
- Ažuriranje ABC modela zbog promjena okolnosti nije lagan posao,
- Model je teoretski netočan pošto ignorira potencijal nekorištenih kapaciteta.

Nadogradnju klasičnog ABC modela daju Kaplan i Anderson (2007) kroz TDABC (engl. Time –Driven Activity-Based Costing) model. Njihov model je jednostavan, jeftiniji i mnogo jači za praktičnu uporabu. On pridružuje troškove resursa direktno objektima koristeći okvir koji zahtijeva:

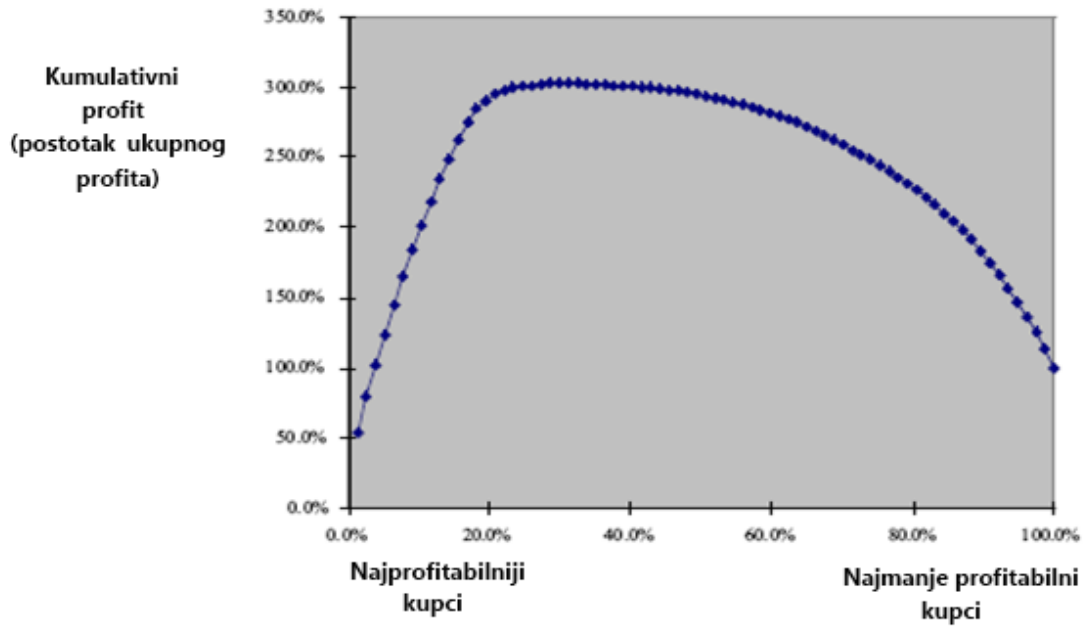
- procjenu troškova kapaciteta resursa i
- procjenu vremena korištenja kapaciteta za pojedine akcije.

Procjena vremena korištenja kapaciteta za pojedine akcije iskazuje se kroz vremensku jednadžbu (engl. time equation). Prilagodbe modela, koju zahtijevaju promjene u okruženju (internom i eksternom), primjenjuju se kroz dosta jednostavan izračun jediničnih troškova kapaciteta resursa i/ili kroz modifikaciju vremenske jednadžbe.

4.2.2.2 Distribucija profitabilnosti

Prethodna razmatranja o troškovima i njihovom rasporedu na kupce pokazala su da u bazi kupaca postoji određeni dio profitabilnih i određeni dio neprofitabilnih kupaca. Stupanj diferencijacije ove dvije grupe kupaca zavisi od više činitelja, a grafičku prezentaciju distribucije najvjernije je dočarao Kaplan tzv. „krivuljom kita“ (engl. whale curve) ili Kanthal krivuljom (engl. Kanthal curve). Krivulja je dobila naziv po kompaniji Kanthal na kojoj su Kaplan (1989-1, 1989-2) i Cooper (Kaplan i Cooper, 1998) izučavali profitabilnost kupaca uporabom ABC metode obračuna troškova. „Kanthal efekt“ izražava svu ozbiljnost potrebe izračuna profitabilnosti pojedinačnog kupca.

Slika 4.5 Kanthal krivulja



Preuzeto iz (Kaplan, Narayanan, 2001)

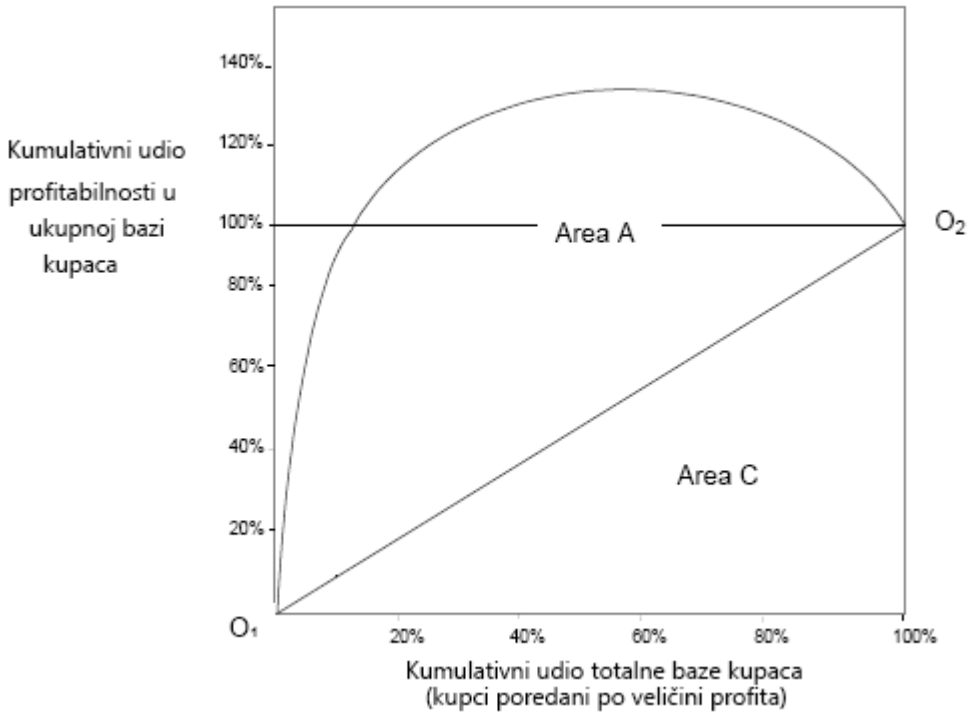
Zajedničko obilježje sustava u kojima se manifestira u svom ekstremnom obliku jesu kompanije koje obračun troškova rade uporabom ABC metode i zadovoljavaju slijedeća dva uvjeta (Kaplan i Atkinson, 1998):

1. U strukturi troškova veliki je udio indirektnih troškova i
2. Postoji velika diversifikacija proizvoda, kupaca i procesa.

Ordinata grafa izražava kumulativni profit sustava izražen u postotnom udjelu od ukupnog profita, dok apscisa prikazuje ukupnost baze kupaca poredanu po veličini profita koji ostvaruju. Krivulja daje zanimljive rezultate: 40% kupaca je profitabilno i donosi 250% od totalnog profita. Zadnjih 10% generira gubitak od gotovo 150% vrijednosti ukupnog profita, a ostatak baze kupaca se nalazi oko točke pokrića (krivulja završava u točki 100% profita).

Stobachoff krivulja (slika 4.6) daje kompletniju analizu distribucije profitabilnosti kupaca uvodeći dva pokazatelja: Stobachoff indeks i indeks proporcije profitabilnih kupaca.

Slika 4.6 Stobachoff krivulja



Preuzeto iz (Storbacka, 1997)

Uz pretpostavku da su svi kupci jednako profitabilni, krivulja profitabilnosti bi bila linearna funkcija predstavljena crtom ograničenom točkama O_1 i O_2. Kako ovakva situacija nije realna, stvarna krivulja kumulativne profitabilnosti pokazuje devijaciju baze kupaca u

odnosu na početnu pretpostavku (idealna baza jednako profitabilnih kupaca). Storbacka izvodi Stobachoff indeks postavljajući u odnos područje između stvarne krivulje profitabilnosti i osi x (T=A+C) i područje koje zatvara idealna krivulja profitabilnosti sa osi x (područje C):

$$S = \frac{A}{T} \qquad (4.1)$$

Indeks izražava stupanj devijacije profitabilnosti stvarne u odnosu na idealnu bazu kupaca. Vrijednost poprima u intervalu između 0 i 1. Ekstremne vrijednosti su slučaj baze jednako profitabilnih kupaca (idealan slučaj u kome je vrijednost indeksa 0; A=0) i ipak samo teoretskog maksimuma kada je indeks jednak 1 (jedan neograničeno profitabilan kupac, jedan neograničeno negativno profitabilan kupac i svi ostali s nultom profitabilnošću).

Drugi bitan pokazatelj je proporcija profitabilnosti, a predstavlja omjer profitabilnih kupaca i ukupne baze kupaca:

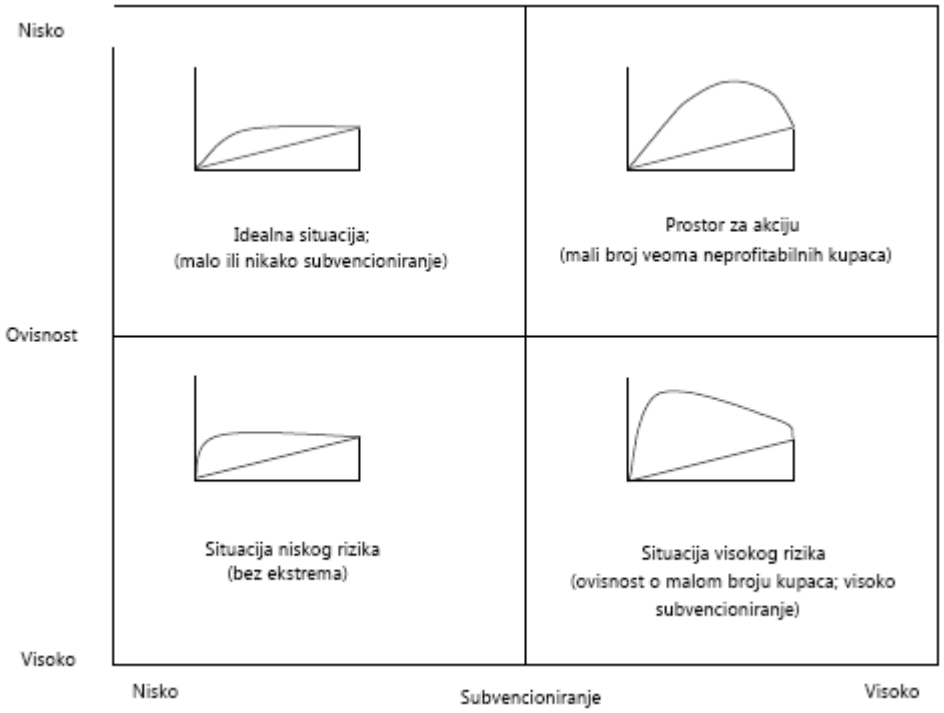
$$P = \frac{n - n_u}{n} \qquad (4.2)$$

U navedenom odnosu n je ukupan broj kupaca, a n_u je broj neprofitabilnih kupaca.

Stobachoff indeks je mjera tzv subvencionirajućeg efekta (engl. subsidising effect). U situaciji neravnomjerno raspoređene profitabilnosti, očigledno je da kupci s pozitivnim profitnim efektom nadoknađuju negativni efekt grupe neprofitabilnih kupaca. Analizirajući broj profitabilnih kupaca (4.2) moguće je napraviti portfolio analizu baze kupaca, definirati prednosti i opasnosti postojeće strukture kupaca, te ukazati na moguće korake menadžmenta u upravljanju kupcima (van Raaij et al. 2003).

Slika 4.7 prikazuje poduzeće s istom razinom profitabilnosti svoje baze kupaca, ali s četiri različita oblika krivulje. Kvadranti II i IV predstavljaju situacije s visokim Stobachoff indeksom, odnosno situacije s velikim efektom subvencioniranja.

Slika 4.7 Portfolio analiza baze kupaca



Preuzeto iz (van Raaij et al. 2003)

U slučaju kvadranta II, proporcija profitabilnih kupaca je velika jer je očigledno da je velik udio profitabilnih kupaca, a mali udio neprofitabilnih. Situacija je to koja zahtijeva akcija koja bi neprofitabilne kupce prevela u grupu profitabilnih ili ih se pak riješila. Potpuno je drugačija situacija u kvadrantu IV s veoma malim brojem profitabilnih kupaca i velikim brojem neprofitabilnih. Gubljenje nekog od kupaca iz grupe profitabilnih ozbiljno bi ugrozilo poslovanje, te su oni u jako povoljnoj pregovaračkoj poziciji u odnosu na ostale kupce. Potencijalno pregovaranje s ovakvim kupcima koje bi ishodilo za njih povoljnije uvjete kupnje (prevođenje u grupu manje profitabilnih ili pak neprofitabilnih kupaca) na jasan način ukazuje kako bi nepoznavanje strukture profitabilnosti kupaca moglo ozbiljno ugroziti poslovnu poziciju poduzeća na tržištu. Situacije iz kvadranata I i III su one s malo rizika jer

ukazuju na bazu profitabilno ravnomjerno raspoređenih kupaca bez ekstrema koji mogu ugroziti poslovanje.

4.2.3 Značajke kupca

Profitabilnost je kompleksna karakteristika kupca koja se izvodi iz skupa značajki kupca. Prethodno navedeni faktori profitabilnosti, troškovi i cijene, upravo su ovisni o značajkama kupca. U svom modelu mjerenja profitabilnosti Niraj et al. (2001) navode veličinu kupca i faktore kompleksnosti kao glavne značajke kupaca koje utječu na njihovu profitabilnost. Veličinu kupca promatraju kroz veličinu prometa koji kupac čini, a kompleksnost kroz druge karakteristike poput broja narudžbi, broja proizvoda, stupnja prilagođavanja proizvoda, broja lokacija isporuke i slično. Istina, ove značajke promatraju pod *ceteris paribus* uvjetima i donose zaključke da „*veličina prometa kupca obično ima pozitivnu vezu s veličinom ukupnih troškova servisiranja kao i s veličinom profitne margine*“, te i da „*kompleksnost rezultira većim servisnim troškovima i manjim profitom od kupca*“.

Veličina kupca se redovno nametala kao glavna značajka kupca (Farris et al., 2006) koja pozitivno determinira profitabilnost. Ipak, istraživanja smjera i veličine ove veze dali su različite rezultate. Shapiro et al. (1987) tako daju primjer eminentnog proizvođača kapitalne opreme koji, upravo s velikim kupcima, ostvaruje najveće gubitke. Storbacka (1997) u istraživanjima koristi bankarski sektor i dolazi do zaključka da „*ono što pokreće profitabilnost jeste volumen veze s kupcem*“. Kaplan i Cooper (1998) otkrivaju da veliki kupci su oni koji prave najveći pozitivni, ali i negativni učinak promatrajući ukupnost profitabilnosti poslovanja. Različitost rezultata istraživanja je, prije svega, ukazala na neophodnost istraživanja profitabilnosti pojedinačnog kupca. Veliki kupci su generalno gledajući profitabilniji od manjih kupaca. Ipak, veličina ne vodi jednosmjernoj vezi ka većoj profitabilnosti.

Zanimljiv konceptualni model u istraživanju utjecaja veličine kupca na njegovu profitabilnost postavlja van Triest (2005). Model ima dvije osnovne varijable: veličinu kupca i marginu proizvoda, te dvije posredne varijable: efikasnost razmjene i investicija u poslovni odnos s kupcem. Autor ističe da se profitabilnost ne može promatrati kao funkcija jedne od ovih varijabli nego upravo suprotno. Uporabom SEM (engl. Structured Equation Model)

metode, postavlja varijable u međusobni odnos i promatra njihovu međuzavisnost u djelovanju na profitabilnost kupca. Rezultati analize pokazuju da se „*utjecaj veličine kupca na profitabilnost ogleda poglavito kroz efikasnost procesa razmjene*“, te da je ova „*efikasnost dovoljno velika da prevlada negativne utjecaje niže margine proizvoda i većih troškova potpore za velike firme*“.

4.2.4 Značajke poduzeća

Poduzeće, u kom se mjeri profitabilnost kupaca, karakterizira niz osobina koje nisu direktno povezane s kupcima, ali su direktni pokazatelji kvalitete organizacije i uspješnosti poslovanja. Uzimajući u obzir kriterije izvora podataka i način iskazivanja, mogu se podijeliti na financijske (računovodstvene) i nefinancijske pokazatelje (Krstić i Sekulić, 2007).

Financijski pokazatelji izražavaju se u novcu. Izvor podataka za njihov izračun čine povijesni podaci iz računovodstvenog informacijskog sustava. Za razliku od njih, nefinancijski pokazatelji izražavaju se nemonetarno, kvalitativne su prirode i izvor podataka nije im računovodstveni informacijski sustav poduzeća. Između ova dva skupa pokazatelja postoji uzročno-posljedična veza na način da poboljšanje nefinancijskih pokazatelja poduzeća vodi ka poboljšanim financijskim pokazateljima. Sa stanovišta upravljačkih procesa, oba skupa pokazatelja potrebno je promatrati u okvirima strateških ciljeva poduzeća, inače njihova mjerenja nisu značajna.

Krstić i Sekulić (2007) mjerila performansi poduzeća promatraju sa stanovišta nastanka i funkcionalne pripadnosti, te ih dijele na:

- Mjerila ulaganja
- Mjerila efekata
- Mjerila efektivnosti
- Mjerila efikasnosti
- Ostala mjerila
    - Mjerila tržišne vrijednosti
    - Mjerila performansi zasnovana na vremenu
    - Mjerila inovativnosti

- Mjerila fleksibilnosti
- Mjerila kvaliteta proizvoda.

Mjerila ulaganja iskazuju vrijednosni izraz ulaganja u poslovne procese poduzeća, a mjerila efekata iskazuju vrijednosti ostvarenih rezultata poduzeća. Neki od pokazatelja efekata poslovanja su ukupan prihod , prihod od prodaje i poslovna dobit. Mjerila efektivnosti stavljaju u odnos ostvarene i planirane rezultate., dok mjerila efikasnosti čine odnos između objektivno potrebitih ulaganja i ostvarenih ulaganja.

Od ostalih mjerila zanimljivu grupu čine mjerila tržišne vrijednosti poduzeća koja se izražavaju npr. u cijeni dionica. Slijedeću grupu pokazatelja definira vrijeme potrebito za izvršavanje određenih procesa, a sa stanovišta mjerenja profitabilnosti kupaca zanimljivi su primjerice vrijeme procesiranja narudžbi i vrijeme transporta. Inovativnost, kao jedna od pretpostavki opstanka na aktivnom tržištu današnjice, mjeri se pokazateljima koji pokazuju stupanj inovativnosti poduzeća, poput broja novih proizvoda, ulaganja u istraživanje i razvoj, ulaganja u edukaciju, broja inovacija u određenom periodu i sl. Grupa pokazatelja vezanih za kvalitetu proizvoda izražava zadovoljstvo i lojalnost kupaca (Donio et al, 2006), imidž marke proizvoda, stopu dobijanja/gubljenja kupaca itd.

4.2.5 Okolina

Poduzeće djeluje u eksternom okruženju koje karakterizira pretežito indirektan utjecaj na poduzeće. Niz varijabli ovog okruženja manje ili više djeluje na konačni rezultat profitabilnosti pojedinačnog kupca ili grupe kupaca. Buble (2006) definira dva ključna segmenta ove okoline:

1. opću ili socijalnu okolinu
    a. političko-pravna okolina,
    b. ekonomska okolina,
    c. socijalno-kulturna okolina,
    d. tehnološka okolina.
2. poslovnu ili okolinu zadatka
    a. konkurenti,

b. kupci,
c. dobavljači,
d. sindikati,
e. vlasnici,
f. partneri.

U općoj okolini bitno mjesto zauzima fiskalna politika države. Zakonski okvir koji definira porezna opterećenja, te i druga slična davanja, direktno utječe na prihode poduzeća i konačnu profitabilnost poslovanja.

Uz varijable ovog tipa jako su bitni i zakonski propisi koji reguliraju konkurentnost tržišta. Varijable ekonomske okoline odražavaju globalno stanje nacionalne ekonomije. One vrše neposredan utjecaj na uvjete i rezultate poslovanja a neke od bitnih su inflacija, kamatne stope i stopa nezaposlenosti. Od socijalno-kulturnih značajki okoline bitno je istaknuti demografsku sliku koja je osnova tržišta radne snage ali i potrošačke snage okruženja. Tehnološka osnovica poslovanja pak djeluje kroz stvaranje uvjeta za poboljšanje produktivnosti i proizvodnosti rada uz konačan rezultat u povećanju profitabilnosti poslovanja. Ovo su varijable na koje poduzeće ne može djelovati. Ipak, može ih promatrati, razumjeti njihov utjecaj i adekvatno reagirati kako bi optimalno iskoristili pozitivne efekte, a one negativne minimizirali.

Poslovna okolina je mnogo bliža poduzeću od socijalne i na nju može i mora utjecati svakodnevno. U svjetlu razmatranja profitabilnosti kupaca, konkurencija i dobavljači čine značajan segment utjecaja na poslovanje poduzeća. Konkurencija utječe dvojako: na području plasmana, ali i na području pribavljanja neophodnih resursa za proizvodnju. S druge strane dobavljači direktno utječu samo na nabavu resursa za proizvodnju, a preko njih i na krajnji rezultat poslovanja.

Niti jedan od ovih faktora ne bi se trebao promatrati odvojeno. Sinteza različitih varijabli iz okoline i njihovo djelovanje kroz ostale varijable općeg modela za mjerenje profitabilnosti, ukazuju na kompleksnost promatranog modela i na neophodnost uporabe kompleksnih statističkih i metoda umjetne inteligencije u procesu istraživanja.

4.3 RFM model

Prospektivna koncepcija analize profitabilnosti promatra kupca kao „nematerijalnu imovinu“ poduzeća i sukladno tome promatra mu „buduću vrijednost“ u formi vjerojatnog profita koji će se ostvariti u poslovanju s njim u nekom definiranom budućem periodu (Gupta, 2003, Kim et al., 2006). Ona definira i pojam CLV-a (engl. Customer Lifetime Value – CLV) kao trenutnu vrijednost očekivanih zarada u predstojećem poslovanju s kupcem. Sukladno tome, izračunati pokazatelji CLV-a čine osnovu za strateško marketinško i prodajno djelovanje menadžmenta poduzeća nad bazom svojih kupaca. Jedna od brojnih klasifikacija u marketinškoj literaturi, Gupta et al. (2006), definira sljedeće pristupe modeliranju CLV-a:

- RFM modeli;
- Modeli vjerojatnosti (engl. Probability Models) bazirani na Pareto/NBD (engl. Negative Binomial Distibution) modelu i Markovljevom lancu;
- Ekonometrijski modeli, također bazirani na NBD modelu;
- Modeli dosljednosti (engl. Persistence Models).

Gupta sa koautorima (2006) ističe RFM model kao najjednostavniji. Klasificira ga također pri tome i kao najmoćniji model za prognoziranje profitabilnosti kupaca. U literaturi se prepoznaje kao jedna od popularnijih metoda za mjerenje vrijednosti pojedinačnog kupca na osnovama povijesnih zapisa o ostvarenom prometu (Miglautsch, 2001; Reinartz i Kumar, 2003; Cheng i Chen, 2009).

4.3.1 Komponente RFM modela

RFM analitički model, koji upotrebljava tri jednako važne dimenzije za segmentiranje baze kupaca, originalno je predstavio Hughes 1994. godine (Hughes, 1994). R (engl. Recency) dimenzija odnosi se na dužinu vremenskog perioda proteklog od zadnje kupnje. F (engl. Frequency) se odnosi na broj kupovina u definiranom vremenskom periodu analiziranja kupovnih navika kupca, a M (engl. Monetary) na novčanu vrijednost kupnje

promatrane u spomenutom vremenskom periodu. Hughes definira vrijednost kupca rezultatom analize u kojoj generira RFM rezultat sintezom sva tri pokazatelja. Ovako vrednovane kupce moguće je segmentirati u različite klastere, te svakom od njih pristupiti odgovarajućom marketinškom strategijom.

U osnovi su dva načina segmentiranja kupaca. Prema prvom, sveukupni broj kupaca podijeli se u pet jednakih grupa na osnovu RFM vrednovanja (Miglautsch, 2002). Svaka od komponetnti modela vrednuje se jednako, odnosno ima jednake težinske faktore. Nakon definiranja ovih grupa, analiziraju se pojedinačno uz definiranje odgovarajućih marketinških strategije za djelovanje. Drugi pristup ne pridaje istu važnost svim komponentama modela, te na osnovama regresijske analize uključuje izračune relativnih težinskih faktora. Promatrajući različito njihovu važnost, nastaju prošireni ponderirani RFM (engl. Weighted RFM) modeli koji od strane različitih autora pojedinim dimenzijama daju različite težinske faktore, odnosno različit redosljed važnosti dimenzija. Stone (1995) i Chuang i Shen (2008) smatraju kako se pojedine dimenzije ne mogu jednako tretirati u različitim industrijama te se kao pitanje postavlja raspored važnosti R-F-M komponenti modela. Novija istraživanja, sasvim obrnuto navedenom, postavljaju M kompontentu na prvo mjesto, a R komponentu na zadnje mjesto (Chuang i Shen, 2008). Nastaju i neki novi modeli koji tri navedene varijable dopunjuju dodatnim, poput RFMTC (engl. Recency, Frequency, Monetary, Time since first purchase, and Churn probability) modela (Cheng i Chen, 2009) koji uključuje vrijeme od prve kupovine i vjerojatnost zadržavanja kupca u bazi kupaca poduzeća.

4.3.2 RFM model u analizi vrijednosti kupaca

Nekoliko modela je prepoznato u modeliranju analize vrijednosti kupca. Hiziroglu i Sengul (2012) klasificiraju ih u dvije grupe na osnovu modeliranja ponašanja kupaca: *past* i *future-past* modeli. Razlike u modelima su u osnovi dvojake. Prva razlika odnosi se na pretpostavku o (ne)aktivnosti kupca u budućnosti, a druga na (ne)uključenost troškova kupca u model. RFM model spada u prvu grupu zajedno sa SOW (engl. Share of wallet) i PCV (engl. Past customer value) modelima. U analizi vrijednosti kupaca koristi se preko 30 godina (Gupta et al., 2006). Kroz sve ovo vrijeme, različiti autori koristili su osnovni model, definirali pojedine komponente na osnovu svojih pretpostavki ili su proširivali i dopunjavali osnovni model koristeći ga u različitim područjima poslovnog djelovanja.

Hughes (1994) dijeli bazu kupaca u pet dijelova, kvartila, sukladno ukupnom RFM skoriranju. Pojedinačne komponente su također podijeljene u kvartile. Tako npr. F varijabla dobiva vrijednost 5 za 20% kupaca koji su imali transakcije u najskorije vrijeme, a vrijednost 1 za „najstarije“ zadnje transakcije. Ostale varijable su definirane istom logikom sukladno metrici same varijable. Bult i Wansbeck (1995) svoju analizu temelje na definiciji R varijable kao vremena proteklog od zadnje kupovine, F kao broj kupovina u određenom periodu vremena, a M kao monetarnu vrijednost koju kupac utroši u tijeku istog vremenskog perioda. Stone (1995) analizira kupce koji svoju kupnju obavljaju upotrebom kreditnih kartica i smatra da vrijednost pojedinačnih varijabli nije jednaka u kompoziciji RFM rezultata. Razmišljanja je da težinska vrijednost faktora zavisi od karakteristika industrije koja se promatra, ali i dolazi do zaključka da najveća težina pripada frekvenciji djelovanja kupca, a najmanja monetarnoj vrijednosti kupnje.

Shih i Liu (2003) promatraju poduzeće koje se bavi prodajom hardvera. U fokusu promatranja im je relativna važnost pojedinačnih komponenti modela, a koriste se AHP (engl. Analytic hierarchy process) metodom kako bi došli do svojih zaključaka. Gupta sa suradnicima (Gupta et al., 2006) nalazi R, F i M vrijednosti kupca kao uspješne prediktore, a RFM model smatra veoma primjenjivim modelom. Ipak, zanimljiva je njegova teza kako je konačni skor pojedinog kupca, prije rezultat marketinških djelovanja promatranog poduzeća u prošlosti, a ne monetarne vrijednosti ostvarene u poslovnom odnosu kupca i poduzeća.

Iz retroaktivne analize zasnovane na promatranjima prošlih transakcija, vođene zahtjevima tržišta za bržim reagiranjem i prilagođavanjem novim aktivnostima dinamičnog tržišta, nastaju proaktivne analize na osnovama RFM modela. Deterministički RFM modeli, za razliku od stohastičkih metoda, tretiraju dobit po proizvodu kao konstantnu kategoriju. Ovo je ograničenje koje predstavlja veliki problem u nastojanjima da se predvidi vjerojatnost kupovine pojedinačnog proizvoda za promatranog kupca. Zbrajanjem svih pojedinačnih prometa i dobiti po proizvodu zbraja se i greška spomenute pretpostavke. Ipak, metode rudarenja podataka i strojnog učenja nadilaze ovu pogrešku i čine RFM model primjenjivim u području predviđanja vrijednosti pojedinačnog kupca.

4.3.2 Metode strojnog učenja i statističke analize u RFM analizi

Primjena metoda strojnog učenja nije limitirana na specifična područja. Mnogobrojna su i veoma različita polja na kojima su ove metode iskušane i potvrđene u praksi. Paliwal i Kumar (2009) klasificiraju dotadašnju primjenu neuronskih mreža na područja financija i računovodstva, zdravstva i medicine, inžinjerstva i proizvodnje , marketinga, te područje generalne primjene ukazujući na postignute rezultate. Njihov rad je dao usporedbu MLFF (engl. Multilayer Feed Forward) mreža sa standardnim statističkim metodama regresijske analize, logističke regresije i diskiriminacijske analize. Komparativna studija rezultata predviđanja i klasifikacije u području marketinga daje evidentnu prednost metodama neuronske mreže nad metodama statističke analize. Neke od primjena su se odnosile na segmentaciju tržišta, predviđanje odgovora tržišta, lansiranje novog proizvoda, predviđanje prodaje, predviđanje odabira kupca i slično. Na drugim područjima primjene nije bila toliko dominantna njihova uloga, već su rezultati korisnosti pojedinih modela ukazali na blagu prednost metoda neuronske mreže ističući i dobru primjenjivost metoda klasične statističke analize u ovim područjima. Ukazujući na prednosti i nedostatke obje grupe metoda sugerira se njihovo kombiniranje u cilju postizanja optimalnih rezultata čime ih se svrstava u komplementarne, a ne konkurentne metode.

Značajnu primjenu metode strojnog učenja imaju na području procjene kreditnog rizika, kritične aktivnosti s kojom se susreće menadžment komercijalnih banaka. Bekhet i Eletter (2014) su u istraživanju za skup komercijalnih banaka koristili RBF-ANN i LR kao risk scoring metode. Obje metode su postigle dobre rezultate i došli su do zaključka da ne postoji preferirana metoda za evaluaciju kreditnih aplikacija.

Joseph et al.(2011) suprotstavljaju regresijskim modelima metode neuronske mreže u predviđanju recesijskih tijekova na osnovu promjena kamatne stope kao prediktora budućih ekonomskih tijekova. Utvrđuju prednost modeliranja metodama neuronskih mreža, ali ih i deklariraju kao sofisticiranije metode u dizajnu i primjerene za setove podataka kod kojih se pretpostavlja da su nelinearni i vremenski promjenjivi poput ekonomskih i financijskih mjera promatranih spomenutim istraživanjem.

Sposobnost neuronskih mreža da prepoznaju kompleksne relacije u skupu podataka s velikim brojem varijabli čini ih superiornim u odnosu na konvencionalne statističke metode. Ipak, sposobnosti logističke regresije dolaze do izražaja sa smanjenjem kompleksnosti

promatranog sustava, tj. nakon redukcije broja ulaznih i izlaznih varijabli (Pourshahriar, 2012). Dosta istraživanja i literature zadnjih godina jasno ukazuju na neuronske mreže kao odličan alat za problem klasifikacije i predviđanja (Panian i Klepac, 2003; Klepac i Mršić, 2006; Panian, 2007; Chiu i Tavella, 2008; Howson, 2008; Ngai et al, 2009; Vercellis, 2009). Gotova sva istraživanja navode sposobnost aproksimacije bilo koje nelinearne matematičke funkcije kao prednost nad ostalim metodama strojnog učenja, te ih stoga tretiraju kao posebno pogodne za modeliranje sustava koji operiraju nad varijablama među kojima nisu poznate međusobne relacije ili su utvrđene kao kompleksne i nepogodne za obradu klasičnim statističkim metodama. Svojstvene su im i neke osobine koje otežavaju njihovu uporabu (određivanje optimalnog broja slojeva, broja neurona i sl.) i čine ih vremenski zahtjevnim metodama.

U literaturi se susreću dva tipa modela za predviđanje profitabilnosti kupaca. Prvi model svoja predviđanja temelji na podacima dobivenim istraživanjima ponašanja kupca u poslovnom odnosu s konkurencijom i poduzećem koja je u centru istraživanja (Rust et al. 2004). Drugi model promatra samo poslovanje kupca s poduzećem koje predviđa njegovu vrijednost zanemarujući pri tome kupčevo poslovanje s konkurencijom (Niraj et al. 2001, Fader et al. 2005). U dosadašnjim istraživanjima predviđanja profitabilnosti po drugom modelu, a koji će koristiti i ovo istraživanje, ipak nisu postignuti izvrsni rezultati uporabom metoda strojnog učenja za razliku od primjena opisanih u prethodnom poglavlju.

Campbell i Frei (2004) koristili su tekuću profitabilnost za predviđanje buduće, te su istraživanjem utvrdili kako predviđanje pogrešno klasificira mnoge kupce. Malthouse i Blattberg (2005) u svom istraživanju koriste regresijski model i neuronske mreže. Svoje rezultate prezentirali su kovanicama 20-55 i 80-15. Dakle, od 20% visokoprofitabilnih kupaca pogrešno su klasificirali 55% njih u projekciji buduće profitabilnosti i osigurali im pogrešan marketinški tretman, dok su od 80% niskoprofitabilnih kupaca pogrešno klasificirali njih 15%. Donkers et al. (2007) promatraju CLV vrijednosti kupaca u industriji osiguranja. Grade nekoliko modela i ustvrđuju da jednostavni modeli daju najbolje rezultate (naivni model preuzimanja profita iz zadnjeg perioda promatranja i njegova neograničena projekcija na buduće periode za pojedinačnog kupca), te predviđaju bolje od regresijskih modela i modela kojima su eksplicitno modelirali stopu zadržavanja kupaca i vjerojatnost ponovljene kupovine.

Postojeća istraživanja predviđanja profitabilnosti kupaca uz uporabu metoda strojnog učenja i na osnovi povijesnih podataka koja poduzeća posjeduju u svojim bazama podataka, nisu pokazala potpunu pouzdanost i veliki uspjeh. Neke jednostavnije metode klasične statističke analize uz korištenje smanjenog broja nezavisnih varijabli, pokazala su se jednako dobrim, ponekad i boljim od ovih složenijih. Na ovom tragu, Rust et al. (2011), razvijaju model čija je osnova BG/NBD (engl. Beta-Geometric/Negative Binomial Distribution) model razvijen od Fadera (Fader et al. 2005) i koji nadilazi jednostavne upravljačke heurističke modele predviđanja buduće profitabilnosti kupca. Sukladno navedenom, postoji mnogo prostora za istraživanje u cilju pronalaženja primjenjivog modela za izračun tekuće i predviđanje buduće profitabilnosti pojedinačnog kupca iz postojećih podataka, a što je i predmet ovog rada.

4.4 Karakteristike industrije sušenog voća, orašastih proizvoda, sjemenki, mahunarki i žitarica

Poduzeće koje je ustupilo podatke za potrebe istraživanja djeluje na tržištu 16 godina. Osnovna djelatnost mu je uvoz, prerada, pakiranje i trgovina na veliko raznovrsnog asortimana prehrambenih proizvoda od orašastih plodova, sjemenki, mahunarki, žitarica, te suhog i egzotičnog voća. Kroz poslovnu strategiju, vlasnici i menadžment poduzeća definirali su slijedeće poslovne ciljeve:

- Striktna kontrola sirovina praćena kompletnom zdravstvenom dokumentacijom po aktuelnim standardima EU.
- Konstantno unapređenje proizvodnje, prodaje i marketinga.
- Organizacijsko i kadrovsko usavršavanje.
- Praćenje zahtjeva i trendova suvremenog tržišta.
- Pružanje kvalitetnog i neprekinutog servisa svojoj bazi kupaca.

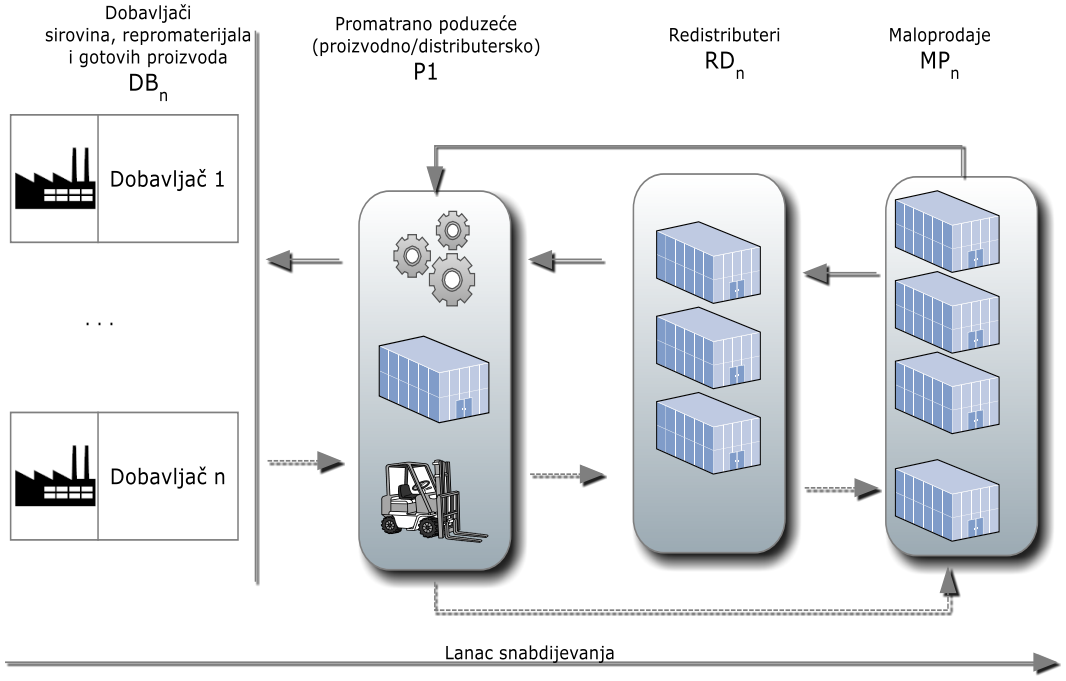
Poduzeće opsjeduje skladišno-proizvodni kompleks na 8000 m^2, od čega je 3000 m^2 skladišta s kontroliranom temeraturom od 0 do 8°C. Proizvodni kapaciteti temelje se najsuvremenijoj tehnologiji i jednim od najmodernijih postrojenja slične vrste u Europi. Stroga kontrola kvalitete podržava se suvremeno opremljenim vlastitim laboratorijem u

kojem se izvode organoleptičke i mikrobiološke analize, te analize prisutnosti aflatoksina. Kontrola se vrši neprekidno u svim fazama koje prolazi roba, od ulaska sitrovina u tvornicu do isporuke gotovih proizvoda. Svi procesi vezani za robu podržavaju sustav sljedivosti. U svakom momentu moguće je odgovoriti na upit vezan za porijeklo svake pojedinačne komponente koja čini gotov proizvod ove tvornice.

4.4.1 Industrijski procesi

Osnovna djelatnost promatranog poduzeća prikazana je slikom 4.8. Predmetno poduzeće prikazano je oznakom *P1*, a na pripadajućem dijelu lanca snabdijevanja nalazi se samo ono. Prethodnicu mu čini grupa dobavljača sirovina, repromaterijala i gotovih proizvoda koje poduzeće *P1* koristi u obavljanju svoje djelatnosti.

Slika 4.8 Tijek podataka i roba u lancu snabdijevanja



Prilagođeno iz (Niraj et al., 2001)

S desne strane poduzeća *P1* nalaze se grupe distributera i maloprodaja. Bitno je naglasiti da jedan distributer može posjedovati više distributerskih skladišta, ali i više maloprodajnih objekata. Kako izgleda tijek robe i što mu prethodi? Na slici su uočljiva dva

smjera kretanja robe i informacija među učesnicima lanca snabdijevanja. Punom crtom su označene narudžbe robe, a isprekidanom crtom isporuke robe. Upravo narudžbe iniciraju tijek robe u lancu snabdijevanja.

Proizvodno-distributersko poduzeće generira narudžbe sirovina, repromaterijala i gotovih proizvoda od svojih dobavljača označenih s DB_n. U odgovoru na spomenutu narudžbu događa se tijek robe DB_n-*P1*. Poduzeće *P1* skladišti primljenu robu, koristi je u procesu proizvodnje, te takovu (ili u izvornom obliku) prosljeđuje daljnjim tijekom lanca snabdijevanja do redistributera RD_n. Za jednog redistributera može postojati više lokacija isporuke. Slijedeće u lancu su maloprodaje koje u svom sastavu imaju jedno ili više prodajnih mjesta. Već je napomenuto da poduzeća iz kolone redistributera mogu imati svoje maloprodajne objekte. Sa slike je moguće uočiti i situaciju da maloprodaje naručuju robu direktno od poduzeća *P1,* ali i da poduzeće *P1* izvršava direktne isporuke robe tim maloprodajama.

U svim pobrojanim aktivnostima nastaju troškovi bez kojih se aktivnosti ne mogu niti provesti. Čitav niz troškova koji su uzrokovani odabirom dobavljača robe, generiranjem narudžbi, dostavom robe, cijenom robe, skladištenjem, preradom, isporukom, administrativnim akcijama, održavanjem kupaca i dr.aktivnostima, čine mozaik troškova koje je neophodno rasporediti na kupca kako bi se izračunala njegova profitabilnost. Naredni odjeljak analizira troškove koji nastaju u poduzeću *P1* i koje je potrebno uključiti u izračunu profitabilnosti pojedinačnog kupca.

4.4.2 Proizvodi

Jedna od komparativnih prednosti koja diferencira poduzeće na tržištu jeste širina asortimana. Cjelokupan asortiman poduzeća podijeljen je u 9 grupa:

1. Impulsna linija
2. Zdrava hrana
3. Čokolada + Suho voće i orašasti plodovi
4. Orašasti plodovi
5. Sjemenke
6. Mahunarke i žitarice
7. Mješavine mahunarki i žitarica

8. Suho voće
9. Egzotično voće.

Podaci o poslovanju poduzeća promatrani su za period 2008-2013. Stalno nastojanje poduzeća da prati tržišne trendove održavajući i unapređujući kvalitetu proizvoda rezultiralo je konstantnim povećanjem asortimana proizvoda. U 2008.godini osnovni asortiman je iznosio 278 artikala, a do 2013.godine ova brojka se povećala na 442 artikla.

Posebna pozornost u poduzeća poklanja se osluškivanju potreba tržišta i rekacijama na plasirane proizvode. Poduzeće je inovativno u segmentu ponude. Tijekom promatranog perioda razvilo je jedan dio asortimana potpuno novih proizvoda i plasiralo ih na na domaćem i inozemnom tržištu. Posebno značajno mjesto u ovom asortimanu zauzima grupa *zdrave hrane*. Vodeći se preporukama vodećih nutricionista i organizacija koje vrše i objavljuju istraživanja, prepoznati su pozitivni zdravstveni učinci orašastih plodova, te je na tom tragu oformljena čitava grupa proizvoda. U ovu grupu spadaju i proizvodi zasnovani na relativno novim proizvodima na ovom tržištu, a čije je originalno porijeklo prilično egzotično (proizvodi poput npr.chia, quinoa, mnogo poznatije aronije i sl.). Ulagalo se i postiglo uspjeha i na razvoju novih linija proizvoda. Ovi proizvodi nisu novi na tržištu, ali su osigurali nove tržišne segmente poduzeću. Istraživanja provedena kod krajnjih korisnika proizvoda dovela su do čitavog niza poboljšanja postojećih proizvoda kroz fine izmjene receptura, ali i korištenje novih sirovina i poluproizvoda u procesima proizvodnje.

Cijena gotovog proizvoda opterećena je direktnim i indirektnim troškovima. Osnovnu komponentu direktnih troškova čini fakturna cijena sirovinske osnove. Kako se sirovina većim dijelom uvozi, troškovi nabave čine znantnu komponentu nabavne cijene, a čine ih uglavnom prijevozni troškovi, carinska opterećenja, usluge obrade dokumentacije i troškovi različitih kontrola ispravnosti sukladno interno i eksterno postavljenim zahtjevima za kvalitetu i ispravnost proizvoda. Prosječna vrijednost zavisnih troškova nabave iznosi 6,10 % od fakturne cijene proizvoda.

Indirektni troškovi predstavljaju kategoriju čiju je točnost, kako je i navedeno u prethodnim razmatranjima, dosta teško odrediti. Interna analiza poduzeća u ovoj domeni definirala je strukturu ovih troškova na sljedeći način:

- Prženje i pakiranje

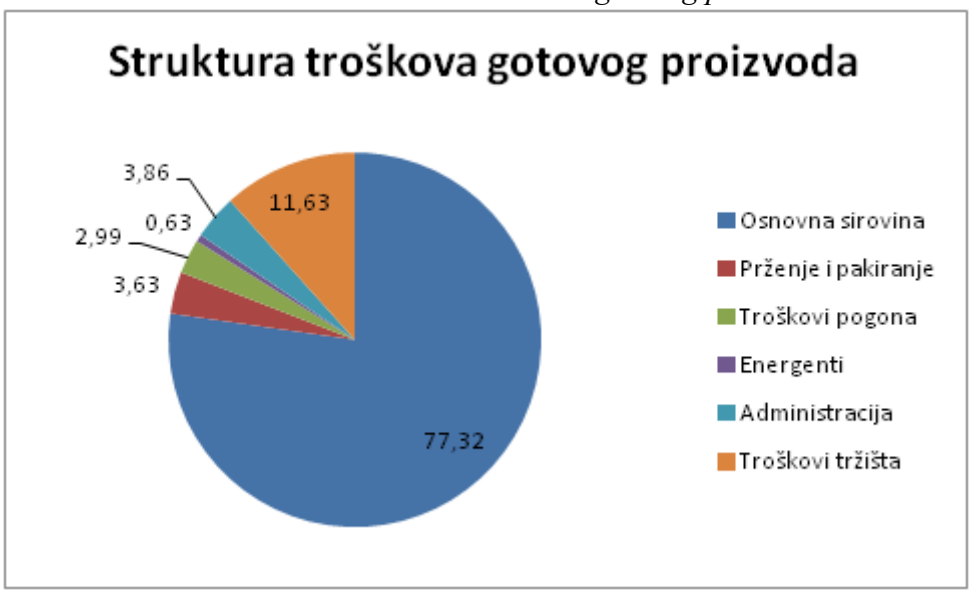
- Troškovi pogona
- Energenti
- Troškovi administracije
- Troškovi tržišta.

Analiza strukture troškova jednog od profitabilnih artikala prikazuje odnos direktnih i indirektnih troškova kao 77,32% : 22,68%. Grafički i tabelarni prikaz ovih podataka ukazuje na troškove prerade u iznosu od oko 7,25% ukupnih troškova, ali posebno i visok udio postproizvodnih troškova u iznosu od 15,49%, a koji se odnose na troškove plasiranja artikla na tržište i vlastite administracije.

Tablica 4.1 Prosječna struktura troškova iz grupe artikala s velikim udjelom u ukupnom prometu

Trošak	Udio (%)
Osnovna sirovina	77,32
Prženje i pakiranje	3,63
Troškovi pogona	2,99
Energenti	0,56
Administracija	3,86
Troškovi tržišta	11,63

Slika 4.9 Struktura troškova gotovog proizvoda



Izvor: podaci promatranog poduzeća

4.4.3 Dobavljači i kupci

Osnova dobrog proizvoda je kvalitetna i prvoklasna sirovina. Vodeći se ovim principima, menadžment poduzeća uspostavlja kontakte s vodećim proizvođačima sirovine na globalnom svjetskom tržištu. Posjećuju se njihovi nasadi i proizvodni kapaciteti, te se putem burzi ove robe kupuju sirovine i dogovaraju isporuke. Dobavljači su sa svih kontinenata osim Australije, sukladno širini asortimana robe. Obzirom na udaljenosti, isporuka robe vrši se uglavnom kontejnerskim brodskim prijevozom do jedne od luka na Jadranu.

Tablica 4.2 Broj dobavljača po kontinentima u 2013.godini

Kontinent	**Broj dobavljača**
Europa	51
Azija	17
Amerika	19
Afrika	1

Poduzeće posvećuje i veliku pozornost domaćem tržištu sirovina. Koristi svaku priliku da pomogne užu i širu lokalnu zajednicu proizvođača sirovina računajući na kvalitet domaćih dobavljača, ali i pokazujući društvenu svijest osjetljivu na potrebe održanja i razvoja domaćih proizvođača.

Nakon kvalitete sirovine koja je određena kao primarni kriterij, sljedeći važni kriteriji je cijena. Poduzeće nastoji dobiti povoljne nabavne cijene postupcima poput dugoročnih ugovaranja, narudžbama odgovarajućih količina robe, vođenjem optimalnih zaliha. Ugled dobavljača je također bitna komponenta njihovog odabira. Posebno mjesto zauzimaju prošla iskustva u poslovanju s dobavljačem u smislu točnosti isporuka i kvalitete isporučene robe.

Kupci gotovih proizvoda promatranog poduzeća većinom su s područja regiona (BiH, Hrvatska, Srbija, Crna Gora i Makedonija). Poduzeće ima svoje distribucijske centre na području Hrvatske i Srbije. Ostatak baze kupaca je s područja Europske unije i Rusije. Distribucijski centar za EU je na području Italije. Postoji i jedan kupac s područja Kanade gdje poduzeće nastoji stvoriti novo tržište. Distribucijski centri se brinu za pojedinačne kupce unutar regija koje pokrivaju.

Tablica 4.3 Broj kupaca po regijama u 2013.godini

Područje	Broj kupaca
Uži region	1528
EU	5
Rusija	1
Kanada	1

Zanimljiva je struktura kupaca obzirom na njihov tip. Većina prometa odvija se preko velikih nacionalnih trgovačkih lanaca. Oni čine 52,73% kupaca. Bitno je napomenuti da jedna grupa ovih kupaca centraliziano nabavlja robu, a potom ju vlastitim kanalima distribucije raspodjeljuje. Kod ostatka ovog segmenta roba se direktno distribuira na njihova prodajna mjesta.

Tablica 4.4 Struktura kupaca u 2013.godini

Tip kupca	Udio
NACIONALNI TRGOVAČKI LANCI	52,73%
MANJI TRGOVAČKI LANCI LOKALNOG KARAKTERA	14,06%
MANJE TRGOVINE, BENZ.PUMPE, KIOSCI	14,19%
MARKETI OD 50 - 200 M2	11,26%
VELETRGOVCI	6,18%
PROIZVOĐAČI	1,56%

Većinski dio ostatka baze kupaca čine manji trgovački lanci, pojedinačne trgovine i marketi. Zadnji, manji dio čine veletrgovci koji distribuiraju robu na području koje pokrivaju, ali i proizvođači koji uzimaju sirovinsku osnovu ili poluproizvode koje prerađuju i/ili pakiraju u svojim pogonima.

Više podataka o dobavljačima, kupcima i artiklima prikazano je u dijelu deskriptivne statistike poglavlja o empirijskim rezultatima istraživanja.

4.4.4 Opće karakteristike industrije

Od promatranog poduzeća dobiveni su podaci o uvozu i izvozu ove grane industrije u BiH u 2014.godini. Podaci su prikupljani od strane analitičko-marketinške službe poduzeća, a

izvor su zvanični izvori podataka (Statistički zavod BiH, Vanjskotrgovinska komora BiH, Carinska uprava, Eurostat...).

Tablica 4.5 Uvoz u BiH 2014.godine

Uvoz u BiH - 2014.godina		
Carinska tarifa	Vrijednost (KM)	Količina (kg)
07131090 Grašak / mahunarke - grašak zeleni	28.705,62	34.554,00
07132000 Mahunarke – slanutak	53.941,81	44.379,09
07133200 Grah t. crveni	34.851,19	11.503,10
07133390 Grah	11.075.751,05	6.241.721,42
07134000 Leća	163.642,76	100.810,09
07135000 Bob drobljeni	527,79	140,00
08012200 Brazilski orah oljušteni	139.341,42	9.999,22
08021190 Badem ljuska	1.326,04	269,00
08021290 Badem jezgra	3.236.460,06	266.959,50
08022100 Lješnjak ljuska	3.649,41	450,00
08022200 Lješnjak jezgra sirovi	2.569.291,39	182.355,90
08023100 Orah ljuska	3.185,04	300,00
08023200 Orah jezgra	6.111.285,39	592.968,56
08025100 Pistacio sirovi u ljusci	1.367.442,41	85.028,16
08025200 Pistacio sirovi oljušteni	75.189,12	2.556,14
08029050 Pinjoli	52.227,43	924,94
08041000 Datula	1.300.472,63	554.749,92
08042090 Lokum / smokva	1.532.987,22	422.530,78
08062010 Grožđica suha	1.079.239,42	345.732,06
08131000 Marelica / marelica suha	642.258,19	109.427,58
08132000 Šljiva	1.093.625,06	279.601,09
08133000 Jabuka	22.579,86	4.234,42
08134095 Goji berries	254.638,00	22.549,09
10059000 Kukuruz kokičar	70.570.671,54	247.142.078,45

11042295 Krupnik	17.044,05	5.340,81
11063090 Kokosovo brašno	1.744.125,70	479.624,49
12024100 Kikiriki ljuska sirovi	675.815,26	345.945,00
12024200 Kikiriki jezgra sirovi	1.325.425,49	584.506,80
19041010 Kukuruz prženi	8.903.874,21	1.610.489,82
19041030 Japanski snack / rice crackers	193.370,79	33.957,75
20060038 Kandirana aronija	89.903,72	10.008,69
20081198 Kikiriki sa preljevom / papete	1.075.899,48	193.924,00
20081912 Mix egzotik više od 1000 g	15.367,62	997,00
20081919 Badem s preljevom / Lješnjak s preljevom	2.448.918,43	165.054,18
20081992 Mix egzotik manje od 1000 g	25.097,04	2.149,27
20081993 Pistacio prženi slani / badem sa aromom dima	901.359,99	55.664,13
20081995 Indijski orah prženi / lješnjak jezgra prženi	668.819,61	57.352,83
20081999 Bundeva i suncokret	379.147,32	92.096,76
20082059 Ananas	62.942,20	20.442,02
20088050 Jagoda više od 1000 g	39.839,83	5.203,00
20088070 Jagoda manje od 1000 g	34.193,45	13.718,46
20089391 Brusnica nto mase veće od 1 kg	432.318,09	88.215,88
20089393 Brusnica nto mase manje od 1 kg	6.680,33	597,90
20089941 Ginger više od 1000 g	723,53	36,80
20089948 Banana čips / kokos čips/ kokos kocka / mango feta više od 1000 g/ mango kocka više od 1000 g / papaja više od 1000 g	43.306,13	8.146,26
20089949 Brusnica / Jagoda	205.888,40	55.093,03
20089951 Ginger manje od 1000 g	25.525,36	4.820,60
20089963 Mango feta manje od 1000 g/ mango kocka manje od 1000 g/ papaja manje od 1000 g	2.092,81	200,00
20089985 Pop corn microwave	521.093,22	159.422,80
UKUPNO	121.252.061,91	260.448.830,79

Izvor: istraživanje promatranog poduzeća

Tablica 4.5 prikazuje ukupnu vrijednost uvoza pojedinih skupina proizvoda zanimljivih sa stanovišta ovog rada i predmetnog poduzeća čiji se podaci koriste u istraživanju. Tablica 4.6 prikazuje ukupnu vrijednost izvoza promatrane skupine proizvoda u 2014. godini. Jasno je vidljivo da je pokrivenost uvoza izvozom 21,92%.

Tablica 4.6 Izvoz iz BiH 2014.godine

Izvoz iz BiH - 2014.godina		
Carinska tarifa	Vrijednost (KM)	Količina (kg)
7131090 Grašak / mahunarke - grašak zeleni	3.069,71	1.032,00
7132000 Mahunarke – slanutak	7.228,33	2.241,00
07133390 Grah	189.714,70	48.644,00
07134000 Leća	10.013,76	3.678,00
08012200 Brazilski orah oljušteni	51.900,82	2.400,00
08021290 Badem jezgra	257.639,81	13.050,32
08022200 Lješnjak jezgra sirovi	66.391,02	3.940,23
08023200 Orah jezgra	361.097,60	23.806,40
08029050 Pinjoli	38.869,78	515,20
08041000 Datula	103.014,91	22.429,70
08042090 Lokum / smokva	332.789,96	46.517,60
08062010 Grožđica suha	96.555,66	19.369,09
08131000 Marelica / marelica suha	240.175,75	26.808,74
08132000 Šljiva	345.494,83	81.739,00
08133000 Jabuka	84.637,12	9.584,00
08134095 Goji berries	232.101,68	14.716,90
10059000 Kukuruz kokičar	20.961.662,80	53.083.742,40
11042295 Krupnik	1.315,51	360,00
11063090 Kokosovo brašno	142.707,84	29.886,00
12024200 Kikiriki jezgra sirovi	42.384,72	17.000,00
19041010 Kukuruz prženi	333.131,87	81.578,73
19041030 Japanski snack / rice crackers	19.773,78	1.800,60
20060038 Kandirana aronija	23.260,42	1.363,75
20081198 Kikiriki sa preljevom / papete	113.181,58	30.549,75
20081912 Mix egzotik više od 1000	45.716,81	5.054,82
20081919 Badem s preljevom / Lješnjak s preljevom	407.851,80	26.721,68
20081993 Pistacio prženi slani / badem sa aromom dima	865.579,53	38.477,41
20081995 Indijski orah prženi / lješnjak jezgra prženi	239.200,10	11.708,86
20081999 Bundeva i suncokret	880.226,77	92.828,93
20082059 Ananas	18.720,65	2.436,69
20088070 Jagoda manje od 1000 g	81,25	15,75

20089391 Brusnica nto mase veće od 1 kg	10.764,20	1.701,00
20089393 Brusnica nto mase manje od 1 kg	35.428,53	4.339,94
20089948 Banana čips / kokos čips/ kokos kocka / mango feta više od 1000 g/ mango kocka više od 1000 g / papaja više od 1000 g	10.856,48	1.264,74
20089949 Brusnica / Jagoda	9.034,89	1.192,80
20089985 Pop corn microwave	855,71	168,00
UKUPNO	**26.582.430,68**	**53.752.664,03**

Izvor: istraživanje promatranog poduzeća

Najveći uvozni partner BiH za promatrane artikle je Srbija s kojom se ostvaruje 63,7% ukupnog uvoza, a potom slijede SAD s 5,34% i Turska s 4,44%. Turska je najveći trgovinski partner u ovoj industriji. U nju BiH izvozi 78,44% ukupnog izvoznog asortimana, a na drguom mjestu je Hrvatska s 13,92%.

Tablica 4.7 Države partneri s najvećim prometom u uvozu i izvozu za 2014.godinu

Izvoz iz BiH - 2014.godina (Top 5)		
Država	Vrijednost (KM)	Količina (kg)
TURSKA	20.848.841,04	52.430.750,00
HRVATSKA	3.700.679,88	939.070,19
SRBIJA	778.688,76	107.760,40
CRNA GORA	460.623,84	43.249,65
SLOVENIJA	263.748,19	157.021,60
Uvoz u BiH - 2014.godina (Top 10)		
Država	Vrijednost (KM)	Količina (kg)
SRBIJA	77.232.801,13	239.417.655,25
SAD	6.474.020,46	887.342,85
TURSKA	5.389.247,94	849.344,67
RUMUNJSKA	4.214.571,60	335.074,50
KINA	4.007.003,98	2.057.050,76
ARGENTINA	3.058.203,82	1.682.537,88
HRVATSKA	2.824.894,94	8.605.484,02
KANADA	2.544.743,39	1.338.768,55
IRAN	1.818.458,09	507.671,30
UKRAJINA	1.590.238,80	646.457,36
EGIPAT	1.419.257,95	954.520,00
INDONEZIJA	1.364.654,81	387.516,51

Izvor: istraživanje promatranog poduzeća

Poduzeće je tijekom 2014.godine provelo istraživanje na uzorcima od po 310 ispitanika u BiH, Srbiji i Hrvatskoj. Ispitanici su krajnji potrošači koji su kupovali proizvode kategorije suhog voća i orašastih plodova, bilo da su njihovi vlastiti proizvodi ili proizvodi drugih proizvođača. Kupci su većim dijelom (70%) radno aktivno stanovništvo (25-35, 36-45 i 46-60 godina starosti), dok 30% otpada na osobe mlađe dobi (18-24) i starije od 60 godina. U svim zemljama, kao osnovni kriterij za kupovinu ovih artikala kupci su istakli kvalitetu proizvoda. Cijena je na drugom mjestu, a tek potom dolaze proizvođač i pakiranje. Zanimljivo je da je raspored važnosti kriterija isti u svim zemljama u kojima se vršilo istraživanje.

Tablica 4.8 Države partneri s najvećim prometom u uvozu i izvozu za 2014.godinu

Kriterij kupovine	BiH	Hrvatska	Srbija
Kvalitet	60,00%	55,00%	59,68%
Cijena	34,57%	30,00%	28,06%
Proizvođač/marka	2,00%	8,00%	5,81%
Pakiranje	1,14%	1,00%	2,90%
Ništa od navedenog	2,29%	6,00%	3,55%

Izvor: Istraživanje tržišta provedeno od strane promatranog poduzeća

Provedeno istraživanje potvrdilo je vodeću poziciju na domaćem tržištu u BiH gdje se 32,86% kupaca izjasnilo da kupuje njihovu marku promatranih proizvoda. U Hrvatskoj se, prema istraživanju, najviše kupuju robne marke trgovačkih lanaca (45% kupaca), a ovo poduzeće zauzima petu poziciju s 2,25% kupaca koji su se izjasnili da kupuju njihovu marku proizvoda. U Srbiji najviše mjesto zauzima lokalni proizvođač ovih grupa proizvoda (62,90%), robne marke trgovačkih lanaca kupuje 14,52% kupaca, a promatrano poduzeće zauzima četvrto mjesto s 4,19% kupaca. Uzevši u obir da jedan se dio proizvodnje poduzeća odnosi na usluge proizvodnje i pakiranja trgovačkih marki, može se reći da poduzeće zauzima dosta značajnu pozicju na tržištu užeg regiona.

5. MODEL PREDVIĐANJA PROFITABILNOSTI KUPCA

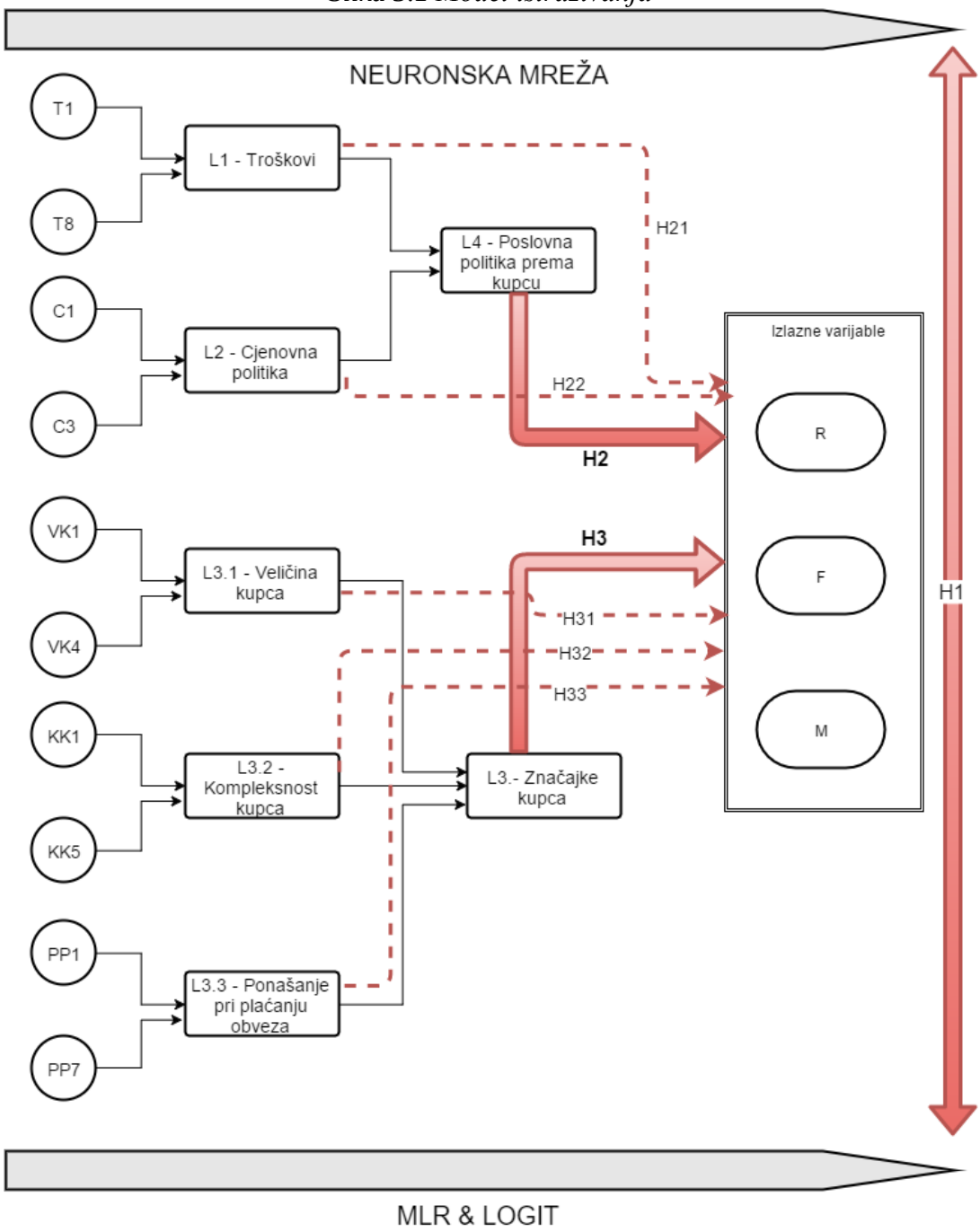
Osnovu modela istraživanja čine podaci o prošloj profitabilnosti kupaca promatranog poduzeća. Ciklus poslovanja započinje nabavom sirovina i repromaterijala od ugovornih dobavljača. Ova roba se skladišti, a onda i upotrebljava u proizvodnom procesu. Gotovi proizvodi se isporučuju distributerima ili maloprodajnim poslovnim subjektima gdje roba nalazi i svog potrošača. Svaka od pobrojanih aktivnosti generira troškove različite vrste. Dok je troškove sirovina relativno jednostavno izračunati, dosta je komplicirano alocirati čitav skup indirektnih troškova na pojedinačnog kupca i artikle koji su predmet kupoprodajnog odnosa s promatranim poduzećem. Direktni i indirektni troškovi jedna su komponenta jednadžbe izračuna profitabilnosti, a drugu čine prihodi od kupca. Postoje i neki drugi pokazatelji koje treba uzeti u obzir u procesu izračuna kao bitne i utjecajne. To su setovi podataka koji se odnose na značajke pojedinog kupca iskazane kroz njegovu kompleksnost i veličinu. Poduzeće provodi poslovne politike koje direktno utječu na odnos s pojedinačnim kupcem, a tu su i utjecaji poslovne i socijalne okoline. Pretpostavka modela uključuje primjenu metoda strojnog učenja i multivarijantne statističke analize nad pobrojanim podacima s ciljem predviđanja buduće profitabilnosti kupaca. Ovo poglavlje definira pojedinačne komponente modela i metode kojima se provodi istraživanje.

5.1 Definiranje okvira za predviđanje profitabilnosti kupca

Jedan od ciljeva istraživanja jeste stvaranje modela za predviđanje profitabilnosti kupaca, pri čemu će se testirati prednost specifične metode strojnog učenja, neuronske mreže, u odnosu na metode MSA analize, te pokušati i dokazati utjecaj provođenja marketinških politika i karakteristika kupca na mjere profitabilnosti.

Polaznu osnovu za predviđanje profitabilnosti čini izračun tekuće, odnosno prošle profitabilnosti pojedinačnih kupaca. Tekuća profitabilnost, ulazni podaci za njen izračun te ostali podaci koji čine nezavisne varijable modela, činiti će podatkovnu osnovu za primjenu metoda strojnog učenja i multivarijacijske statističke analize u pokušaju izgradnje modela kojim će se predviđati profitabilnost kupaca.

Slika 5.1 Model istraživanja



Nezavisne varijable općeg modela za predviđanje profitabilnosti razvrstane su (slika 5.1) u sljedeće grupe prema srodnosti:

- Podaci o troškovima (L1)
- Podaci o cjenovnoj politici (L2)
- Podaci o značajkama kupca (L3).

Izlazne varijable su proizvod modela za predviđanje. Njihova definicija u ovom istraživanju zasnovana je na RFM modelu, a imenovanje je izvršeno sukladno definiciji modela: R, F i M varijabla.

5.1.1 Troškovi (L1)

Osnovu za što točnije iskazivanje pokazatelja profitabilnosti kupaca čini adekvatan raspored troškova. Svakako da je prethodnica rasporedu utvrđivanje onog dijela troškova koji se mogu razvrstati po kupcima. ABC metodom definiraju se aktivnosti koje troše resurse poduzeća, a odgovarajuća razina ovih aktivnosti pridružuje se pojedinačnom kupcu uporabom troškovnih pokretača (engl. *cost drivers*).

U izgradnji modela za izračun profitabilnosti pojedinačnog kupca, Niraj et al. promatraju aktivnosti poduzeća u vremenskom periodu **t** i pri tome koriste sljedeću notaciju:

I = broj kupaca indeksiran po **i;**

S = ukupan broj proizvoda obrađen od strane poduzeća $\mathbf{P_1}$, indeksiran po **s;**

$\mathbf{N_i}$ = broj prodajnih mjesta Kupca **i**;

$\mathbf{m_{iS}}$= potražnja u broju jedinica za Proizvod **s** za svako od prodajnih mjesta Kupca **i**;

$\mathbf{k_i}$ = broj narudžbi od Kupca **i**;

$\mathbf{D_i}$ = broj lokacija isporuke za Kupca **i**;

$\mathbf{S_{iN}}$ = skup proizvoda koje Kupac **i** prima kao normalne isporuke od poduzeća $\mathbf{P_1}$ (isporuka distributeru);

$\mathbf{S_{iD}}$ = skup proizvoda koje Kupac **i** prima kao direktne isporuke od poduzeća $\mathbf{P_1}$ (direktna isporuka maloprodaji);

$\mathbf{S_{iT}}$ = skup svih proizvoda koje je Kupac **i** kupio od poduzeća $\mathbf{P_1}$;

$\mathbf{P_{iS}}$ = neto cijena proizvoda Kupcu **i** za Proizvod **s**;

$\mathbf{K_S}$ = broj narudžbi koje je poduzeće $\mathbf{P_1}$ dobilo za Proizvod **s**;

$\boldsymbol{\beta_S}$ = dio potražnje za proizvodom **s** koju poduzeće $\mathbf{P_1}$ drži na zalihi;

$\mathbf{C_S}$ = trošak proizvoda **s** za poduzeće $\mathbf{P_1}$ prema dobavljaču robe;

$\mathbf{Vol_S}$ = fizička zapremina po jedinici proizvoda **s** izražena u npr. m^3.

Kako Niraj et al. koriste model za izračun tekuće profitabilnosti u poduzeću koje se isključivo bavi distribucijom i prodajom robe, ali ne i vlastitom proizvodnjom, model će se proširiti varijablom $\mathbf{CP_S}$. Ona će predstavljati vrijednost troška proizvodnje za proizvode iz vlastitog asortimana poduzeća $\mathbf{P_1}$.

U svom modelu Niraj et al. definiraju sedam grupa indirektnih troškova koje pridružuju kupcima kao i jednadžbe za njihovu alokaciju.

I. Troškovi prodaje i direktnog marketinga.

Prodajno osoblje troši svoje radno vrijeme u rutinskim posjetima kupaca, prihvatu njihovih narudžbi, generiranju narudžbi za pojedine kupce koji ih ne generiraju samostalno itd. Svi ovi troškovi vremena, koje prodajno osoblje utroši u poslovima prodaje i direktnog marketinga, alociraju se na pojedinačnog kupca kao funkcija njegove veličine (prihoda) i kompleksnosti transakcija iskazane kroz broj različitih isporučenih proizvoda, broj lokacija isporuke i frekvenciju isporuka. Ako se udjeli troškova s obzirom na veličinu prihoda, broj isporučenih proizvoda, broj lokacija isporuke i frekvenciju isporuka označe sa $\mathbf{SF_1}$ do $\mathbf{SF_4}$, tada će jednadžba za trošak prodaje i direktnog marketinga za kupca **i** biti:

$$SC_{i1} = SF_1 \times (\sum_{s:s\in S_{iT}} P_{is} \times m_{is} \times N_i.) + SF_2 \times D_i + SF_3 \times S_{iT} + SF_4 \times k_i \qquad (5.1)$$

II. Troškovi procesiranja narudžbi kupaca.

Procesiranje svake pojedinačne narudžbe, bez obzira na njenu veličinu, podliježe skupu aktivnosti koji generiraju određeni iznos troškova. Ovi troškovi se

alociraju na pojedinačnog kupca na osnovi broja narudžbi. Dakle, označi li se troškovni udio u ovim troškovima s **OR**, a s $\mathbf{k_i}$ broj narudžbi kupca ***i*** u promatranom vremenskom periodu, trošak procesiranja narudžbi za kupca ***i*** izračunavati će se prema jednadžbi:

$$SC_{i2} = OR \times k_i \qquad (5.2)$$

III. Troškovi isporuke kupcu.

Isporuka kupcu generirana je aktivnostima koje proizvode troškove svrstane u četiri kategorije. Prva kategorija uključuje troškove vezane za aktivnosti poput lociranja narudžbe, skeniranja stavki, provjere točnosti itd., pri čemu će se troškovni udio u pojedinoj narudžbi označiti s $\mathbf{HS_1}$. Drugu kategoriju uključuju troškovi vezani za fizičko procesiranje pojedinačne stavke narudžbe poput internog transporta od mjesta skladištenja do prijevoznog sredstva, utovar na prijevozno sredstvo i slično. Udio troška se određuje na razini jedinice pojedinačne stavke narudžbe i označen je s $\mathbf{HS_2}$. Treću kategoriju čine troškovi procesiranja dokumenata koji prate isporuku kao što su otpremnica, izlazna faktura, transportni papiri. Udio troška u pojedinačnoj narudžbi označiti će se s **AR**. Zadnju kategoriju ovih troškova čine troškovi direktne isporuke u maloprodajne objekte i udio u pojedinoj takvoj narudžbi označiti će se s **DR**. Troškovi isporuke kupcu **i** izračunavati će se prema jednadžbi:

$$SC_{i3} = HS_1 \times k_i + HS_2 \times \sum_{s:s \in S_{iN}} m_{iS} \times N_i + AR \times k_i + DR \times \sum_{s:s \in S_{iD}} k_i \qquad (5.3)$$

IV. Troškovi održavanja dobavljača.

Poduzeće $\mathbf{P_1}$ generira tzv. troškove održavanja dobavljača (engl. *vendor maintenance*) koji uključuju neke poput identificiranja dobavljača, ugovaranja cijena

za proizvode i održavanja računa dobavljača. Troškovi ovog tipa alociraju se na kupce sukladno udjelu u isporučenim jedinicama proizvoda (koji je naručen od dobavljača i uskladišten prije same prodaje). Za razliku od Niraj et al. modela, poduzeće **P_1** određeni dio prodajnog asortimana samostalno proizvodi. Za ove proizvode poduzeće **P_1** je „dobavljač" te će se i oni kao i prethodno opisani troškovi rasporediti na kupce na isti način. Označi li se udio ovih troškova u jedinici proizvoda s **VM**, tada je jednadžba za njihov izračun na razini kupca **i** slijedeća:

$$SC_{i4} = VM \times \left[\sum_{s:s \in S_{iT}} \frac{m_{is} \times N_i}{\sum_{i=1}^{I} (m_{is} \times N_i)} \right] \qquad (5.4)$$

V. Troškovi pripreme narudžbi od dobavljača.

Priprema svake od narudžbi prema dobavljaču, od strane poduzeća **P_1**, generira određene troškove. Pretpostavka je da za poduzeće **P_1** postoji normalna frekvencija narudžbi svakog pojedinačnog proizvoda. Postoje određene specijalne narudžbe pri kojima se maloprodajnim kupcima vrše direktne isporuke od dobavljača poduzeća **P_1**. Poduzeće **P_1** je samo posrednik u isporuci ove robe. Dakle, ovi troškovi se razlikuju ako se rade za normalne isporuke ili pak za direktne isporuke. Alociraju se na kupce proporcionalno udjelu isporuke pojedinačnog proizvoda kupcu u odnosu na cjelokupnu isporuku istog. Označi li se udio ovog troška u svakoj narudžbi s **OI**, udio pojedinog kupca računa se sukladno udjelu u ukupnoj isporuci za pojedini proizvod. U obzir treba uzeti i direktne isporuke, pa je jednadžba za izračun troškova pripreme narudžbi od dobavljača za kupca ***i*** slijedeća:

$$SC_{i5} = OI \times \left[\sum_{s:s \in S_{iT}} k_s \times \frac{m_{is} \times N_i}{\sum_{i=1}^{I} (m_{is} \times N_i)} + \sum_{s:s \in S_{iD}} k_i \right] \qquad (5.5)$$

VI. Troškovi manipulacije i prijevoza robe od dobavljača.

Ovi troškovi dijele se u tri kategorije. Prvu čine troškovi zaprimanja narudžbe i provjere njene točnosti. Udio ovih troškova označen je s $\mathbf{HR_1}$. Sljedeću kategoriju, $\mathbf{HR_2}$ čine troškovi fizičkog manipuliranja pojedinačnim proizvodima i promatraju se na transakcijskoj razini za pojedinačni proizvod. Svaku od isporuka robe od dobavljača, prate i troškovi na izradi papirnate dokumentacije vezane za isporuku. Udio ovih troškova označava se s **AP** po isporuci. Alociranje troškova manipuliranja i prijevoza robe od dobavljača na kupca ***i*** računa se prema jednadžbi:

$$SC_{i6} = HR_1 \times \left[\sum_{s:s \in S_{iT}} k_s \times \frac{m_{is} \times N_i}{\sum_{i=1}^{I} (m_{is} \times N_i)} \right] + HR_2 \times \left[\sum_{s:s \in S_{iN}} (m_{is} \times N_i) \right] + AP \times \left[\sum_{s:s \in S_{iT}} k_s \times \frac{m_{is} \times N_i}{\sum_{i=1}^{I} (m_{is} \times N_i)} + \sum_{s:s \in S_{iD}} k_i \right] \quad (5.6)$$

VII. Troškovi skladišnog prostora i troškovi kapitala.

Troškovi skladišnog prostora izračunavaju se prema jedinici zapremine prostora koja čini skladište, a označeni su s **ST**. **HO** označava trošak kapitala po jedinici valute uskladištene robe. Jednadžba za izračun troškova skladišnog prostora i kapital po kupcu ***i*** je:

$$SC_{i7} = ST \times \left[\sum_{s:s \in S_{iN}} Vol_s \times \beta_s \times m_{iS} \times N_i \right] + HO \times \left[\sum_{s:s \in S_{iN}} C_s \times \beta_s \times m_{is} \times N_i \right] \quad (5.7)$$

Nirajev model ne alocira troškove investicija u nove kupce, troškove oglašavanja i drugih marketinških aktivnosti poput onih na izgradnji brenda i informiranosti kupaca. Ovi troškovi imaju značajnost u dužem vremenskom periodu, ali svakako nije beznačajna i njihova značajnost za period u kojem se izračunava profitabilnost pojedinačnog kupca. Iz navedenog razloga, proširiti će se polazni model na način da uključi i ove troškove u profitabilnost kupaca.

VIII. Troškovi investicija u nove kupce, oglašavanje i druge marketinške aktivnosti

Raspored ovih troškova na pojedine kupce napravljen je sukladno udjelu pojedinog kupca u ukupnom prihodu. Ako se s **OC** označi jedinica ovih troškova po jedinici ostvarenog prihoda, tada je jednadžba za raspored ovih troškova na kupca:

$$SC_{i8} = OC \times \sum_{s:s \in S_{iT}} P_{is} \times m_{is} \times N_i . \qquad (5.8)$$

Nakon prikaza svih osam grupa troškova, ukupna vrijednost troškova za pojedinačnog kupca ***i,*** u promatranom vremenskom periodu t, dobije se jednadžbom:

$$CSC_i = \sum_{j=1}^{8} SC_{ij} \qquad (5.9)$$

Ako se s $\mathbf{TR_i}$ (jednadžba 5.10) označi ukupan prihod ostvaren prodajom proizvoda kupcu ***i*** u promatranom vremenskom periodu, s $\mathbf{CTR_i}$ (jednadžba 5.11) vrijednost robe prodane kupcu **i** iz asortimana **trgovačke robe** (tj. robe koja ne podliježe internim proizvodno-prerađivačkim procesima), a s $\mathbf{CVP_i}$ (jednadžba 5.12) vrijednost robe prodane kupcu ***i*** iz asortimana **vlastitih proizvoda** (tj. asortimana koji se proizvodi/pakira u vlastitim prerađivačkim pogonima), tada se prihod kupca ***i*** u promatranom vremenskom periodu ***t***,

označen s $\mathbf{CP_i}$, može izračunati umanjivanjem ukupnog prihoda za troškove robe (jednadžba 5.13).

$$TR_i = \sum_{s:s \in S_{iN}} P_{is} \times m_{is} \times N_i \qquad (5.10)$$

$$CTR_i = \sum_{S:S \in S_{iT}} C_{iS} \times m_{iS} \times N_i \qquad (5.11)$$

$$CVP_i = \sum_{S:S \in S_{iT}} CP_{iS} \times m_{iS} \times N_i \qquad (5.12)$$

$$CP_i = TR_i - CTR_i - CVP_i \qquad (5.13)$$

Niraj et al. u opisu svog modela, kao nedostatak, naglašavaju linearnost rasporeda troškova. Ona postavlja upitnim model sa stanovišta točnosti. Ipak, činjenica je da u postojećim informacijskim sustavima postoji nedostatak odgovarajućih podataka potrebitih za točno izračunavanje troškova pojedinačnog kupca. Model uvelike nadomješta ovo ograničenje, uključujući osobitosti pojedinačnog kupca iskazane kroz veličinu kupnje, cjenovnu marginu, te faktore kompleksnosti i efikasnosti kupca. Kvaliteta ovakvog modela za izračun tekuće profitabilnosti, dokazana je (Niraj et al., 2001) empirijskom usporedbom s tri dodatna modela (BASE, MULH i REDC). Zaključak je da se model pozitivno ponaša u odnosu na uspoređivane modele iz marketinške literature, te da ukazuje na značajnu povezanost između karakteristika kupca i njegove profitabilnosti. Ovaj zaključak je bitan sa stanovišta izgradnje modela za predviđanja profitabilnosti: osnovu ovakvih modela čine točni podaci iz prošlosti na temelju kojih sustav uči i predviđa buduće ponašanje.

Na osnovama modificiranog Niraj modela, definirana je prva grupa nezavisnih varijabli modela istraživanja. Ona sadrži osam nezavisnih varijabli kojima su na razini pojedinačnog kupca iskazani troškovi nastali u različitim fazama poslovnog odnosa pojedinačnog kupca i poduzeća, a koristiti će se za predviđanje profitabilnosti:

T_1	Trošak prodaje i direktnog marketinga
T_2	Trošak procesiranja narudžbi kupaca

T_3	Trošak isporuke kupcu
T_4	Trošak nabave od dobavljača i skladištenja
T_5	Trošak pripreme narudžbi od dobavljača
T_6	Trošak manipulacije i prijevoza robe od dobavljača
T_7	Trošak skladišnog prostora i troškovi kapitala
T_8	Trošak investicija u nove kupce i oglašavanje

5.1.2 Cjenovna politika (L2)

Sljedeća grupa ulaznih podataka vezana je za cijenu proizvoda, te s njom povezanu bruto maržu. Umnožak cijene i količine kupljenih proizvoda, od strane pojedinačnog kupca, daje ukupan prihod ostvaren u prometovanju s njim. Pri tome, cijena nije fiksna kategorija već se razlikuje od kupca do kupca. Nezavisne varijable koje definiraju odstupanje pojedinačnih cijena po kupcima od standardnih cijena proizvoda čine ovu grupu varijabli. Promatrano poduzeće u internim pravilima poslovanja razlikuju tri kategorije rabata, te su sukladno internom poslovnom pravilu određivanja prodajnih cijena, definirane sljedeće ulazne varijable modela:

C_1	Individualni rabat
C_2	Teritorijalni rabat
C_3	Količinski rabat

Prva varijabla sadrži ukupni rabat koji kupac dobije u pojedinačnim transakcijama na osnovu prethodno ugovorenih stopa rabata u dogovaranju poslovnog odnosa. Teritorijalni rabat čini iznos rabata dobiven na osnovu teritorijalne pripadnosti kupca. Politika poduzeća je da stimulira kupovinu u određenim regijama svog tržišta davanjem dodatnih rabata. Uobičajena praksa u ovom načinu poslovanja je da se ugovaraju i količinski rabati koje kupac ostvaruje kupovinom iznad dogovorene količine/vrijednosti određenih proizvoda.

Sve tri varijable dobivaju se iz transakcijskih podataka o prodaji i izražene su u apsolutnoj vrijednosti domicilne valute.

5.1.3 Značajke kupca (L3)

U predviđanju profitabilnosti kupca svakako da bitnu komponentu čine značajke pojedinačnog kupca. Postoji niz varijabli koje opisuju samog kupca, a za potrebe istraživanja kao značajne utvrđene su karakteristike veličine (VK), kompleksnosti (KK) i ponašanja pri plaćanju obveza (PO). Sukladno ovoj kategorizaciji, nezavisne varijable sustava čine isto tako nazvane tri grupe. Do podataka se dolazi agregiranjem podataka o financijskim i robnim tijekovima iz transakcijskih podataka poduzeća, te preuzimanjem pojedinih atributa podataka o kupcima iz baze kupaca.

5.1.3.1 Veličina kupca

Veličina kupca direktno je povezana s kategorijom troškova. Pod veličinom se ovdje podrazumijeva veličina prometa koji kupac ostvaruje. Kako je opisano u troškovnom modelu varijabli, određeni troškovi variraju s količinom kupljenih proizvoda, drugi s brojem transakcija, a treći pak s nekim karakteristikama povezanim s veličinom kupca. Iz navedenog definirane su varijable koje su direktno povezane s količinom prometa (VK_1 i VK_2) i one koje se odnose na obilježja kupca bez obzira na njegov promet (VK_3 i VK_4):

VK_1	Veličina prometa s kupcem
VK_2	Bruto marža ostvarena u prometovanju s kupcem
VK_3	Veličina kupca (malo/srednje/veliko poduzeće)
VK_4	Domicilna/inozemna kompanija

Vrijednosti za prve dvije od navedenih varijabli, dobivene su agregiranjem podataka o financijskim i robnim tijekovima iz transakcijskih podataka poduzeća, dok su ostale dvije dobivene preuzimanjem pojedinih atributa podataka o kupcima iz baze kupaca. Veličina prometa i bruto marža su izražene u apsolutnoj vrijednosti domicilne novčane jedinice, te su kao takve nepogodne za upotrebu u neuronskoj mreži i metodama multivarijacijske analize.

Njihova normalizacija provedena je izračunom udjela u ukupnoj vrijednosti za promatrani vremenski.

Veličina kupca (VK_3) određena je važećim zakonom o privrednim subjektima, te su vrijednosti varijable određene nominalnom mjernom ljestvicom sa sljedećim vrijednostima:

1 – Malo poduzeće,
2 – Srednje poduzeće,
3 – Veliko poduzeće.

Ovakve varijable nije potrebno normalizirati za upotrebu u neuronskim mrežama. One se definiraju kao simboličke varijable, a softver u pripremi podataka generira nove varijable. Generira ih onoliko koliko ima mogućih vrijednosti nominalne ljestvice, naziva ih prema opisu iz ljestvice i dodjeljuje vrijednost 0 ili 1.

Na sličan način definirana je i varijabla teritorijalne pripadnosti kupca (VK_4). Nominalna ljestvica koja opisuje ovaj atribut kupca ima također tri vrijednosti:

1 – Domicilno poduzeće,
2 – Inozemno poduzeće iz EU,
3 – Inozemno poduzeće van EU.

Za upotrebu ovih podataka vrijedi isto pravilo kao i za prethodno opisanu VK_3 varijablu.

5.1.3.2 Kompleksnost kupca

Faktori kompleksnosti kupca odnose se na karakteristike kupca koje nisu povezane s veličinom prometa. U istraživanju su odabrane karakteristike koje bi trebale direktno utjecati na povećanje troškova isporuke kupcu:

KK_1	Broj različitih kupljenih proizvoda
KK_2	Broj narudžbi koje kupac napravi u promatranom vremenskom periodu
KK_3	Broj lokacija isporuke

KK_4	Vrijednost povrata robe
KK_5	Broj transakcija povrata robe u promatranom vremenskom periodu

Varijabla KK_3 poprima vrijednosti iz nominalne ljestvice s vrijednostima 1,2 i 3. Vrijednost 1 podrazumijeva da promatrano poduzeće ima samo jednu lokaciju isporuke, vrijednost 2 odnosi se na 2-10 lokacija isporuke, a vrijednost 3 na poduzeća s preko 10 lokacija isporuke.

Ostale varijable poprimaju apsolutne vrijednosti izražene u domicilnoj novčanoj jedinici, a dobivene su kao i ostale ovog tipa, agregiranjem podataka o financijskim i robnim tijekovima iz transakcijskih podataka poduzeća. Normalizacija varijabli je napravljena iskazivanjem udjela u ukupnoj vrijednosti pojedinačnog pokazatelja za promatrani vremenski period.

5.1.3.3 Ponašanje kod plaćanja

Definirane su sljedeće varijable koje karakteriziraju navike i ponašanje kod plaćanja pojedinog kupca:

PP_1	Prosječno vrijeme kašnjenja u plaćanju kupca
PP_2	Iznos potraživanja naplaćen na vrijeme
PP_3	Iznos potraživanja naplaćen s zakašnjenjem do 30 dana
PP_4	Iznos potraživanja naplaćen s zakašnjenjem do 90 dana
PP_5	Iznos potraživanja naplaćen s zakašnjenjem do 180 dana
PP_6	Iznos potraživanja naplaćen s zakašnjenjem preko 180 dana
PP_7	Dospjeli dug kupca na kraju promatranog razdoblja

Prosječno vrijeme kašnjenja u plaćanju obveza definirano je kao vrijeme manje (vrijednost 0) ili veće (vrijednost 1) od prosječnog vremena kašnjenja, a ostale vrijednosti iskazane su u apsolutnom iznosu novčane jedinice.

5.2. Dimenzije profitabilnosti

Izlazne varijable su proizvod modela za predviđanje. Njihova definicija u ovom istraživanju zasnovana je na RFM modelu (Hughes, 1994), a imenovanje je izvršeno sukladno definiciji modela:

- R varijabla
- F varijabla
- M varijabla.

R varijabla se odnosi na vremenski interval između kupovina pojedinačnog kupca, *F varijabla* se odnosi na broj kupovina i *M varijabla* predstavlja monetarnu vrijednost kupovine.

Osnovu za predviđanje profitabilnosti kupca čine podaci iz vremenskog perioda t kojeg čini jedna kalendarska godina. Predviđanje se vrši za budući vremenski period t+1 (slijedeća godina u odnosu na podatke koji čine osnovu za predviđanje) u kome kupac možda više neće niti biti kupac, odnosno u prometovanju promatranog poduzeća neće učestvovati u kupoprodajnom odnosu. U tijesnoj vezi s (ne)aktivnosti kupca su stopa zadržavanja kupaca i njoj suprotna stopa odlazaka kupaca. One stavljaju u odnos broj zadržanih, odnosno izgubljenih kupaca s ukupnom bazom kupaca određenog perioda.

Ukoliko kupac nastavi ovu ulogu u narednom periodu (R varijabla), zanimljivo je predvidjeti broj transakcija koje će nastati u tom odnosu (F varijabla). Ovaj pokazatelj sadrži sintezu stope zadržavanja, ponavljanja kupnje i rasta (ili pada) računa pojedinačnog kupca.

Jedan od financijskih pokazatelja budućeg prometa je i očekivani promet (M varijabla). Varijabla je to koja u sebi sadrži više komponenti. Svakako je bitan miks proizvoda kojima se prometuje. Količina je bitan činitelj, ali i cjenovna politika u odnosu s pojedinačnim kupcem.

5.3 Normalizacija podataka

U istraživanju se koriste metode višestruke logističke regresije za predviđanje M varijable, te binarne logističke regresije za predviđanje F i R varijabli. Metoda neuronske mreže koristi se za predviđanje sve tri izlazne varijable modela, te se rezultati istraživanja uspoređuju u cilju dokazivanja postavljenih hipoteza sustava koje stavljaju u odnos metode multivarijacijske statističke analize i strojnog učenja.

Sve pobrojane metode su matematičke metode, te se sve ulazne i izlazne varijable modela, koje su nenumeričke, moraju pretvoriti u numeričke. Istovremeno, sve numeričke varijable modela potrebno je normalizirati, odnosno svesti na raspon od 0 do 1, odnosno od -1 do 1.

Normalizacija **kategorijalnih** varijabli rađena je formiranjem novih binomnih varijabli u broju jednakom broju kategorija promatrane varijable. Ovako definirane varijable poprimaju vrijednost 0 ili 1 ovisno da li vrijednost varijable odgovara promatranoj kategoriji. Npr. VK3-Veličina kupca može poprimiti vrijednosti 1,2 ili 3 (1-malo poduzeće, 2-srednje poduzeće ili 3-veliko poduzeće). Formirane su nove varijable VK_{31} (0 za VK_3<>1, 1 za VK_3=1), VK_{32} (0 za VK_3<>2, 1 za VK_3=2) i VK_{33} (0 za VK_3<>3, 1 za VK_3=3).

Normalizacija **numeričkih** varijabli modela odrađena je svođenjem na raspon vrijednosti od -1 do 1, upotrebom sljedeće jednadžbe:

$$Var_x = 2 * \left[\frac{(\mathrm{var}_x - \min(\mathrm{var}_x))}{(\max(\mathrm{var}_x) - \min(\mathrm{var}_x) + 1)} \right] - 1 \qquad (5.14)$$

pri čemu je

Var$_x$ normalizirana numerička varijabla iz sustava ulaznih varijabli,

var$_x$ originalna vrijednost promatrane varijable za pojedinačnog kupca u periodu t

max(var$_x$) maksimalna vrijednost varijable var_x po svim kupcima za period t

min(var$_x$) minimalna vrijednost varijable var_x po svim kupcima za period t

Ulazne varijable sustava podijeljene su u 3 grupe, te analogno nazivu i označene velikim slovima abecede:

- Troškovi (**T**)
- Cjenovna politika (**C**)
- Značajke kupca
    - Veličina kupca (**VK**)
    - Kompleksnost kupca (**KK**)
    - Ponašanje kod plaćanja obveza (**PP**)

T – Troškovi

Ako se na razini vremenskog perioda **t,** koji odgovara jednoj kalendarskoj godini, minimalan iznos pojedinačnih kategorija troškova u poduzeću označi s **$Tmin_1$ .. $Tmin_8$**, maksimalan iznos pojedinačnih kategorija troškova u poduzeću označi s **$Tmax_1$ .. $Tmax_8$**, te ako se iznos pojedinačnog troška za kupca **i** označimo s **t_i**, tada će vrijednost ulaznih varijabli grupe troškova biti izražena formulom

$$T_{i=1..8} = 2 * \left[\frac{(t_i - T_{\max i})}{(T_{\max i} - T_{\min i} + 1)} \right] - 1 \qquad (5.15)$$

Sve varijable grupe troškova prikazane su tablicom 5.1.

Tablica 5.1 Nezavisne varijable grupe troškova

Oznaka	Opis	Vrijednosti
T_1	Troškovi prodaje i direktnog marketinga	$-1 \le T_1 \le 1$
T_2	Troškovi procesiranja narudžbi kupaca	$-1 \le T_2 \le 1$
T_3	Troškovi isporuke kupcu	$-1 \le T_3 \le 1$
T_4	Troškovi nabave od dobavljača i skladištenja	$-1 \le T_4 \le 1$
T_5	Troškovi pripreme narudžbi od dobavljača	$-1 \le T_5 \le 1$
T_6	Troškovi manipulacije i prijevoza robe od	$-1 \le T_6 \le 1$

	dobavljača	
T_7	Troškovi skladišnog prostora i troškovi kapitala	$-1 \leq T_7 \leq 1$
T_8	Troškovi prodane robe	$-1 \leq T_8 \leq 1$

C – Cjenovna politika

Ako se na razini vremenskog perioda **t** minimalan iznos pojedinačnih kategorija rabata u poduzeću označi s $\mathbf{Cmin_1}$, $\mathbf{Cmin_2}$, $\mathbf{Cmin_3}$, maksimalan iznos pojedinačnih kategorija rabata u poduzeću označi s $\mathbf{Cmax_1}$, $\mathbf{Cmax_2}$, $\mathbf{Cmax_3}$ te ako se iznos pojedinačnog rabata za kupca **i** označi s $\mathbf{c_i}$, tada će vrijednost ulaznih varijabli grupe rabata biti izražena formulom

$$C_{i=1..3} = 2 * \left[\frac{\left(c_i - C_{\max i}\right)}{\left(C_{\max i} - C_{\min i} + 1\right)} \right] - 1 \qquad (5.16)$$

Tablica 5.2 Nezavisne varijable grupe cijena

Oznaka	Opis	Vrijednosti
C_1	Individualni rabat	$-1 \leq C_1 \leq 1$
C_2	Teritorijalni rabat	$-1 \leq C_2 \leq 1$
C_3	Količinski rabat	$-1 \leq C_3 \leq 1$

Značajke kupca

U skupu varijabli koje opisuju značajke kupca, četiri su kategorijalne varijable $\mathbf{VK_3}$, $\mathbf{VK_4}$, $\mathbf{KK_3}$ i $\mathbf{PP_1}$. Ostale varijable su numeričke i njihova normalizacija urađena je prilagodbom jednadžbe 5.14. Tablica 5.3 daje pregled svih varijabli ove grupe s pripadajućim skupovima vrijednosti koje poprimaju.

Tablica 5.3 Nezavisne varijable grupe značajki kupaca

Oznaka	Opis	Način izračuna	Vrijednosti
Veličina kupca			
VK_1	Veličina prometa s kupcem	(jednadžba 5.14)	$-1 \leq VK_1 \leq 1$
VK_2	Bruto marža ostvarena u prometovanju s kupcem	(jednadžba 5.14)	$-1 \leq VK_2 \leq 1$
VK_3	Veličina kupca	1-malo 2-srednje 3-veliko	$VK_3 \in \{1,2,3\}$
VK_4	Domicilna/inozemna kompanija	1-domicilna 2-inozemna EU 3-inozemna nonEU	$VK_4 \in \{1,2,3\}$
Kompleksnost kupca			
KK_1	Broj različitih kupljenih proizvoda	(jednadžba 5.14)	$-1 \leq KK_1 \leq 1$
KK_2	Broj narudžbi koje kupac napravi u promatranom vremenskom periodu	(jednadžba 5.14)	$-1 \leq KK_2 \leq 1$
KK_3	Broj lokacija isporuke	1 – jedna 2 –od 2 do 10 3- preko 10	$KK_3 \in \{1,2,3\}$
KK_4	Vrijednost povrata robe	(jednadžba 5.14)	$-1 \leq KK_4 \leq 1$
KK_5	Broj transakcija povrata robe u promatranom vremenskom periodu	(jednadžba 5.14)	$-1 \leq KK_5 \leq 1$
Ponašanje kod plaćanja			
PP_1	Prosječno vrijeme kašnjenja u plaćanju kupca	**0**-manje od medijan vrijednosti vremena za sve kupce **1**-veće ili jednako od medijan vrijednosti vremena za sve kupce	$PP_1 \in \{0,1\}$
PP_2	Iznos potraživanja naplaćen na vrijeme	(jednadžba 5.14)	$-1 \leq PP_2 \leq 1$
PP_3	Iznos potraživanja naplaćen s zakašnjenjem do 30 dana	(jednadžba 5.14)	$-1 \leq PP_3 \leq 1$
PP_4	Iznos potraživanja naplaćen s zakašnjenjem do 90 dana	(jednadžba 5.14)	$-1 \leq PP_4 \leq 1$
PP_5	Iznos potraživanja naplaćen s	(jednadžba 5.14)	$-1 \leq PP_5 \leq 1$

	zakašnjenjem do 180 dana		
PP_6	Iznos potraživanja naplaćen s zakašnjenjem preko 180 dana	(jednadžba 5.14)	$-1 \leq PP_6 \leq 1$
PP_7	Dospjeli dug kupca na kraju promatranog razdoblja	(jednadžba 5.14)	$-1 \leq PP_7 \leq 1$

Izlazne varijable sustava (tablica 5.4) pokazatelji su poslovanja s kupcem u period t+1. Dakle, ulazni vektor podataka odnosi se na godinu t, a izlazni vektor podataka odnosi se na godinu t+1. Varijable **R i F** mogu poprimiti vrijednosti 0 i 1. Ako varijabla **R** ima vrijednost 0 to znači da promatrani kupac neće obavljati kupovinu u periodu **t+1**, a vrijednost 1 znači da promatrani kupac obavlja kupovinu i u spomenutom narednom periodu.

Ako s $\mathbf{F_{med}}$ označimo median vrijednost, odnosno broj koji dijeli skup različitih vrijednosti varijable **F** na dva jednaka dijela, tada uporabom ove vrijednosti određujemo i vrijednost varijable **F**. Varijabla F ima vrijednost 0 ako je broj kupovina promatranog kupca u periodu **t+1** manji od $\mathbf{F_{med}}$, odnosno vrijednost 1 ako je broj kupovina veći od $\mathbf{F_{med}}$.

Normalizacija vrijednosti varijable **M,** koja predstavlja novčani iskaz vrijednosti prometa pojedinačnog kupca u vremenskom periodu **t+1**, normalizirana je uporabom jednadžbe 5.14.

Tablica 5.4 Zavisne varijable modela

Oznaka	Opis	Način izračuna	Vrijednosti
R	Nastavak kupovine promatranog kupca	1-Da 0-Ne	$R \in \{0,1\}$
F	Broj kupovina	Usporedba s prosječnim brojem kupovina: *0-manji broj od median vrijednosti* *1-veći ili jednak od median vrijednosti*	$F \in \{0,1\}$
M	Novčani iskaz prometa promatranog kupca	(jednadžba 5.14)	$-1 \leq M \leq 1$

5.4 Statističke metode za modeliranje profitabilnosti kupca

Opisani model za predviđanje profitabilnosti kupaca podrazumijeva skup od 27 nezavisnih varijabli i tri zavisne varijable. Ovakav skup nezavisnih varijabli, zapravo više od dvije istovremeno, predmet su analiza multivarijacijske statističke analize. Cilj njihove primjene jeste analiza djelovanja više nezavisnih varijabli na jednu ili više zavisnih varijabli.

Odabir odgovarajuće statističke metode limitiran je pretpostavkama koje ona postavlja pred model. Jedna od pretpostavki jeste priroda zavisne varijable. M varijabla modela je numerička varijabla. Upravo kod višestruke linearne regresije zavisna varijabla je numerička i izražava se kao linearna kombinacija nezavisnih varijabli. R i F varijable poprimaju vrijednost iz skupa s dvije vrijednosti: 0 i 1. S obzirom na binarnu prirodu ovih varijabli, binarna logistička regresija je odgovarajuća metoda. U znanstvenim krugovima posljednjih godina logistička regresija postaje sve popularnija zamjenjujući konkurentsku diskriminacijsku analizu (Rozga, 2010).

Prva hipoteza istraživanja pretpostavlja da metode neuronske mreže mogu predvidjeti profitabilnost kupaca jednako dobro ili bolje od statističkih metoda logističke i višestruke linearne regresije. Druga hipoteza tvrdi da politika koju poduzeće vodi prema pojedinom kupcu direktno utječe na njegove mjere profitabilnosti. Naredno izlaganje predstavlja način na koji su dvije spomenute metode multivarijantne statističke analize korištene u istraživanju.

5.4.1 Višestruka linearna regresija

Metoda višestruke linearne regresije koristi se u dijelu istraživanja koji pokušava dokazati obje glavne hipoteze istraživanja. Za potrebe prve hipoteze vršiti će se predviđanje zavisne varijable M. Za provođenje ovog dijela istraživanja koristi se Weka programski paket. Osnovne polazne postavke modela regresije su sljedeće:

1. Ulazne varijable regresije su sve varijable definirane okvirom za profitabilnost kupca u periodu t.
2. Zavisna varijabla sustava je podatak o relativnom udjelu kupovine kupca u vremenskom periodu t+1.

3. Enter metoda se koristi kao metoda za ulazak varijabli u sustav. Metoda inicijalno uvodi sve nezavisne varijable u model.

Kako bi statističke i metode strojnog učenja stavili u iste uvjete operiranja nad podacima, formirana su dva skupa podataka. Set podatka za učenje obuhvaća podatke iz perioda 2008. - 2011. godine, a služi za formiranje regresijskog modela. Testni set čine podaci iz 2013. godine na koje se primjenjuje formirani model, a rezultati predviđanja se potom uspoređuju s rezultatima koji se dobiju uporabom metode neuronske mreže.

Druga hipoteza istraživanja dokazuje se kroz dvije pomoćne hipoteze upotrebom SPSS programskog paketa. Poslovna politika, u smislu pojma kojeg definira ovo istraživanje, ima dvije komponente: (1) troškove i (2) cjenovnu politiku. Kako bi analizirali efekte ovih grupa nezavisnih varijabli na komponentu M profitabilnosti kupca, a time i dokazali ili opovrgnuli prvu pomoćnu hipotezu, definirana su dva regresijska modela:

$$M = f(T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8) \qquad (5.17)$$

$$M = f(C_1, C_2, C_3) \qquad (5.18)$$

I kod ovih modela se koristi Enter metoda za ulazak, inicijalno svih, varijabli u sustav.

Značajke kupca definirane modelom odnose se na: (1) veličinu, (2) kompleksnost i (3) ponašanje kupca pri plaćanju obveza. Analiziranje efekata ovih grupa nezavisnih varijabli na veličinu profitabilnosti kupca u periodu t+1, provedeno je kroz sljedeće regresijske modele:

,

$$M = f(VK_1, VK_2, VK_3, VK_4) \qquad (5.19)$$

$$M = f(KK_1, KK_2, KK_3, KK_4, KK_5) \qquad (5.20)$$

$$M = f(PP_1, PP_2, PP_3, PP_4, PP_5, PP_6, PP_7) \qquad (5.21)$$

5.4.2 Binarna logistička regresija

Binarna logistička regresija koristi se za dokazivanje prve hipoteze istraživanja, pri čemu se njenom uporabom nastoji predviđati vrijednost R i F zavisnih varijabli modela. Kao i kod prethodne metode, formirana su dva skupa podataka: set za učenje s podacima iz perioda 2008. - 2011. godine te testni set s podacima iz 2013. godine. Model se formira učenjem nad prvim skupom podatka, a rezultati predviđanja se potom testiraju nad testnim setom podatka. Porgramski paket Weka korišten je za formiranje ova dva regresijska modela.

Osnovne postavke korištene za primjenu binarne logističke regresije u predviđanju R varijable su sljedeće:

1. Ulazne varijable regresije su sve varijable definirane okvirom za profitabilnost kupca u periodu t.
2. Zavisna varijabla sustava je podatak da li kupac nastavlja s prometom (biti kupac) u vremenskom periodu t+1.
3. Enter metoda se koristi kao metoda za ulazak varijabli iz sustava. Metoda inicijalno uvodi sve nezavisne varijable sustava u analizu bez njihovog ispadanja u narednim koracima.
4. Kod opcija regresije odabrane su sljedeće polazne postavke:
    a. Razina značajnosti za ulazak varijable u model je 0,05
    b. Razina značajnosti za izlazak varijable iz modela je 0,10
    c. Klasifikacijska vjerojatnost je postavljena na 0,5
    d. Maksimalni broj iteracija je postavljen na 200
    e. Odabrana je uporaba Hosmer-Lemeshow goodness-of-fit testa
5. Za referentne vrijednosti nezavisnih kategoričkih varijabli odabrane su prve vrijednosti:
    a. Varijabla ***KK3***: 1- Jedna lokacija isporuke
    b. Varijabla ***VK3***: 1- Malo poduzeće
    c. Varijabla ***VK4***: 1- Domicilna kompanija
    d. Varijabla ***PP1***: 0- Manje od prosječnog vremena

F varijabla mjeri broj kupovina u periodu t+1, a potom na osnovu usporedbe s medijan vrijednosti istog perioda poprima vrijednost 0 za manji broj kupovina, odnosno 1 za veći ili jednak broj kupovina. Osnovne postavke regresijskog modela pri predviđanju

varijable F vrijednosti su većinom iste kao i kod prethodnog modela. Razlika je u odabiru metode za ulazak/izlaza varijabli iz sustava. One su sljedeće:

1. Ulazne varijable regresije su sve varijable definirane okvirom za profitabilnost kupca u periodu t.
2. Zavisna varijabla sustava je podatak da li kupac nastavljajući s prometom, u vremenskom periodu t+1, ostvaruje takav promet da je on manji ili veći od medijan vrijednosti perioda.
3. Forward Stepwise Conditional metoda se koristi kao metoda za ulazak varijabli u sustav. Metoda uvodi u koracima pojedine nezavisne varijable modela.
4. Kod opcija regresije odabrane su sljedeće polazne postavke:
    a. Razina značajnosti za ulazak varijable u model je 0,05 (defaultna vrijednost)
    b. Razina značajnosti za izlazak varijable iz modela je 0,10 (defaultna vrijednost)
    c. Klasifikacijska vjerojatnost je postavljena na 0,5 (defaultna vrijednost)
    d. Maksimalni broj iteracija je postavljen na 200 (defaultna vrijednost)
    e. Odabrana je uporaba Hosmer-Lemeshow goodness-of-fit testa
5. Za referentne vrijednosti nezavisnih kategoričkih varijabli odabrane su prve vrijednosti:
    a. Varijabla ***KK3***: 1-Jedna lokacija isporuke
    b. Varijabla ***VK3***: 1-Malo poduzeće
    c. Varijabla ***VK4***: 1-Domicilna kompanija
    d. Varijabla ***PP1***: 0-Manje od prosječnog vremena

5.5 Modeliranje neuronske mreže za predviđanje profitabilnosti kupca

Istraživanje metodom NN odrađeno je upotrebom softverskog paketa NeuroSolutions ver.6.2, uz dodatak paketa NeuroSolutions for Excel. Prethodnica izgradnji modela je priprema podataka.

5.5.1 Varijable modela

Nezavisne varijable modela, odnosno ulazne varijable mreže, definirane su i prikazane u tablicama 5.1, 5.2 i 5.3. Nakon uvoza podataka u MS Excel, spomenute su varijable deklarirane kao ulazne varijable sustava. Varijable VK_3, VK_4, KK_3 i PP_1 su deklarirane kao simboličke kolone. Deklaracijom simboličkih kolona potpuno se normalizira model stvarajući od svake simboličke kolone skup novih kolona. Broj novih kolona jednak je brojnosti skupa iz kojeg varijabla poprima vrijednost. Npr. od kolone $VK_3 \in \{1,2,3\}$ nastati će tri nove varijable: $VK_3(1)$, $VK_3(2)$ i $VK_3(3)$. Ako varijabla u jednoj instanci podataka ima vrijednost $VK_3=1$, tada će simboličke kolone imati sljedeće vrijednosti $VK_3(1)=1$, $VK_3(2)=0$ i $VK_3(3)=0$. Ista logika odnosi se na sve simboličke varijable.

Neuronska mreža uči na osnovu primjera iz povijesti. Budući su ti primjeri inputi mreže, oni trebaju zadovoljiti određene kriterije kako bi bili dobra osnova za projektiranje mreže:

- Dovoljan broj podataka
- Adekvatnu prezentaciju uvjeta u kojima su nastali
- Uključenost svih ulaznih varijabli koje su u relaciji s nezavisnim varijablama
- Učenje pod nadzorom zahtijeva dobre izlazne varijable

Ulazni set, prije svega, treba biti sačinjen od dovoljnog broj primjera za učenje. Tablica 5.5 daje informaciju da kompletan set ulaznih podataka čini 1911 slučajeva iz vremenskog perioda od 6 uzastopnih godina. Nadalje, podaci trebaju adekvatno predstavljati normalne operacijske uvjete u kojima su i nastali. Kako bi se ispunio ovaj kriterij, u obzir su uzeti svi podaci bez ikakvog izostavljanja ili zanemarivanja određenih instanci podataka. Početni set podataka sadržavao je pored troškova, cijena i podataka o kupcima još dvije grupe podataka: o promatranom poduzeću i o poslovnom i socijalnom okruženju. Ovi posljednji podaci, na razini jedne kalendarske godine, isti su za svakog pojedinačnog kupca. Ovo je razlog vjerovanja da kao takvi nisu u direktnoj relaciji sa zavisnim varijablama, te su i ispali iz modela.

Zavisne varijable modela, odnosno varijable čije će se vrijednosti predviđati upotrebom neuronske mreže, definirane su modificiranim RFM modelom i prikazane u tablici

5.4. Varijable R i F su deklarirane kao simboličke varijable i definirane kao izlazne varijable sustava. Varijabla M je numerička varijabla koja je normalizacijom svedena na skup vrijednosti $-1 \leq M \leq 1$. Ovakva priroda izlaznih varijabli definira predviđanje R i F varijabli kao klasifikacijski problem, dok se predviđanje varijable M tretira kao regresijski problem.

Prema pretpostavci istraživanja, podaci iz 2008. - 2011. godine označeni su kao trening set podataka (65,08%), podaci iz 2012. godine kao validacijski set (17,01%), a oni iz 2013. godine kao set podataka za testiranje (17,11%) dobivene arhitekture.

Tablica 5.5 Definicija setova podataka za neuronsku mrežu

Rbr	Set podataka	Godina	Broj slogova	%
1	Učenje	2008-2011	1259	65,08%
2	Validacija	2012	325	17,01%
3	Test	2013	327	17,11%
			1911	100,00%

Promatrajući u cjelini set podataka, može se reći da je on čist, bez šuma i da sadrži dovoljno podataka da se provede istraživanje i postignu njegovi ciljevi.

5.5.2 Arhitektura mreže

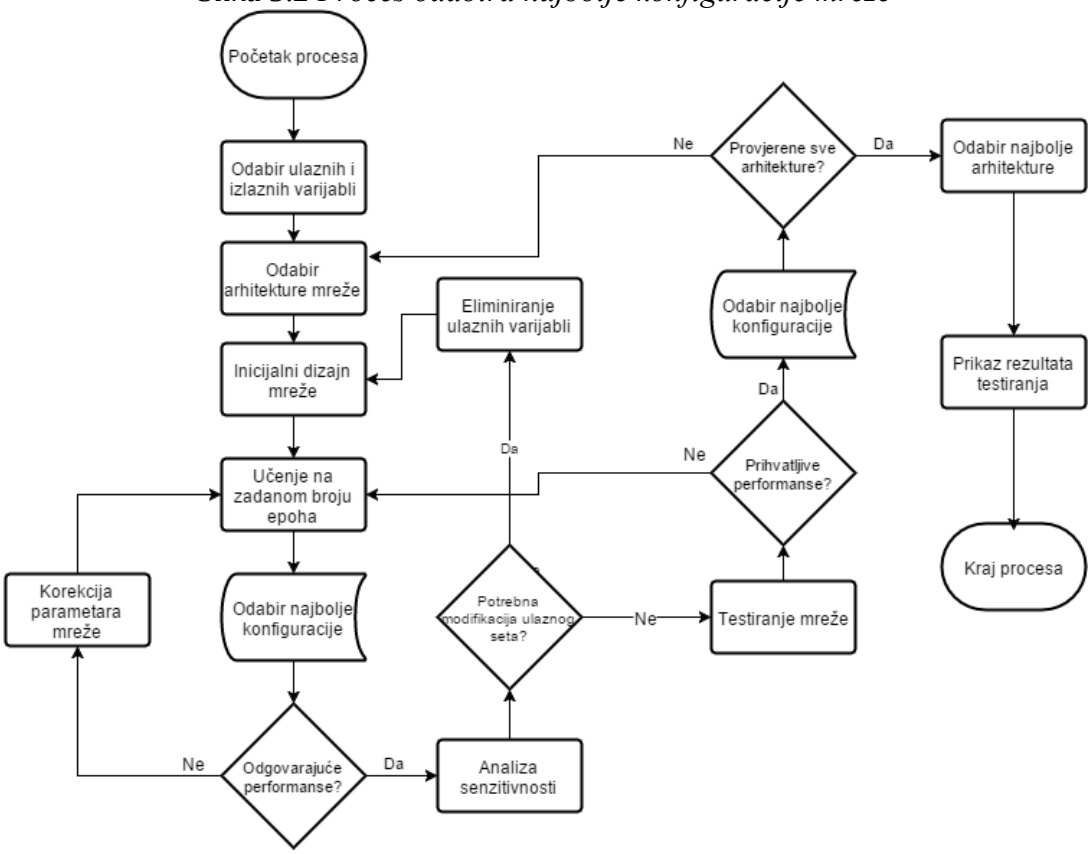
Po formiranja skupova podataka za učenje, testiranje i konačnu validaciju mreže, upotrebom NeuroSolutions paketa provodi se testiranje sljedećih arhitektura mreže:

1. MLP-Multi Layer Perceptron,
2. LR-Linear Regression,
3. PNN-Probabilistic Neural Network,
4. RBF-Radial Basis Function,
5. MLPPCA-Multi Layer Perceptron with Principle Component Analysis,
6. SVM- Classification Support Vector Machine,
7. GFF- Generalized Feedforward,

8. TDNN- Time-Delay Network,
9. TLRN- Time-Lag Recurrent Network,
10. RN- Recurrent Network.

Dijagram tijeka (slika 5.2) prikazuje korake koji se provode u testiranju pojedinih arhitektura mreže. Testiranje svake od njih započinje potpunim skupom ulaznih podataka. Formira se inicijalni dizajn mreže i provodi se učenje na 1000 epoha. Provedbom učenja odabire se konfiguracija koja ima najbolje pokazatelje.

Slika 5.2 Proces odabira najbolje konfiguracije mreže



Performanse sustava definirane su sljedećim pokazateljima:

a. MSE (engl. *Mean Squared Error*),
b. *NMSE* (engl. *Normalised Mean Square Error*),
c. Korelacijski koeficijent ***r*** (engl. *Correlation Coeficient*), oznaka ***r***,
d. Postotak pogreške ***%Error***,

e. *AIC* (engl. *Akaike's information criterion*).

Ovi pokazatelji će se koristiti za usporedbu modela generiranih metodama neuronske mreže i multivarijacijske statističke analize. Oni su pobliže opisani i definirani u poglavlju 5.5 koji opisuje načine komparacije modela izgrađenih metodama koje se uspoređuju u ovom istraživanju.

Jedna konfiguracija je definirana skupom težinskih faktora mreže. Za klasifikacijske probleme, kakvi su predviđanje R i F varijable, najbolja konfiguracija je ona kod koje je najveći postotak točnih predviđanja unutar CV skupa podataka. Za regresijske probleme, a u ovom istraživanju to je previđanje M varijable, najbolja konfiguracija je ona s najmanjim vrijednostima MSE i MAE uz najveći koeficijent korelacije *r*. Ukoliko nisu postignute zadovoljavajuće performanse mreže, vrši se korekcija parametara mreže i ponovo provodi učenje na testnom i CV setu podataka. Zadovoljavajuće performanse kandidiraju model za proces analize senzitivnosti.

Moguća eliminacija nekih od varijabli ulaznog skupa vraća proces na ponovno testiranje s novim setom ulaznih varijabli, u suprotnom se provodi testiranje modela na testnom skupu podataka. Ako prođe ova, konačna provjera modela, on ulazi u skup potencijalno najboljih arhitektura. Proces se nastavlja odabirom nove arhitekture sve do provjere svake od njih. Konačni korak procesa je odabir najbolje od provjerenih arhitektura i prikaz rezultata testiranja. U narednom tekstu prikazaće se rezultati testiranja arhitektura u predviđanju svake od pojedinačnih zavisnih varijabli sustava.

5.5.2.1 Topologija mreže za predviđanje R varijable

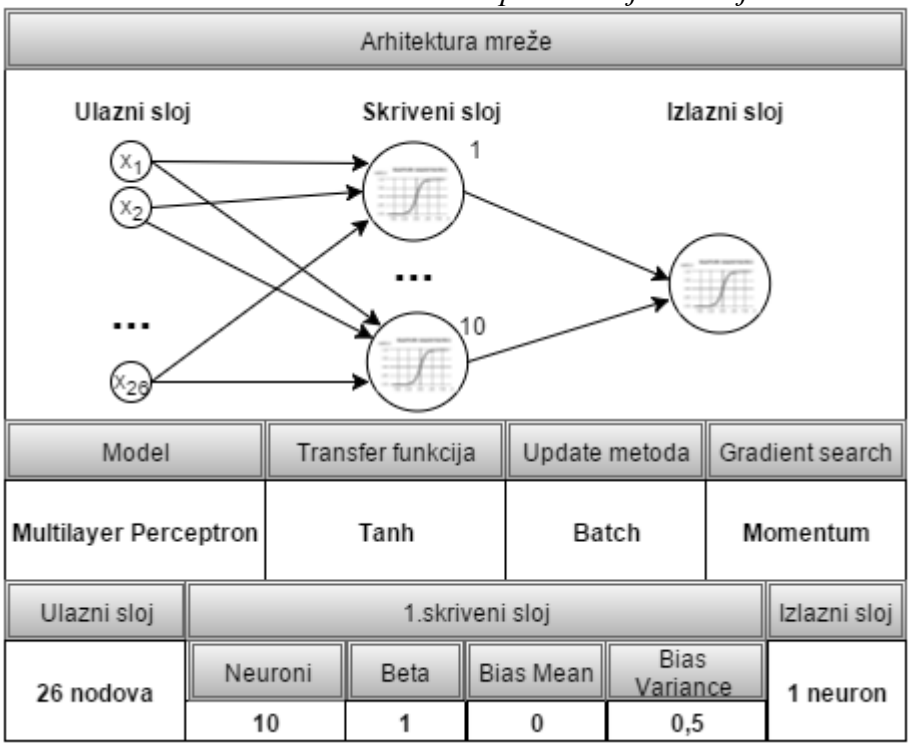
Kako R varijabla prima vrijednost iz skupa podataka koji sadrži vrijednosti 0 i 1, ovaj problem je definiran kao klasifikacijski problem.

Tablica 5.6 Rezultat usporedbe različitih arhitektura mreže u predviđanju R varijable

Performance Metrics

	Testing		
Model Name	**MSE**	**r**	**Correct**
MLP-1-O-M (Multilayer Perceptron)	0.235207	0.098317	69.11%
LR-0-B-M (Linear Regression)	0.213865	0.191351	69.11%
LR-0-B-L (Linear Regression)	0.222414	0.166637	69.42%
MLP-1-B-L (Multilayer Perceptron)	0.196784	0.37832	70.95%
PNN-0-N-N (Probabilistic Neural Network)	0.235803	0.223939	66.97%
RBF-1-B-L (Radial Basis Function)	0.221496	0.207875	69.42%
GFF-1-B-L (Generalized Feedforward)	0.207499	0.311349	70.34%
MLPPCA-1-B-L (MLP with PCA)	0.20972	0.294545	71.25%
SVM-0-N-N (Classification SVM)	0.209139	0.287497	70.34%
TDNN-1-B-L (Time-Delay Network)	0.226376	0.083033	69.11%
TLRN-1-B-L (Time-Lag Recurrent Network)	0.221134	0.100155	67.58%
RN-1-B-L (Recurrent Network)	0.320603	0.064212	49.54%
MLP-2-B-L (Multilayer Perceptron)	0.207415	0.30852	70.95%
MLP-1-B-M (Multilayer Perceptron)	0.207645	0.30914	71.56%
MLP-2-O-M (Multilayer Perceptron)	0.225575	0.138676	68.50%
MLP-2-B-M (Multilayer Perceptron)	0.217125	0.126189	67.58%
MLPPCA-1-O-M (MLP with PCA)	0.228347	0.231526	66.36%
MLPPCA-1-B-M (MLP with PCA)	0.225685	-0.04263	67.58%
GFF-1-O-M (Generalized Feedforward)	0.236556	0.294791	70.34%
GFF-1-B-M (Generalized Feedforward)	0.212802	0.195341	68.50%
RBF-1-O-M (Radial Basis Function)	0.224934	0.047447	67.58%
RBF-1-B-M (Radial Basis Function)	0.22485	-0.12237	67.58%
TDNN-1-O-M (Time-Delay Network)	0.228788	0.075211	67.58%
TDNN-1-B-M (Time-Delay Network)	0.221115	0.078367	67.58%
RN-1-O-M (Recurrent Network)	0.212384	0.259172	70.64%
RN-1-B-M (Recurrent Network)	0.240265	0.031716	68.20%
TLRN-1-O-M (Time-Lag Recurrent Network)	0.236978	-0.00157	66.06%
TLRN-1-B-M (Time-Lag Recurrent Network)	0.22676	0.006721	67.58%

Slika 5.3 Arhitektura mreže za predviđanje R varijable



Proces učenja različitih arhitektura NN pokazao je višeslojni preceptron (MLP-1-B-L) s jednim skrivenim slojem, back propagation algoritmom učenja uz uporabu Momentum pravila učenja, kao najbolju arhitekturu za promatrani set podataka (tablica 5.6).

5.5.2.2 Topologija mreže za predviđanje F varijable

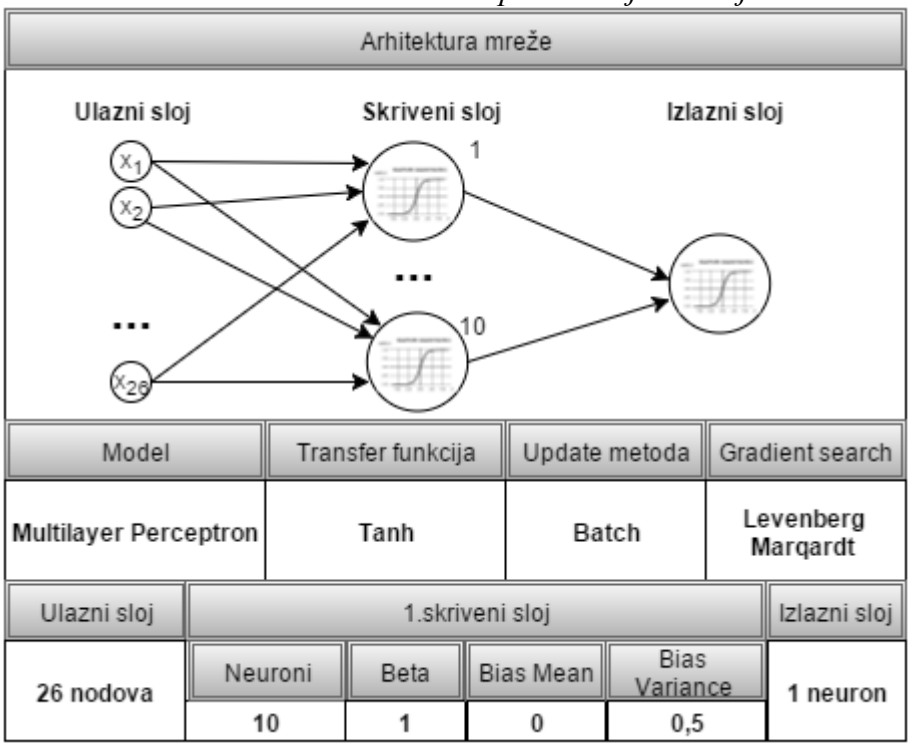
F varijabla, baš kao i R iz prethodnog poglavlja, poprima vrijednost iz skupa podataka koji sadrži vrijednosti 0 i 1. Problem je stoga u postavkama za učenje mreže prepoznat kao klasifikacijski problem.

Tablica 5.7 Rezultat usporedbe različitih arhitektura mreže u predviđanju F varijable

Performance Metrics

	Testing		
Model Name	**MSE**	**r**	**Correct**
MLP-1-O-M (Multilayer Perceptron)	0.197086	0.499196	72.17%
LR-0-B-M (Linear Regression)	0.191061	0.532916	70.95%
LR-0-B-L (Linear Regression)	0.213333	0.428729	71.56%
MLP-1-B-L (Multilayer Perceptron)	0.110982	0.747695	87.16%
PNN-0-N-N (Probabilistic Neural Network)	0.196113	0.50986	71.87%
RBF-1-B-L (Radial Basis Function)	0.170484	0.581361	74.92%
GFF-1-B-L (Generalized Feedforward)	0.133042	0.696311	83.49%
MLPPCA-1-B-L (MLP with PCA)	0.246876	0.359287	46.79%
SVM-0-N-N (Classification SVM)	0.193318	0.480917	66.67%
TDNN-1-B-L (Time-Delay Network)	0.210794	0.42062	68.20%
TLRN-1-B-L (Time-Lag Recurrent Network)	0.250949	0.026798	52.91%
RN-1-B-L (Recurrent Network)	0.23384	0.357749	51.68%
MLP-2-B-L (Multilayer Perceptron)	0.123823	0.713831	85.63%
MLP-1-B-M (Multilayer Perceptron)	0.166203	0.577421	76.45%
MLP-2-O-M (Multilayer Perceptron)	0.182847	0.520525	72.17%
MLP-2-B-M (Multilayer Perceptron)	0.169321	0.572325	74.31%
MLPPCA-1-O-M (MLP with PCA)	0.267174	0.103033	46.48%
MLPPCA-1-B-M (MLP with PCA)	0.164838	0.602722	75.23%
GFF-1-O-M (Generalized Feedforward)	0.186486	0.568681	75.84%
GFF-1-B-M (Generalized Feedforward)	0.176708	0.572106	74.92%
RBF-1-O-M (Radial Basis Function)	0.200182	0.475198	68.20%
RBF-1-B-M (Radial Basis Function)	0.247779	0.236243	52.91%
TDNN-1-O-M (Time-Delay Network)	0.249061	0.047939	54.74%
TDNN-1-B-M (Time-Delay Network)	0.208859	0.412738	66.36%
RN-1-O-M (Recurrent Network)	0.171959	0.564609	75.54%
RN-1-B-M (Recurrent Network)	0.170088	0.599442	76.45%
TLRN-1-O-M (Time-Lag Recurrent Network)	0.220068	0.347592	65.75%
TLRN-1-B-M (Time-Lag Recurrent Network)	0.207271	0.412928	67.89%

Slika 5.4 Arhitektura mreže za predviđanje R varijable



Proces učenja različitih arhitektura NN pokazao je višeslojni preceptron (MLP-1-B-L) s jednim skrivenim slojem (slika 5.4), back propagation algoritmom učenja uz uporabu Levenberg-Marquardt pravila učenja, kao najbolju arhitekturu za promatrani set podataka (tablica 5.7).

5.5.2.3 Topologija mreže za predviđanje M varijable

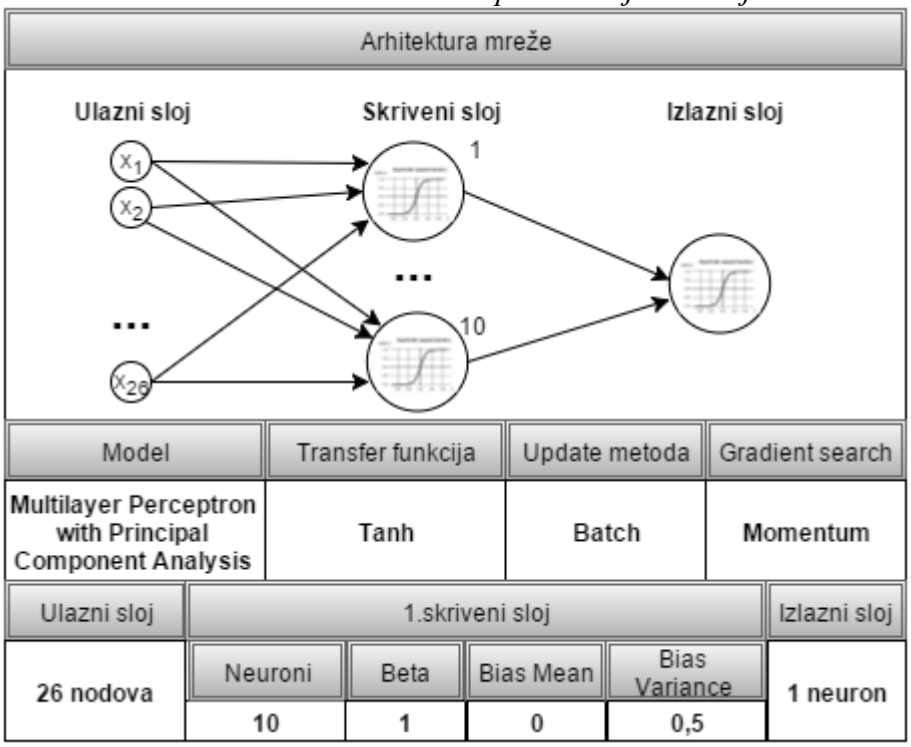
Za razliku od prethodne dvije varijable, M varijabla, poprima vrijednost iz skupa podataka između 0 i 1, pri čemu vrijednost predstavlja relativni udio kupca u ukupnom prometu narednog perioda. Problem je dakle regresijski u odnosu na prethodna dva klasifikacijska problema.

Tablica 5.8 Rezultat usporedbe različitih arhitektura mreže u predviđanju M varijable

Performance Metrics (variable M)

	Testing		
Model Name	**MSE**	**r**	**MAE**
MLP-1-O-M (Multilayer Perceptron)	0,001837	0,935655	0,007706
LR-0-B-M (Linear Regression)	0,000627	0,981068	0,007569
LR-0-B-L (Linear Regression)	0,00023	0,992171	0,003682
MLP-1-B-L (Multilayer Perceptron)	0,000437	0,985253	0,006474
PNN-0-N-N (Probabilistic Neural Network)	0,0005	0,982807	0,005098
RBF-1-B-L (Radial Basis Function)	0,000995	0,966235	0,009357
GFF-1-B-L (Generalized Feedforward)	0,002829	0,900337	0,008133
MLPPCA-1-B-L (MLP with PCA)	0,000223	0,992341	0,004838
SVM-0-N-N (Classification SVM)	72,4709	0,169421	7,004463
TDNN-1-B-L (Time-Delay Network)	0,002751	0,912656	0,013242
TLRN-1-B-L (Time-Lag Recurrent Network)	0,039172	0,237173	0,093127
RN-1-B-L (Recurrent Network)	0,15255	-0,01875	0,181154
MLP-2-B-L (Multilayer Perceptron)	0,000696	0,975715	0,006289
MLP-1-B-M (Multilayer Perceptron)	0,002551	0,939279	0,012762
MLP-2-O-M (Multilayer Perceptron)	0,000544	0,982803	0,006215
MLP-2-B-M (Multilayer Perceptron)	0,008567	0,941197	0,024338
MLPPCA-1-O-M (MLP with PCA)	0,000167	0,995535	0,005115
MLPPCA-1-B-M (MLP with PCA)	0,009829	0,598091	0,040351
GFF-1-O-M (Generalized Feedforward)	0,000211	0,992743	0,006142
GFF-1-B-M (Generalized Feedforward)	0,027855	0,595707	0,124947
RBF-1-O-M (Radial Basis Function)	0,002539	0,960925	0,014426
RBF-1-B-M (Radial Basis Function)	0,013185	0,395944	0,026864
TDNN-1-O-M (Time-Delay Network)	0,00123	0,977942	0,00814
TDNN-1-B-M (Time-Delay Network)	0,015412	-0,04499	0,032228
RN-1-O-M (Recurrent Network)	0,001449	0,949564	0,012235
RN-1-B-M (Recurrent Network)	0,009975	0,839229	0,084318
TLRN-1-O-M (Time-Lag Recurrent Network)	0,002191	0,94415	0,009963
TLRN-1-B-M (Time-Lag Recurrent Network)	0,020663	-0,01089	0,07804

Slika 5.5 Arhitektura mreže za predviđanje M varijable



Kroz testiranje različitih arhitektura mreže, kao najbolja se pokazala arhitektura višeslojne preceptron mreže uz principal komponent analizu (MLPPCA-1-O-M) s jednim skrivenim slojem i Momentum pravilom učenja (slika 5.5).

Skriveni sloj mreže ima 10 neurona. Kao transfer funkcija odabran je Tanh Axon uz vrijednosti Beta=1, Bias Mean vrijednost 0 i Bias varijancu 0,5.

5.5.3 Optimizacija mreže

Softver kojim su projektirane mreže generirao je inicijalno najbolje arhitekture mreže kako je prikazano u prethodnim razmatranjima. Parametri ovih arhitektura mogu se dodatno unaprijediti koristeći neke od poznatih tehnika optimizacije. Za potrebe ovog istraživanja korištene su sljedeće metode dodatnih testiranja kojima se pokušalo dodatno optimizirati projektirane arhitekture:

1. Ponavljanje učenja *n* puta (engl. *Train N Times*)
2. Varijacija parametra (engl. *Vary A Parameter*)
3. Izostavljanje *n* primjera (engl. *Leave N Out*)

4. Učenje uz upotrebu genetičkog algoritma (engl. *Train Genetic*).

Efikasnost pojedine metode provjerava se usporedbom rezultata prije pokretanja testa i nakon njenog provođenja. Rezultati primjene prve metode, *ponavljanje učenja n puta*, dani su u tablici 5.9. Učenje se provodi kroz pet ciklusa. Svaki ciklus podrazumijeva novi inicijalni skup težinskih faktora mreže definiran random metodom. Najbolji set težinskih faktora se pamti za svaki ciklus i epohu kada je najmanja greška na setu za unakrsnu validaciju, a na kraju testiranja među njima se odabire najbolji, a rezultati odabrane konfiguracije prikazani su u tablici.

Tablica 5.9 Usporedni rezultati učenja mreže ponavljanjem testiranja u više ciklusa

Ponavljanje učenja *n* puta		
Broj epoha	1000	
Broj ciklusa (n)	5	
Korištenje CV seta	Da	
Prekid CV testiranja nakon *x* epoha bez unapređenja mreže	Bez prekida	
Ravnomjerno ponderiranje klasa	Ne	
Varijabla	**R**	
Arhitektura	**MLP-1-B-M (Multilayer Perceptron)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,207645101	0,208811943
R	0,309139574	0,302454619
Correct	71,56%	71,25%
Varijabla	**F**	
Arhitektura	**MLP-1-B-L (Multilayer Perceptron)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,110981999	0,124743247
R	0,74769475	0,723637276
Correct	87,16%	83,18%
Varijabla	**M**	
Arhitektura	**MLPPCA-1-O-M (MLP with PCA)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,000167159	0,000255666
R	0,995534603	0,991660048
MAE	0,005115457	0,005966829

Upotrebom *metode učenja uz n ponavljanja* nisu se postigli bolji rezultati od inicijalnog testiranja, naprotiv oni su neznatno lošiji.

Varijacija parametara je proces kojim se nadograđuje prethodni model optimizacije mreže. Uz ponavljanje epoha u određenom broju ciklusa, mijenja se i određeni parametar mreže. U procesu istraživanja odlučeno je da će se mijenjati broj neurona u skrivenom sloju. Proces učenja započinje s dva neurona u skrivenom sloju mreže, a potom se povećava za 1. Ukupan broj varijacija je 15, a to znači da će se testirati konfiguracije koje imaju između 2 i 17 neurona u skrivenom sloju.

Tablica 5.10 Usporedni rezultati učenja mreže varijacijom parametara

Varijacija parametra		
Broj epoha	1000	
Broj ciklusa (n)	3	
Korištenje CV seta	Da	
Prekid CV testiranja nakon *x* epoha bez unapređenja mreže	Bez prekida	
Ravnomjerno ponderiranje klasa	Ne	
Parametar	Broj neurona skrivenog sloja	
Startna vrijednost	2	
Inkrement	1	
Broj varijacija	15	
Varijabla	**R**	
Arhitektura	**MLP-1-B-M (Multilayer Perceptron)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,207645101	0,206438963
R	0,309139574	0,315776607
Correct	71,56%	70,34%
Neuroni skrivenog sloja	10	13
Varijabla	**F**	
Arhitektura	**MLP-1-B-L (Multilayer Perceptron)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,110981999	0,1442569
R	0,74769475	0,672312154
Correct	87,16%	82,87%

Neuroni skrivenog sloja	10	7
Varijabla	**M**	
Arhitektura	**MLPPCA-1-O-M (MLP with PCA)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,000167159	0,00057894
R	0,995534603	0,981568918
MAE	0,005115457	0,007076672
Neuroni skrivenog sloja	10	15

Zamjetno je da nisu postignute bolje performanse mreže u odnosu na početno testiranje koje je dalo početne najbolje konfiguracije.

Algoritam *izostavljanja N primjera* prolazi kroz učenje mreže u više ciklusa, a u svakom ciklusu izostavlja zadani (N) broj zapisa. Izostavljeni zapisi u ciklusu čine set podataka za testiranje dobivene mreže. Na ovaj način se izbjegava stalno korištenje istog seta podataka za učenje, CV i testiranje, ali se postiže i efekat korištenja svih podataka u sve tri svrhe (učenje, unakrsna validacija i testiranje).

Tablica 5.11 Usporedni rezultati učenja mreže varijacijom parametara

Pomjeranje *n* primjera		
Broj epoha	1000	
Broj ciklusa (n)	5	
Korištenje CV seta	Da	
Broj CV zapisa	325	
Izostavljanje zapisa (n)	325	
Varijabla	**R**	
Arhitektura	**MLP-1-B-M (Multilayer Perceptron)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,207645101	0,204843072
R	0,309139574	0,30836145
Correct	71,56%	70,34%
Varijabla	**F**	
Arhitektura	**MLP-1-B-L (Multilayer Perceptron)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,110981999	0,15141847
R	0,74769475	0,65600177
Correct	87,16%	83,18%

Varijabla	M	
Arhitektura	**MLPPCA-1-O-M (MLP with PCA)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,000167159	0,000586451
R	0,995534603	0,97986559
MAE	0,005115457	0,005945229

Na kraju se dobivena konfiguracija testira na cjelokupnom setu podataka, a rezultati dobiveni ovom metodom prikazani su u tablici 5.11. Proces treniranja mreže ovom metodom, baš kao i prethodni nije postigao nikakva značajna poboljšanja početne konfiguracije koja se pokazala najboljom.

Zadnji algoritam optimiziranja arhitekture odabrane neuronske mreže zasniva se na još jednoj metodi strojnog učenja, *genetičkom algoritmu*. Proces učenja koristi ovaj prirodni proces za optimiziranje: i. broja ulaznih varijabli, ii. veličine koraka, iii. vrijednosti momentuma i iv. broja neurona u skrivenim slojevima mreže. Optimiziranje pretpostavlja pronalaženje odgovarajuće arhitekture i skupa težinskih faktora koji će proizvesti minimalnu grešku mreže.

Opisani proces započinje kreiranjem inicijalnih populacija. Broj populacija je ulazni parametar procesa. Svaka populacija predstavlja slučajno generiranu ahitekturu mreže s prethodno pobrojanim parametrima. Svaka od populacija prolazi potom fazu učenja kroz odabrani broj epoha, još jedan od ulaznih parametara procesa optimizacije. Na kraju procesa učenja svaka populacija ima pokazatelje svoje uspješnosti. Karakteristike „dobrih mreža“ inicijalnog kruga podliježu potom operacijama selekcije, križanja i mutiranja kako bi se kreirale nove populacije. Proces se ponavlja kroz zadani broj generacija, a prekida se kada mreža dostigne jedan od zadanih parametara: maksimalni broj generacija ili maksimalno vrijeme evaluacije. Ovaj algoritam je jako zahtjevan sa stanovišta procesorske snage i vremena, pa je stoga i vrijeme evaluacije jedan od zadanih parametara. Inicijalni parametri zajedno s rezultatima procesa optimizacije prezentirani su u tablici 5.12.

Rezultati primjene genetičkog algoritma ukazuju na njegovu efikasnost u procesu optimiziranja neuronske mrežne arhitekture modeliranog sustava. Klasifikacijsko predviđanje R i F varijabli poboljšano je smanjenjem mjera srednje vrijednosti greške, povećanjem

koeficijenta korelacije i postotka ispravnih predviđanja. Regresijski problem predviđanja M varijable postigao je jednako dobre rezultate novo projektiranom arhitekturom.

Tablica 5.12 Usporedni rezultati učenja mreže upotrebom genetičkog algoritma

Učenje upotrebom genetičkog algoritma		
Broj epoha	1000	
Broj populacija	50	
Broj generacija	100	
Max. vrijeme evaluacije	360 minuta	
Optimizacija	Od	do
Step size	0	1
Momentum	0	1
Broj neurona u skr.slojevima	1	20
Varijabla	**R**	
Arhitektura	**MLP-1-B-M (Multilayer Perceptron)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,207645101	0,196827646
R	0,309139574	0,329725049
Correct	71,56%	75,34%
Broj nezavisnih varijabli	27	17
Broj neurona skrivenog sloja	10	3
Varijabla	**F**	
Arhitektura	**MLP-1-B-L (Multilayer Perceptron)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,110981999	0,109852544
r	0,74769475	0,749959088
Correct	87,16%	89,62%
Broj nezavisnih varijabli	27	12
Broj neurona skrivenog sloja	10	2
Varijabla	**M**	
Arhitektura	**MLPPCA-1-O-M (MLP with PCA)**	
Pokazatelj	*Prije testa*	*Nakon testa*
MSE	0,000167159	0,000588195
r	0,995534603	0,979626097
MAE	0,005115457	0,006811475
Broj nezavisnih varijabli	27	16
Broj neurona skrivenog sloja	10	6

5.6 Metode usporedbe različitih metoda za predviđanje profitabilnosti kupaca

Rezultati dobiveni primjenom različitih metoda neuronske mreže nad testnim setom podatka koriste se za konačnu procjenu pojedinih modela i odabir najboljeg. Klasifikacijski problemi (predviđanje R i F varijabli) u središte promatranja postavljaju klasifikacijske matrice i postotak korektnih rezultata. Regresijski problem prisutan kod predviđanja M varijable u fokus postavlja konfiguraciju s najmanjom srednjom vrijednošću kvadrata greške (MSE). Procjena se vrši usporedbom pokazatelja performansi generiranih NN i logističkih modela. Odabrani pokazatelji za potrebe ovog istraživanja su:

f. MSE (engl. *Mean Squared Error*),
g. *NMSE* (engl. *Normalised Mean Square Error*),
h. Korelacijski koeficijent ***r*** (engl. *Correlation Coeficient*),
i. Postotak pogreške ***%Error***,
j. *AIC* (engl. *Akaike's information criterion*).

MSE je pokazatelj mjere srednje vrijednosti kvadrata greške (Principe et al., 2010). Greška je iznos razlike stvarne i predviđene vrijednosti mreže. Poželjno je da MSE vrijednost bude što manja. Ako je *P* broj izlaznih varijabli, *N* broj instanci podataka u skupu podataka, y_{ij} izlazni rezultat mreže za instancu podataka i na izlaznom neuoronu j te d_{ij} željena vrijednost izlaza za istu instancu i izlaznu varijablu, tada se MSE računa sljedećom jednadžbom:

$$MSE = \frac{\sum_{j=0}^{p} \sum_{i=0}^{N} (d_{ij} - y_{ij})^2}{NP} \qquad (5.22)$$

NMSE je procjenitelj sveukupne devijacije između predviđene i stvarne vrijednosti izlazne varijable. Kao i kod prethodnog pokazatelja vrijedi pravilo da je poželjna što niža vrijednost pokazatelja NMSE. Uz poznate elemente jednadžbe 5.2 ovaj pokazatelj se računa sljedećom jednadžbom:

$$NMSE = \frac{P*N*MSE}{\sum_{j=0}^{p}\left(\frac{\left(N\sum_{i=0}^{N}\left(d_{ij}^2\right)-\left(\sum_{i=0}^{N}d_{ij}\right)^2\right)}{N}\right)} \qquad (5.23)$$

Korelacijski koeficijent *r* nadoknađuje nedostatak MSE koeficijenta da prikaže u kojem pravcu se kreće predviđena vrijednost u odnosu na stvarnu vrijednost zavisne varijable (Principe et al., 2010). Ako se s *x* označi dobivena vrijednost zavisne varijable, a s *d* željena vrijednost, jednadžba korelacijskog koeficijenta definirana je na sljedeći način:

$$r = \frac{\frac{\sum_i(x_i-\overline{x})(d_i-\overline{d})}{N}}{\sqrt{\frac{\sum_i(d_i-\overline{d})^2}{N}}\sqrt{\frac{\sum_i(x_i-\overline{x})^2}{N}}} \qquad (5.24)$$

Koeficijent *r* prima vrijednost iz ranga [-1,1]. Granične vrijednosti definiraju savršenu pozitivnu linearnu korelaciju (*r=1*) ili savršenu negativnu korelaciju (*r=-1)* dobivene i željene vrijednosti izlaza mreže. Korelacija ne postoji u slučaju kada je *r=0*. Kako bi promatrani model mreže bio što bolje prilagođen podacima, neophodno je da vrijednost ovog koeficijenta bude što bliže vrijednosti 1.

Postotak pogreške ***%error*** definira srednju vrijednost postotka greške željenih (dd_{ij}) i dobivenih vrijednosti (dy_{ij}) mreže za sve slučajeve (N) promatranog seta podataka i broj izlaznih varijabli (P) (jednažba 5.25).

$$\%error = \frac{100}{NP}\sum_{j=0}^{p}\sum_{i=0}^{N}\frac{|dy_{ij}-dd_{ij}|}{dd_{ij}} \qquad (5.25)$$

AIC je pokazatelj koji daje informaciju o tome u kolikoj mjeri je generirani model aproksimacija realnog svijeta, ili još točnije u kojoj mjeri projektirani model mreže minimizira gubitak informacija o realnom (projektiranom) problemu (Principe et al., 2010). On je u osnovi poveznica Kullback-Leibler pokazatelja koji predstavlja gubitak informacije u projektiranom modelu i MLE (engl. *Maximum-likelihood estimation*) pokazatelja procjene parametara statističkog modela.

$$AIC = -2(log - likelihood) + 2K \qquad (5.26)$$

U jednadžbi (5.26) K predstavlja broj parametara modela. Kako je MLE vrijednost pokazatelj generiranog statističkog modela, AIC prikazuje sveukupnu prilagođenost modela, pri čemu veća vrijednost predstavlja bolji pokazatelj.

6. REZULTATI EMPIRIJSKOG ISTRAŽIVANJA

6.1 Ciljevi empirijskog istraživanja

Svrha istraživanja ogleda se u davanju doprinosa pri pronalasku odgovora na znanstveno pitanje kako mjeriti i kojim metodama predviđati profitabilnost pojedinog kupca korištenjem podataka iz postojećih upravljačkih informacijskih sustava. Glavno pitanje istraživanja nameće i niz parcijalnih pitanja koja traže odgovore:

- Koji je opći okvir za definiranje profitabilnosti kupca?
- Kako nadoknaditi nedostatak povijesnih financijskih i transakcijskih podataka za procjenu ukupne vrijednosti kupca?
- Jesu li metode strojnog učenja i multivarijacijske statističke analize primjenjive u predviđanju profitabilnosti kupca?
- Koje metode koristiti u predviđanju profitabilnosti kupca?
- Koje su metode bolje sa stanovišta primjene i točnosti predviđanih mjera profitabilnosti?
- Koje mjere koristiti u izražavanju profitabilnosti kupaca?
- Da li odabrane mjere profitabilnosti stvaraju dobru osnovu za donošenje odluka menadžmenta u svrhu povećanja profitabilnosti kupaca?
- Kako poslovna politika prema pojedinom kupcu utječe na njegovu profitabilnost?
- Kako karakteristike pojedinog kupca utječu na njegovu profitabilnost?

Znanstveni cilj istraživanja definiran je na sljedeći način:

- Razvoj općeg teorijskog modela, utemeljenog na postojećoj znanstvenoj literaturi, za mjerenje i predviđanje profitabilnosti kupaca koristeći podatke iz upravljačkih informacijskih sustava.
- Otkrivanje smjera i intenziteta veze između karakteristika kupca i profitabilnosti kupca.

- Otkrivanje smjera i intenziteta veze između poslovne politike koju poduzeće provodi prema kupcu i profitabilnosti kupca.

Pragmatični cilj istraživanja odnosi se na stvaranje odgovarajućeg podatkovno-aplikacijskog modela za potrebe implementacije u postojeće upravljačke informacijske sustave proizvodno-distributerskih poduzeća sa svrhom predviđanja profitabilnosti postojeće osnove kupaca.

6.2 Polazne hipoteze istraživanja

Istraživanjem se nastoji stvoriti model predviđanja profitabilnosti kupaca, utvrditi prednost specifične metode strojnog učenja, neuronske mreže, u odnosu na metode multivarijacijske statističke analize, te dokazati utjecaj provođenja marketinških politika i karakteristika kupca na mjere profitabilnosti. Sukladno navedenom definiran je model (slika 5.1), tri glavne hipoteze istraživanja i pet pomoćnih koje će se nastojati potvrditi istraživanjem.

H1: Metoda neuronske mreže može predvidjeti profitabilnost kupaca jednako dobro ili bolje od statističkih metoda logističke regresije i višestruke linearne regresije.

Neuronska mreža će se koristiti za predviđanje sve tri izlazne varijable modela, dok će se logit metoda koristiti za predviđanje **R** i **F** varijable, a MLR metoda za predviđanje **M** varijabli. Ova hipoteza stavlja u odnos odabrane metode strojnog učenja i multivarijacijske analize i nastoji dokazati dominaciju metode strojnog učenja u području predviđanja profitabilnosti kupaca.

H2: Politika prema kupcima utječe na profitabilnost kupaca

Istraživačka pretpostavka ove hipoteze u odnos stavlja politiku koju poduzeće provodi prema pojedinom kupcu i tvrdi da ona direktno utječe na veličinu mjera profitabilnosti kupca. Kako prema definiranom modelu istraživanja, politiku prema kupcu čine cjenovna politika i

veličina troškova nastalih poslovnim odnosom s kupcem, formirane su dvije pomoćne hipoteze koje će potpomognuti dokazivanje glavne hipoteze.

H2.1: Troškovna komponenta politike prema kupcu utječe na profitabilnost kupca

H2.2: Cjenovna politika prema kupcu utječe na profitabilnost kupca

Istraživačka pretpostavka treće hipoteze stavlja u odnos heterogenost skupa kupaca te tvrdi da ona direktno utječe na veličinu mjera profitabilnosti kupca:

H3: Karakteristike kupca utječu na profitabilnost kupca

Definiranim modelom istraživanja karakteristike kupca su grupirane u kategorije veličine, kompleksnosti i ponašanja pri plaćanju obveza. Veličinu kupca čine financijski elementi poslovnog odnosa iskazani veličinom prometa i ostvarene marže u poslovnom odnosu s kupcem. Osim ovih čisto financijskih elemenata, ovu kategoriju karakterizira i pripadnost grupi kupaca sukladno veličini (malo, srednje ili veliko poduzeće) i teritorijalnoj pripadnosti (domaće ili inozemno poduzeće). Kompleksnost kupca definira stupanj složenosti poslovnog odnosa s kupcem. Modelirana je kroz varijable koje kvantificiraju količinu narudžbi od kupca, različitost asortimana isporučenih proizvoda, broj mjesta isporuke te količinu i vrijednost transakcija povrata robe. Ponašanje pri plaćanju definira način ponašanja kupca u finalnom koraku poslovnog odnosa na relaciji promatranog poduzeća i pojedinačnog kupca. Varijable ove kategorije određuju iznose potraživanja naplaćene na vrijeme, vremensku strukturu naplate potraživanja van ugovorenog vremena plaćanja, nenaplaćena potraživanja i prosječno vrijeme kašnjenja u plaćanju. Sukladno navedenom, formirane su tri pomoćne hipoteze koje će potpomognuti dokazivanje navedene glavne hipoteze.

H3.1: Veličina kupca utječe na profitabilnost kupca

H3.2: Kompleksnost kupca utječe na profitabilnost kupca

H3.3: Ponašanje pri plaćanju kupca utječe na profitabilnost kupca

6.3 Metodološki aspekti

6.3.1 Definicija podataka

Nezavisne varijable projektiranog modela za predviđanje profitabilnosti grupirane su prema srodnosti u:

- podatke o troškovima,
- podatke o značajkama kupaca,
- podatke o cjenovnoj politici prema kupcu.

Grupa podataka o troškovima sadrži varijable kojima se na razini pojedinačnog kupca iskazuju troškovi nastali u različitim fazama poslovnog odnosa između pojedinačnog kupca i poduzeća. Do njihovih vrijednost dolazi se izračunom tekuće profitabilnosti uporabom modificiranog Niraj modela (Niraj et al. 2001). Izvor podataka za izračun ovih pokazatelja su transakcijski podaci poduzeća. Vrijednosti varijabli iskazane su u novčanim jedinicama na razini pojedinačne godine iz promatranog perioda 2008-2013. Pojedinačne varijable ove grupe podataka s osnovnim deskriptivnim statističkim pokazateljima prikazane su u tablici 6.1.

Tablica 6.1 Varijable troškova

Oznaka	Opis	Min	Max	Mean	Median	stdev
T_1	Troškovi prodaje i direktnog marketinga	60,40	288.060,59	1.547,87	193,05	11.444,20
T_2	Troškovi procesiranja narudžbi kupaca	0,00	6.215,19	27,90	2,66	229,04
T_3	Troškovi isporuke kupcu	0,00	126.918,39	736,34	43,19	5.228,67
T_4	Troškovi nabave od dobavljača i skladištenja	0,00	258.081,09	822,61	15,08	9.466,43
T_5	Troškovi pripreme narudžbi od dobavljača	0,00	57.467,61	230,02	7,46	2.201,67
T_6	Troškovi manipulacije i prijevoza robe od dobavljača	0,00	8.664,05	29,58	0,63	317,26
T_7	Troškovi skladišnog prostora i troškovi kapitala	0,00	505.120,44	2.442,69	75,03	20.003,01

Deskriptivni podaci odnose se na vrijednost nezavisnih varijabli prije procesa normalizacije.

Profitabilnost je kompleksna karakteristika kupca na koju utječu varijable iz skupa podataka o *značajkama kupca*. Prethodno navedeni faktori profitabilnosti, troškovi, upravo su ovisni o značajkama kupca. U tablici 6.2 dan je pregled svih varijabli ove grupe ulaznih podataka uz osnovne statističke pokazatelje.

Tablica 6.2 Varijable značajki kupca

Oznaka	Opis	Min	Max	Mean	Median	stdev
VK1	Veličina prometa s kupcem	0.00	10.227.703,41	71.423,24	2.788,32	579.164,83
VK2	Bruto marža ostvarena u prometovanju s kupcem	-30.190,14	3.566.012,63	23.271,78	1.000,84	198.019,09
VK3	Veličina kupca					
VK4	Domicilna/inozemna kompanija					
KK1	Broj različitih kupljenih proizvoda	1,00	366,00	26,44	17,00	30,95
KK2	Broj narudžbi koje kupac napravi u promatranom vremenskom periodu	1,00	8.278,00	55,78	7,00	385,93
KK3	Broj lokacija isporuke					
KK4	Vrijednost povrata robe	0,00	122.248,65	834,91	0,00	6.235,21
KK5	Broj transakcija povrata robe u promatranom vremenskom periodu	0,00	6.610,00	25,15	0,00	250,08
PP1	Vrijeme kašnjenja u plaćanju kupca					
PP2	Iznos potraživanja naplaćen na vrijeme	0,00	6.215,19	27,90	2,66	229,04
PP3	Iznos potraživanja naplaćen s zakašnjenjem do 30 dana	0,00	126.918,39	736,34	43,19	5.228,67
PP4	Iznos potraživanja naplaćen s zakašnjenjem do 90 dana	0,00	258.081,09	822,61	15,08	9.466,43
PP5	Iznos potraživanja naplaćen s zakašnjenjem do 180 dana	0,00	57.467,61	230,02	7,46	2.201,67
PP6	Iznos potraživanja naplaćen s zakašnjenjem preko 180 dana	0,00	8.664,05	29,58	0,63	317,26
PP7	Dospjeli dug kupca na kraju promatranog razdoblja	0,00	505.120,44	2.442,69	75,03	20.003,01

Četiri varijable promatrane grupe nisu numeričke prirode: VK_3, VK_4, KK_3 i PP_1. Tablica 6.3 daje pregled mogućih vrijednosti i distribuciju unutar skupa ulaznih podataka.

Tablica 6.3 Deskriptivna statistika varijabli VK_3, VK_4, KK_3 i PP_1

VK3 - Veličina kupca		
1 - malo poduzeće	256	36,11%
2 - srednje poduzeće	415	58,53%
3 - veliko poduzeće	38	5,36%
	709	**100,00%**
VK4 - Domicilno/inozemno poduzeće		
1 - domicilno poduzeće	672	97,67%
2 - inozemno poduzeće EU	13	1,89%
3 - inozemno poduzeće nonEU	3	0,44%
	688	**100,00%**
KK3 - Broj lokacija isporuke		
1 - jedna	628	85,79%
2 - od 2 do 10	83	11,34%
3 - preko 10	21	2,87%
	732	**100,00%**
PP1 - Prosječno vrijeme kašnjenja u plaćanju kupca		
1 - veće od median vrijednosti	395	43,60%
0 - manje od median vrijednosti	511	56,40%
	906	**100,00%**

Grupa podataka o cijeni proizvoda, te s njom povezanoj bruto marži, čini sljedeći važan faktor profitabilnosti kupaca. Umnožak cijene proizvoda i količine prodanih proizvoda daje prihod. U izračunu profitabilnosti kupaca, od ostvarenog prihoda po kupcu oduzimaju se troškovi i dobiva se ostvareni bruto profit. Pri tome, cijena nije fiksna kategorija već se razlikuje od kupca do kupca. Stoga će se u modelu profitabilnosti koristiti ulazne varijable koje definiraju odstupanje pojedinačnih cijena po kupcima od standardnih cijena proizvoda. Promatrano poduzeće u internim pravilima poslovanja razlikuje tri kategorije rabata na osnovu kojih su i definirane varijable ove grupe ulaznih podataka (tablica 6.4).

Tablica 6.4 Varijable cijenovne politike

Oznaka	Opis	Min	Max	Mean	Median	stdev
C1	Individualni rabat	0,00	199.959,15	1.648,08	43,28	9.259,96
C2	Teritorijalni rabat	0,00	204.227,07	1.486,45	49,62	10.360,98
C3	Količinski rabat	0,00	325.676,92	587,11	0,00	9.714,43

Prva varijabla je individualni popust koji svaki kupac pojedinačno ostvari na osnovu ugovorenih stopa rabata u dogovaranju poslovnog odnosa. Teritorijalna pripadnost kupca određuje veličinu druge varijable ove grupe. Ovaj tip rabata nastao je na poslovnoj politici poduzeća da stimulira kupovinu u određenim regijama svog tržišta davanjem dodatnih rabata. Količinske rabate kupac ostvaruje kupovinom iznad dogovorene količine/vrijednosti određenih proizvoda na godišnjoj razini.

Zavisne varijable su proizvod modela za predviđanje. Njihova definicija u istraživanju zasnovana je na RFM modelu (Hughes, 1994), a imenovanje je napravljeno prema definiciji modela:

- R varijabla,
- F varijabla,
- M varijabla.

R varijabla se odnosi na vremenski interval između kupovina pojedinačnog kupca. Za pojedinog kupca promatra se prosječna vrijednost ovih vremenskih intervala i medijan vrijednost na razini godine. Ukoliko je prosječna vrijednost za pojedinog kupca manja od medijan vrijednosti varijabla R prima vrijednost 0, a u sprotnom vrijednost 1. F varijabla se odnosi na broj kupovina i po sličnom pravilu poprima vrijednost 0 ili 1 kao što je slučaj s R varijablom. Dakle, ukoliko je ukupni broj kupovina kupca na razini godine manji od medijan vrijednosti tada F varijabla za kupca poprima vrijednost 0, a u suprotnom 1. M varijabla predstavlja monetarnu vrijednost kupovine, a za potrebe istraživanja svedena je na vrijednost od -1 do 1 upotrebom jednadžbe 5.14.

Osnovu za predviđanje profitabilnosti kupca čine podaci iz vremenskog perioda **t** koji čini jedna kalendarska godina. Predviđanje se vrši za budući vremenski period **t+1** (slijedeća godina u odnosu na podatke koji čine osnovu za predviđanje) u kome kupac možda više neće biti kupac, odnosno u prometovanju promatranog poduzeća neće učestvovati u

kupoprodajnom odnosu. U tijesnoj vezi s (ne)aktivnosti kupca su stopa zadržavanja kupaca i njoj suprotna stopa odlazaka kupaca. One stavljaju u odnos broj zadržanih, odnosno izgubljenih kupaca s ukupnom bazom kupaca određenog perioda.

Ukoliko kupac nastavi ovu ulogu u narednom periodu (R varijabla), zanimljivo je predvidjeti broj transakcija koje će nastati u tom odnosu (F varijabla). Ovaj pokazatelj sadrži sintezu stope zadržavanja, ponavljanja kupnje i rasta (ili pada) računa pojedinačnog kupca.

Jedan od financijskih pokazatelja budućeg prometa je i očekivani promet (M varijabla). Varijabla je to koja u sebi sadrži više komponenti. Svakako je bitan miks proizvoda kojima se prometuje. Količina je bitan činitelj, ali i cjenovna politika u odnosu s pojedinačnim kupcem.

6.3.2 Instrumenti istraživanja

Za teorijski dio disertacije korišteni su uobičajeni metodološki postupci (analiza, sinteza, kompilacija, deskripcija, komparacija, klasifikacija, generalizacija, modeliranje). Uporabom nabrojanih metoda nastojalo se dati teorijsku osnovu problema profitabilnosti i predviđanja mjera profitabilnosti.

Prethodno navedene hipoteze istraživanja, postavljaju u međusobni odnos i uspoređuju efikasnost metoda multivarijantne statističke analize i metoda strojnog učenja nad postavljenim problemom profitabilnosti kupaca.

U skladu s postavljenim pretpostavkama, za empirijski dio istraživanja koristile su se metode strojnog učenja i multivarijantne statističke analize (tablica 6.5). Metoda neuronske mreže koristila se za predviđanje sve tri nezavisne varijable modela. Genetički algoritam se dodatno koristio kao metoda za optimiziranje arhitektura neuronske mreže koje su se pokazale kao najbolje u procesu učenja i validacije. Metode strojnog učenja koristile su se samo za dokazivanje prve hipoteze istraživanja. Za ovu hipotezu, ali i za preostale dvije, koristile su se metode višestruke logističke regresije. Logit metoda je korištena u dokazivanju prve hipoteze i to za predviđanje F i R varijabli, a što je i uvjetovano prirodom nezavisnih varijabli. U istoj hipotezi je metoda višestruke logističke regresije korištena za predviđanje M varijable. Za obje grupe metoda, kod dokazivanja prve hipoteze, korišten je kompletan skup

nezavisnih varijabli modela. Pretpostavke druge i treće hipoteze uvjetovale su primjenu višestruke regresije kako bi se dokazala korelacija i utjecaj pojedinačnih nezavisnih varijabli modela s profitabilnošću kupaca.

Tablica 6.5 Primjena metoda po hipotezama istraživanja

Hipoteza	Multivarijacijska statistička analiza	Metode strojnog učenja	Nezavisne varijable	Zavisna varijabla
H_1	Logit		T, C, VK, KK, PP	R
	Logit		T, C, VK, KK, PP	F
	Višestruka linearna regresija		T, C, VK, KK, PP	M
		Neuronska mreža i GA	T, C, VK, KK, PP	R
		Neuronska mreža i GA	T, C, VK, KK, PP	F
		Neuronska mreža i GA	T, C, VK, KK, PP	M
H_{21}	Višestruka linearna regresija		T	M
H_{22}	Višestruka linearna regresija		C	M
H_{31}	Višestruka linearna regresija		VK	M
H_{32}	Višestruka linearna regresija		KK	M
H_{33}	Višestruka linearna regresija		PP	M

Izvor podatka za istraživanje pronađen je u poduzeću iz lanca snabdijevanja u industriji hrane. Poduzeće igra dvostruku ulogu: distributera i proizvođača proizvoda u industriji proizvoda sušenog voća, povrća, orašastih proizvoda, sjemenki, mahunarki i žitarica. Poduzeće svoju proizvodnu djelatnost obavlja na području Bosne i Hercegovine, a kupci su s područja zemalja Europe. Od poduzeća je dobivena dozvola za upotrebu podataka za potrebe istraživanja. Upotrebljavati će se podaci iz vremenskog perioda 2008.-2013. godine. Za pripremu podataka koristilo se prvenstveno skladište podataka poduzeća, ali su za specifične svrhe izračuna indirektnih troškova korištene transakcijske baze podataka.

Sve korištene metode empirijskog istraživanja su matematičke metode te su sve ulazne i izlazne nenumeričke varijable modela pretvorene u numeričke. Istovremeno, sve numeričke varijable modela su normalizirane, odnosno svedene na raspon od -1 do 1 (jednadžba 5.14).

6.3.3 Preliminarni pregled podataka

Promatrano poduzeće je u odabranom periodu prometovalo s 761 različitih kupaca. Struktura kupaca po godinama te atributima veličine i zemlje porijekla prikazana je u tablici 6.6.

Tablica 6.6 Broj kupaca po godinama

Godina	Broj kupaca	Veličina poduzeća			Zemlja porijekla		
		Malo	Srednje	Veliko	BiH	EU	nonEU
2008	346	109	218	19	346	0	0
2009	297	84	196	17	297	0	0
2010	291	81	196	14	291	0	0
2011	336	132	183	21	325	8	3
2012	326	120	190	16	320	3	3
2013	329	120	191	18	320	6	3

U nastavku je dana deskriptivna statistika ostvarenog prometa s kupcima. Podaci iz prometa čine osnovu podatkovnog modela sustava upotrebljenog za predviđanje profitabilnosti kupaca.

Deskriptivna statistika za proizvode sadrži podatke o broju artikala po pojedinim godinama, uz srednju, median, minimalnu i maksimalnu vrijednost za sljedeće pokazatelje ovog prometa:

a. Prodaja pojedinačnog artikla u kg (tablica 6.7)
b. Broj narudžbi (tablica 6.8)
c. Prosječne isporuke u kg po narudžbi (tablica 6.9)

Tablica 6.7 Količina prodaje u kg

Prodaja pojedinačnog artikla (kg)					
Godina	Broj artikala	Mean	Median	Min	Max
2008	278	14.271,60	1.103,00	0,20	313.085,50
2009	342	9.629,30	875,00	0,20	303.313,00

2010	385	8.569,90	848,30	0,10	352.373,40
2011	487	8.846,20	1.107,60	0,40	292.722,80
2012	459	9.016,70	980,00	1,30	358.961,70
2013	442	8.168,90	1.011,00	0,50	268.568,50

Tablica 6.8 Broj narudžbi po godinama

Broj narudžbi					
Godina	Broj artikala	Mean	Median	Min	Max
2008	278	372,20	85,00	1	3190
2009	342	340,00	69,50	1	3773
2010	385	410,90	105,00	1	5021
2011	487	355,20	56,00	1	5427
2012	459	390,70	55,00	1	5228
2013	442	436,20	68,00	1	5979

Tablica 6.9 Prosječna količina po isporuci u kg

Prosječne količine po isporuci (kg)					
Godina	Broj artikala	Mean	Median	Min	Max
2008	278	341,60	55,00	0,20	288.000,00
2009	342	253,30	50,10	0,10	48.000,00
2010	385	201,10	41,80	0,10	24.000,00
2011	487	209,90	45,40	0,20	64.000,00
2012	459	186,00	39,40	0,40	19.313,60
2013	442	148,60	31,30	0,20	21.964,40

Deskriptivna statistika za kupce daje osnovne statističke podatke o broju i strukturi transakcija vezanih za promet po kupcima i promatranim godinama. Provedena je analiza vezana za ukupnu količinu svih naručenih artikala (tablica 6.10), broj narudžbi (tablica 6.11), broj različitih lokacija isporuke (tablica 6.12), broj različitih artikala koje je pojedini kupac naručivao (tablica 6.13) i ukupnu vrijednost narudžbi izraženu u domicilnoj valuti promatranog poduzeća (tablica 6.14).

Tablica 6.10 Prosječna ukupna naručena količina robe po kupcu u KG

Isporučena prosječna količina po kupcu (kg)					
Godina	Broj kupaca	Mean	Median	Min	Max
2008	346	72.672,20	3.816,00	7,80	8.326.458,80
2009	297	76.637,20	3.424,40	10,00	7.643.085,20
2010	291	157.444,80	4.156,40	15,00	13.086.211,80
2011	336	13.847,50	595,30	5,00	1.479.780,60
2012	326	14.110,30	603,90	1,40	1.536.631,10
2013	329	12.658,90	556,40	4,40	1.356.224,80

Tablica 6.11 Prosječni broj narudžbi po kupcu

Broj narudžbi po kupcu					
Godina	Broj kupaca	Mean	Median	Min	Max
2008	346	31,80	6,00	1	2489
2009	297	42,40	7,00	1	3751
2010	291	55,50	8,00	1	531
2011	336	66,00	9,00	1	8261
2012	326	75,90	10,00	1	8192
2013	329	85,20	10,50	1	7720

Tablica 6.12 Broj različitih lokacija isporuke po kupcu

Broj lokacija isporuke					
Godina	Broj kupaca	Mean	Median	Min	Max
2008	346	2,20	1,00	1	79
2009	297	2,40	1,00	1	120
2010	291	2,70	1,00	1	140
2011	336	2,90	1,00	1	153
2012	326	3,00	1,00	1	158
2013	329	3,30	1,00	1	168

Tablica 6.13 Prosječan broj različitih artikala naručenih od pojedinog kupca

Broj različitih artikala za kupca					
Godina	Broj kupaca	Mean	Median	Min	Max
2008	346	2,20	1,00	1	79
2009	297	2,40	1,00	1	120
2010	291	2,70	1,00	1	140
2011	336	2,90	1,00	1	153
2012	326	3,00	1,00	1	158
2013	329	3,30	1,00	1	168

Tablica 6.14 Prosječan promet ostvaren po narudžbi za pojedinog kupca

Ostvareni promet po narudžbi					
Godina	Broj kupaca	Mean	Median	Min	Max
2008	346	1.858,40	444,10	9,70	122.459,90
2009	297	1.591,40	400,10	5,10	180.946,40
2010	291	1.406,10	380,20	5,10	163.743,50
2011	336	1.223,70	285,80	1,00	181.341,30
2012	326	1.150,90	270,20	3,90	193.042,80
2013	329	1.040,60	251,00	1,20	223.950,40

Deskriptivna statistika profitabilnosti kupaca prikazuje profil ukupne profitabilnosti promatranog poduzeća po godinama. Tablica 6.15 daje relativne udjele troškova i profita u ostvarenom neto prihodu. Primjetan je porast grupe indirektnih troškova i sukladno ovome smanjenje neto prihoda. Fiksni troškovi robe (gotovih proizvoda koji se preprodaju ili sirovina za proizvodnju) su relativno stabilni kroz vremenske periode što promatrano poduzeće može zahvaliti kupovini robe i sirovina za preradu na svjetskim burzama ovih proizvoda i ugovaranju po najpovoljnijim cijenama za duže vremenske periode.

Tablica 6.15 Struktura prometa po godinama

Profil profitabilnosti kupaca po godinama	**Godina**					
	2008	**2009**	**2010**	**2011**	**2012**	**2013**
1. Broj kupaca	346	297	291	336	326	329
2. Neto prihod	100,00	100,00	100,00	100,00	100,00	100,00

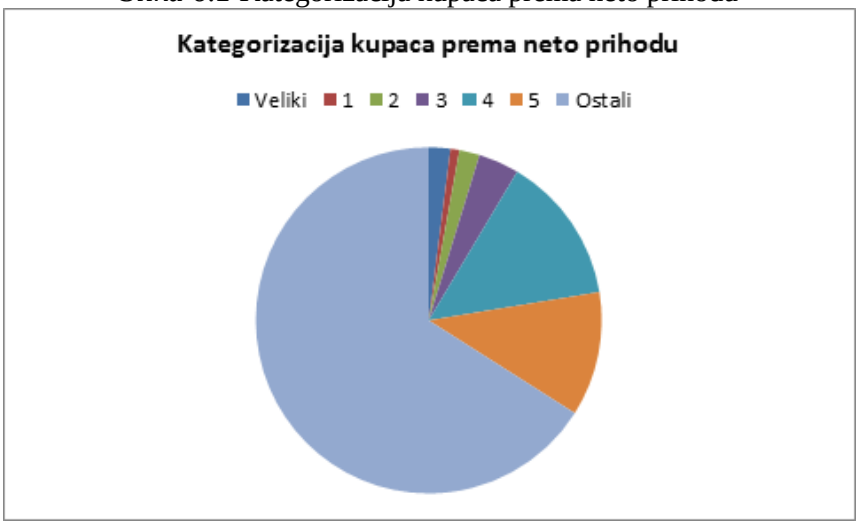
3. Trošak robe	67,96	67,98	69,58	67,68	66,96	65,19
4. Bruto profit	32,04	32,02	30,42	32,32	33,04	34,81
5. Indirektni troškovi narudžbi robe i sirovina od dobavljača i skladištenja (T_4+T_5+T_6+T_7)	6,25	7,08	7,21	5,93	5,48	5,00
6. Troškovi isporuke robe kupcima (T_2+T_3)	2,47	2,44	2,78	3,98	4,65	4,04
7. Troškovi prodaje i direktnog marketinga (T1)	1,97	1,39	1,53	2,75	4,65	4,00
8. Ukupni indirektni troškovi (5+6+7)	10,69	10,91	11,52	12,66	14,78	13,04
9. Neto profit (4-8)	21,35	21,11	18,90	19,66	18,26	21,77

Veličina kupca, određena neto prihodom, čini bitnu odrednicu profitabilnosti kupca. Sukladno veličini neto prihoda, kupci su klasificirani u 7 grupa (tablica 6.16).

Tablica 6.16 Kategorizacija kupaca prema veličini prihoda

Grupa	Neto prihod		Broj kupaca	%
	Od	Do		
Veliki kupac	1.000.000,00		19	2,01
1	500.000,00	1.000.000,00	8	0,84
2	250.000,00	500.000,00	18	1,90
3	100.000,00	250.000,00	36	3,80
4	50.000,00	100.000,00	131	13,83
5	25.000,00	50.000,00	110	11,62
Ostali	0,00	25.000,00	625	66,00

Slika 6.1 Kategorizacija kupaca prema neto prihodu



Podaci iz tablica 6.16 i 6.17 ukazuju na podatak da 2% kupaca ostvaruje 79,68% prometa. Podaci promatranog poduzeća čine ekstremnu varijantu pravila „80:20“. Podaci o neto profitu ukazuju da 46,56% kupaca (441 od ukupno 947) ostvaruje negativan profit.

Tablica 6.17 Portfolio kupaca po veličini prihoda

Grupa	Udio u prihodu (%)	Srednja vrijednost			Rang		
		Margina bruto prihoda (%)	Ind.troškovi kao postotak prodaje (%)	Margina neto prihoda (%)	Margina bruto prihoda (%)	Ind.troškovi kao postotak prodaje (%)	Margina neto prihoda (%)
Veliki kupac	79,68	32,19	17,08	15,11	11,96 - 39,64	6,37 - 28,37	-8,43 - 33,27
1	3,98	33,20	16,45	16,76	27,02 - 37,01	6,74 - 22,45	9,62 - 30,26
2	4,72	28,89	35,49	-6,60	0,03 - 40,15	5,55 - 377,57	-377,55 - 29,29
3	4,21	32,65	29,44	3,22	5,32 - 57,87	5,78 - 270,46	-254,68 - 37,50
4	5,05	32,28	21,45	10,83	-83,75 - 42,42	3,96 - 115,71	-125,94 - 30,37
5	1,26	34,32	32,49	1,84	-9,97 - 61,82	4,58 - 286,42	-254,21 - 43,98
Ostali	1,11	35,90	121,88	-85,98	-0,77 - 52,50	4,61 - 137,22	-137,29 - 36,12

Zanimljivo je istaknuti da jedino kupci iz grupe 1 (kupci s prometom 500.000,00 - 1 mil. KM) ne ostvaruju negativan profit, a u grupi *Velikih kupaca* su 2 koja su „u minusu“. Primjetan je i rast margine bruto prihoda za manje kupce iz razloga većih cijena za (ne)ugovorene manje količine proizvoda.

Tablica 6.18 Prvih 12 kupaca iz grupe „Velikih kupaca“

Rbr kupca	Udio u ukupnoj prodaji (%)	Margina bruto profita (%)	Udio ind.troškova u ukupnoj prodaji (%)	Margina neto profita (%)	Broj lokacija isporuke	Broj narudžbi	Broj različitih artikala
1	42,89	45,36	34,77	59,07	9	2.391	392
2	10,87	11,44	13,25	9,1	322	35.832	475
3	7,03	3,15	6,62	-1,36	2	305	383
4	4,25	4,67	4,02	5,51	115	7.563	404
5	2,86	3,01	2,97	3,07	49	7.079	303
6	2,59	2,55	3,24	1,66	150	7.902	255

7	2,24	2,33	2,45	2,18	30	3.589	281
8	1,81	2,2	0,63	4,25	1	174	83
9	1,58	1,5	0,79	2,41	2	429	99
10	1,49	1,52	0,78	2,47	40	803	132
11	1,1	1,21	0,8	1,74	27	2.429	189
12	0,96	1,07	0,58	1,7	11	1.141	193

Tablica 6.18 prikazuje pojedinačne pokazatelje za prve 2/3 kupaca iz grupe najvećih. Jasno se uočavaju četiri kupca koja čine gotovo 66% prometa ove grupe. Radi se o kompanijama koje djeluju kroz svoj lanac prodaje u cijeloj regiji. Jedna od tih kompanija ostvaruje negativnu neto maržu (-1,36%) bez obzira na veličinu prometa. Kompleksnost kupca i neki posebni zahtjevi generirali su efekt prebacivanja značajnog dijela indirektnih troškova na ovo poduzeće i spuštanje neto marže ispod nule.

6.4 Obrada podataka

Za potrebe obrade podataka sve nekategorijalne vrijednosti podataka su normalizirane, odnosno svedene na vrijednosti između -1 i 1. Podaci u ovom obliku predstavljaju ulazne i izlazne varijable odabranih metoda istraživanja, a njihova deskriptivna statistika prikazana je u tablici 6.19.

Tablica 6.19 Deskriptivna statistika normaliziranih podataka

Deskriptivna statistika						
	N	Minimum	Maximum	Mean		Std. Deviation
	Statistika	Statistika	Statistika	Statistika	Std. Error	Statistika
T1	1911	-1	1	-0,981920984	0,00284268	0,124267709
T2	1911	-1	1	-0,980344846	0,002811191	0,122891157
T3	1911	-1	1	-0,977594976	0,003260159	0,142517798
T4	1911	-1	1	-0,9854427	0,002710635	0,118495378
T5	1911	-1	1	-0,985443223	0,002710577	0,118492842
T6	1911	-1	1	-0,982413396	0,002977147	0,130145922
T7	1911	-1	1	-0,982653061	0,002965823	0,129650923
C1	1911	-1	1	-0,975271062	0,003059954	0,133765857

C2	1911	-1	1	-0,980200419	0,002923991	0,127822235
C3	1911	-1	1	-0,996398221	0,001364683	0,0596571
VK1	1911	-1	1	-0,985443747	0,002710376	0,118484067
VK2	1911	-1	1	-0,981301413	0,002681767	0,117233388
VK3	1911	1	3	1,720041863	0,01277005	0,558242566
VK4	1911	1	3	1,018315018	0,003787464	0,165568945
KK1	1911	-1	1	-0,786804291	0,005937556	0,259560183
KK2	1911	-1	1	-0,980097855	0,002808712	0,122782809
KK3	1911	1	3	1,193092622	0,010875984	0,475443481
KK4	1911	-1	1	-0,977932496	0,003184061	0,139191198
KK5	1911	-1	1	-0,982305599	0,002977622	0,130166702
PP1	1911	0	1	0,468864469	0,011418516	0,499160256
PP2	1911	0	1	0,354604919	0,009195973	0,402001847
PP3	1911	0	1	0,328424908	0,007684444	0,335925352
PP4	1911	0	1	0,172799581	0,00602376	0,263328595
PP5	1911	0	1	0,030104657	0,002727632	0,119238375
PP6	1911	0	1	0,003354265	0,00089974	0,039332115
PP7	1911	0	1	0,111067504	0,004239296	0,185320752
R	1911	0	1	0,73155416	0,010139929	0,443266837
F	1911	0	1	0,442176871	0,011363957	0,496775226
M	1911	-1	1	-0,985336473	0,002736488	0,119625547

6.4.1 Regresija u predviđanju varijabli profitabilnosti

Upotrebom programskog paketa Weka primijenjene su metoda binarne logističke regresije za predviđanje R i F varijabli, te višestruka linearna regresija za predviđanje varijable M. Prethodno je ukupni set podataka podijeljen na set za učenje (podaci od 2008. do 2012. godine) i set za testiranje dobivenog modela (2013. godina). U narednim poglavljima opisane su pretpostavke predviđanja pojedinih varijabli i rezultati dobiveni regresijskim metodama predviđanja zavisnih varijabli modela.

6.4.1.1 Predviđanje R varijable

Osnovne postavke korištene za primjenu binarne logističke regresije su sljedeće:

1. Ulazne varijable regresije su sve varijable definirane okvirom za profitabilnost kupca u periodu t.
2. Zavisna varijabla sustava je podatak je li kupac nastavlja s prometom (biti kupac) u vremenskom periodu t+1.
3. Metoda uvodi sve nezavisne varijable u analizu bez njihovog ispadanja u narednim koracima.
4. Klasifikacijska vjerojatnost je postavljena na 0,5.

Logit model dobiven je na setu podataka za učenje, a potom je primjenjen na testni set podataka. Rezultati primjene dobivenog modela na testni set (tablica 6.20) pokazuju da je ispravno klasificirano 69,33% slučajeva, a neispravno 30,67% slučajeva.

ROC krivulja (engl. *Receiver Operating Characteristic Curve*) definira odnos senzitivnosti i specifičnosti dobivenog modela. Senzitivnost je vjerojatnost da model klasificira poslovnog partnera u grupu onih koji će ponoviti kupovinu u periodu *t+1* ako će je on i stvarno napraviti. Vrijednost ovog pokazatelja je 0,982 što znači da je 98,2% vjerojatnost prepoznavanja kupovine u narednom periodu ako će se ta kupovina zaista i dogoditi. Specifičnost je vjerojatnost da model ne prepozna poslovnog partnera kao kupca u narednom periodu ako on to zaista i neće biti. Vrijednost ovog koeficijenta je jako niska, odnosno 0,094, što znači da je 9,4% vjerojatnosti prepoznavanja nekupovine u narednom periodu ako se ta kupovina i neće dogoditi.

Tablica 6.20 Rezultati Logit regresijskog modela u predviđanju R varijable

	Variable R			
	Training Data Set		Test Data Set	
Total Number of Instances	1261		326	
Correctly Classified Instances	981	77,7954%	226	69,3252%
Incorrectly Classified Instances	280	22,2046%	100	30,6748%
Kappa statistic	0,2798		0,0983	
Mean absolute error	0,3055		0,3724	
Root mean squared error	0,3866		0,4591	
Relative absolute error	79,1794%		89,4147%	
Root relative squared error	88,0366%		97,1098%	

ROC Area	0,8100		0.6700	
Confusion Matrix	**0**	**1**	**0**	**1**
0	87	242	10	96
1	38	894	4	216
TP Rate	0,2640	0,9590	0,0940	0,9820
FP Rate	0,0410	0,7360	0,0180	0,9060
Precision	0,6960	0,7870	0,7140	0,6920
Recall	0,2640	0,9590	0,0940	0,9820
F-Measure	0,3830	0,3830	0,1670	0,8120

ROC površina je mjera diskriminacije modela, odnosno sveukupne sposobnosti modela da razluči kupce perioda *t* na one koji će kupovati i na one koji neće kupovati u periodu *t+1*. Ovaj pokazatelj za testni model iznosi 0,67 te bi se na osnovu ove vrijednosti model mogao ocijeniti kao „zadovoljavajući“.

6.4.1.2 Predviđanje F varijable

Osnovne postavke korištene za primjenu binarne logističke regresije pri predviđanju varijable F većinom su iste kao i kod prethodnog modela. One su sljedeće:

1. Ulazne varijable regresije su sve varijable definirane okvirom za profitabilnost kupca u periodu *t*.
2. Zavisna varijabla sustava je podatak da li kupac nastavljajući s prometom, u vremenskom periodu *t+1*, ostvaruje isti pri čemu je broj kupovina manji ili veći od median vrijednosti perioda.
3. Metoda uvodi sve nezavisne varijable u analizu bez njihovog ispadanja u narednim koracima.
4. Klasifikacijska vjerojatnost je postavljena na 0,5.

Logit model dobiven je na setu podataka za učenje, a potom je primjenjen na testni set podataka. Tablica 6.21 prikazuje rezultate primjene modela na testni set. Ispravno je klasificirano 73,31% slučajeva, a neispravno 26,69% slučajeva.

Tablica 6.21 Rezultati Logit regresijskog modela u predviđanju F varijable

	Variable F			
	Training Data Set		Test Data Set	
Total Number of Instances	1261		326	
Correctly Classified Instances	1021	80,9675%	239	73,3129%
Incorrectly Classified Instances	240	19,0325%	87	26,6871%
Kappa statistic	0,6035		0,4499	
Mean absolute error	0,2985		0,3400	
Root mean squared error	0,3741		0,4156	
Relative absolute error	60,8169%		68,5074%	
Root relative squared error	75,5253%		82,9805%	
ROC Area	0,8840		0,8620	
Confusion Matrix	**0**	**1**	**0**	**1**
0	646	70	168	4
1	170	375	83	71
TP Rate	0,9020	0,6880	0,9770	0,4610
FP Rate	0,3120	0,0980	0,5390	0,0230
Precision	0,7920	0,8430	0,6690	0,9470
Recall	0,0920	0,6880	0,9770	0,4610
F-Measure	0,8430	0,7580	0,7940	0,6200

Senzitivnost modela je vjerojatnost da on klasificira poslovnog partnera u grupu kupaca koji će ostvariti u periodu *t+1* broj kupovina veći od median vrijednosti perioda, ako će je on i stvarno ostvariti takvu. Vrijednost ovog pokazatelja je 0,461, odnosno 46,1% je vjerojatnost prepoznavanja broja kupovina većeg od median vrijednosti u narednom periodu ako će se takve kupovina zaista i dogoditi. Specifičnost je vjerojatnost da model ne prepozna poslovnog partnera kao kupca s brojem ponavljanja kupovina iznad median vrijednosti u narednom periodu ako on to zaista i neće biti. Vrijednost ovog koeficijenta je 0,977, što znači da je 97,7% vjerojatnosti opisane klasifikacije.

Mjera diskriminacije modela kao sveukupne sposobnosti modela da razluči kupce perioda *t* na one koji će imati broj kupovina iznad ili ispod median vrijednosti perioda *t+1* za testni model iznosi 0,862. Ova vrijednost klasificira model u kategoriju „vrlo dobrog“ modela.

6.4.1.3 Predviđanje M varijable

Za predviđanje M varijable, relativni udio prometa kupca u ukupnom prometu perioda *t+1*, koristila se višestruka linearna regresija i Weka programski paket. Osnovne polazne postavke modela regresije su sljedeće:

1. Ulazne varijable regresije su sve varijable definirane okvirom za profitabilnost kupca u periodu *t*.
2. Zavisna varijabla sustava je podatak o relativnom udjelu kupovine kupca u vremenskom periodu *t+1*.
3. Enter metoda se koristi kao metoda za ulazak varijabli u sustav. Metoda inicijalno uvodi sve nezavisne varijable u model.

Regresijski problem previđanja varijable M dobro je modeliran ukoliko su vrijednosti MSE i MAE minimalne uz što veći koeficijent korelacije *r*. Rezultati dobivenog modela ukazuju na visok stupanj korelacije od 99,24% (tablica 6.22) i niske vrijednosti srednje vrijednosti greške (MAE=0,0036) i korijena srednje vrijednosti kvadrata greške (RMSE=0,015). Model pokazuje gotovo savršenu pozitivnu korelaciju između skupa stvarnih i predviđenih vrijednosti zavisne varijable M.

Tablica 6.22 Rezultati modela višestruke linearne regresije u predviđanju M varijable

	Variable M	
	Training Data Set	Test Data Set
Total Number of Instances	1261	326
Correlation coefficient	0,9958	0,9924
Mean absolute error	0,0031	0,0036
Root mean squared error	0,0110	0,0150
Relative absolute error	12,5579%	14,6900%
Root relative squared error	9,1859%	12,4569%

6.4.2 Neuronska mreža u predviđanju varijabli profitabilnosti

Istraživanje metodom NN odrađeno je upotrebom softverskog paketa NeuroSolutions ver.6.3, uz dodatak paketa NeuroSolutions for Excel. Prethodnica izgradnji modela je priprema podataka. Nakon uvoza podataka u MS Excel, deklarirane su ulazne varijable

sustava, pri čemu su varijable VK3, VK4, KK3 i PP1 deklarirane kao simboličke kolone. Prema pretpostavci istraživanja, podaci iz 2008.-2011. godine označeni su kao trening set podaci (65,97%), podaci iz 2012. godine kao validacijski set (16,94%), a oni iz 2013. godine kao set podataka za testiranje (17,09%) dobivene arhitekture.

Po formiranja skupova NeuroSolutions paket osigurava automatsko testiranje sljedećih arhitektura NN: MLP-Multi Layer Perceptron, LR-Linear Regression, PNN-Probabilistic Neural Network, RBF-Radial Basis Function, MLPPCA-Multi Layer Perceptron with Principle Component Analysis, SVM- Classification Support Vector Machine, GFF-Generalized Feedforward, TDNN- Time-Delay Network, TLRN- Time-Lag Recurrent Network and RN- Recurrent Network. Softver na osnovu broja ulaznih varijabli i deklariranih modela NNs, bira početne konfiguracije, provodi učenje nad prva dva seta podataka i validaciju generiranih modela nad testnim setom podataka. Validacija se vrši usporedbom pokazatelja performansi generiranih NN modela. Kod klasifikacijskih problema koriste se korelacijski koeficijent ***r*** i postotak korektnih klasifikacija, a kod regresijskog problema srednja vrijednost kvadrata greške ***MSE***, korelacijski koeficijent ***r*** i srednja vrijednost apsolutne vrijednosti greške ***MAE***.

Popis svih testiranih arhitektura neuronske mreže po predviđanju pojedinih varijabli prikazane su u tablicama 5.6, 5.7 i 5.8. Tablica 6.23 daje pregled najboljih arhitektura dobivenih procesom selekcije i pripadajuće parametre.

Tablica 6.23 Rezultati najboljih modela neuronske mreže

Varijabla	Model	Problem	MSE	r	MAE	% ispravnih
R	MLP-1-B-M (Multilayer Perceptron)	Klasifikacijski	0,2076	0,3091	-	71,56%
F	MLP-1-B-L (Multilayer Perceptron)	Klasifikacijski	0,1109	0,7476	-	87,16%
M	MLPPCA-1-O-M (MLP with PCA)	Regresijski	0,0001	0,9955	0,0051	-

Najbolji modeli mreže podvrgnuti su dodatnim testiranjima i procesu optimizacije. Korištene su metode:

i. ponavljanja učenja n puta,
ii. varijacije parametara,
iii. izostavljanja n primjera i
iv. genetičkog algoritma.

Primjena metode genetičkog algoritma jedina je dala bolje rezultate od polaznih (tablica 6.24), pri čemu su bolji rezultati pokazani na klasifikacijskim problemima predviđanja varijabli R i F.

Tablica 6.24 Rezultati optimiziranih modela neuronske mreže upotrebom genetičkog algoritma

Varijabla	Model	Problem	MSE	r	MAE	% ispravnih
R	MLP-1-B-L (Multilayer Perceptron)	Klasifikacijski	0,1968	0,3297	-	75,34%
F	MLP-1-B-L (Multilayer Perceptron)	Klasifikacijski	0,1098	0,7499	-	89,62%
M	MLPPCA-1-O-M (MLP with PCA)	Regresijski	0,0001	0,9955	0,0051	-

6.4.3 Regresija u ocjeni poslovne politike na profitabilnost kupca

Poslovnu politiku poduzeća prema postavkama ovog istraživanja čine troškovna politika definirana varijablama T_1-T_7 i cjenovna politika definirana varijablama C1-C3. Analiza efekata ovih grupa nezavisnih varijabli na veličinu profitabilnosti kupca (zavisna varijabla M), promatra se kroz sljedeće regresijske modele:

$$M = f(T_1, T_2, T_3, T_4, T_5, T_6, T_7) \qquad (6.1)$$

$$M = f(C_1, C_2, C_3) \qquad (6.2)$$

Sinteza bitnih elemenata definiranih regresijskih modela dana je u tablicama 6.25 i 6.26.

Tablica 6.25 Utjecaj troškova na profitabilnost

Output	Adjusted R^2	F	Durbin-Watson	Sig		Nezavisne varijable – Troškovi				
						T_1	T_2	T_3	T_4	T_7
M	0,9908	41038,3741	1,9778	0,0000	Sig	0,1759	1,32E-16*	1,17E-09*	1,95E-40*	4,30E-10*
					B	0,0892	0,0838	-0,2532	0,8971	0,2597

**statistički značajna varijabla uz 5% signifikantnosti*

Forsiran je ulazak svih varijabli troškova (tablica 6.25) u regresijski model (Enter metoda), a SPSS je iz modela izuzeo varijable T_5 i T_6. Adjusted R^2 promatranog modela je 0,9908 što znači da linearna regresija objašnjava 99,08% varijance u podacima. Durbin-Watson koeficijent d ima vrijednost 1,9778 ($1.5 < d < 2.5$) iz čega se može pretpostaviti da ne postoji linearna autokorelacija u regresijskom modelu. F-test modela linearne regresije ima nultu hipotezu da ne postoji linearni odnos između varijabli i u predmetnom modelu regresije pokazuje značajnu signifikantnost tako da možemo pretpostaviti linearnu relaciju među varijablama modela.

Sve varijable modela su statistički značajni prediktori veličine prihoda od kupca u periodu *t+1* ($p<0,05$) osim varijable T_1-Troškovi prodaje i direktnog marketinga ($p=0,1759 > 0,05$). B vrijednost za sve varijable osim T_3-Troškovi isporuke kupcu je pozitivna. Podatak ukazuje na povećanje relativnog udjela prihoda od kupca za jedinično povećanje promatrane nezavisne varijable troška pri čemu ostale nezavisne varijable ostaju konstantne. Kod varijable T_3, ovaj odnos je negativan odnosno ukazuje na smanjenje promatrane zavisne varijable.

Kod modela koji promatra utjecaj cjenovne politike na profitabilnost (tablica 6.26) također je korištena Enter metoda i forsiran ulazak svih varijabli cjenovne politike u regresijski model. Adjusted R^2 promatranog modela je 0,0843 što znači da linearna regresija objašnjava 8,43% varijance u podacima. Vrijednost Durbin-Watson koeficijenta iznosi $d=2,0092$ ($1.5 < d < 2.5$), iz čega se može pretpostaviti da ne postoji linearna autokorelacija u regresijskom modelu.

Tablica 6.26 Utjecaj cjenovne politike na profitabilnost

Output	Adjusted R^2	F	Durbin-Watson	Sig		Nezavisne varijable - Cjenovna politika		
						C_1	C_2	C_3
M	0,0843	59,6674	2,0092	7,231E-37	Sig	0,1194	0,0028*	0,3792
					B	0,0901	0,1811	0,0387

**statistički značajna varijabla uz 5% signifikantnosti*

Varijable modela C1-Individualni rabat i C3-Količinski rabat statistički nisu značajni prediktori veličine prihoda od kupca u periodu *t+1* ($p>0,05$). Varijabla C2-Teritorijalni rabat je statistički značajna u ovom smislu ($p=0,0028 < 0,05$). B pokazatelj nestandardnog koeficijenta je pozitivan za sve varijable. Podatak ukazuje na povećanje relativnog udjela

prihoda od kupca za jedinično povećanje relativnog udjela datog pojedinog rabata kupcu pri čemu ostale nezavisne varijable (ostali tipovi rabata) ostaju konstantne.

6.4.4 Regresija u ocjeni karakteristika kupca na profitabilnost kupca

Analiza utjecaja varijabli karakteristika kupca (VK-veličina kupca, KK-kompleksnost kupca i PP-ponašanje pri plaćanju obveza) na veličinu profitabilnosti kupca u periodu *t+1*, provedeno je kroz sljedeće regresijske modele:

$$M = f(VK_1, VK_2, VK_3, VK_4) \qquad (6.3)$$

$$M = f(KK_1, KK_2, KK_3, KK_4) \qquad (6.4)$$

$$M = f(PP_1, PP_2, PP_3, PP_4, PP_5, PP_6, PP_7) \quad (6.5)$$

Sva tri modela koriste se Enter metodom za ulazak varijabli u model pri čemu se forsira ulazak svih varijabli.

Analiza utjecaja podataka o veličini kupca na prihod narednog vremenskog perioda (tablica 6.27) pokazuje (Adjusted R^2 = 0,9899) da linearna regresija objašnjava 98,99% varijance u podacima. Vrijednost Durbin-Watson koeficijenta iznosi d=2,0281 (1.5 < d < 2.5), iz čega se pretpostavlja da ne postoji linearna autokorelacija u regresijskom modelu. F-test modela linearne regresije ima nultu hipotezu da ne postoji linearni odnos između veličine kupca i ostvarenog prihoda perioda *t+1*, a u predmetnom modelu regresije pokazuje značajnu signifikantnost tako da možemo pretpostaviti linearnu relaciju među varijablama modela.

Tablica 6.27 Utjecaj veličine kupca na profitabilnost

Output	Adjusted R^2	F	Durbin-Watson	Sig		Nezavisne varijable - Veličina kupca			
						VK_1	VK_2	VK_3	VK_4
M	0,9899	46902,8	2,0281	0,0000	Sig	0,0000*	0,0039*	0,2116	0,0156*
					B	1,0650	-0,0608	0,0006	-0,0043

**statistički značajna varijabla uz 5% signifikantnosti*

Veličina prometa kupca, teritorijalna pripadnosti domaćem ili inozemnom teritoriju i ostvarena bruto marža su statistički značajni prediktori veličine prihoda, dok varijabla veličine kupca (malo/srednje/veliko) nije statistički značajna. Beta koeficijent je negativan uz jako nisku vrijednost za vrijednost ostvarene bruto marže u prometu s kupcem i ppokazatelj domicilne/inozmene kompanije, dok je pozitivan i ima izrazito nisku vrijednost za elemente klasifikacije kupca prema zakonskom okviru (B=0,0006). Sukladno ovome, varijable VK_2, VK_3 i VK_4 gotovo nikako ne utječu na relativni udio prihoda od kupca za jedinično povećanje promatrane nezavisne varijable pri vrijednostima ostalih koje se ne mijenjaju. Očekivano, varijabla VK_1-Veličina prometa s kupcem ima pozitivnu vrijednost (B=1,065) i uz značajnu signifikantnost pokazuje jasan i snažan utjecaj na zavisnu varijablu novčanog iskaza prometa s kupcem u narednom vremenskom periodu.

Model utjecaja podataka o kompleksnosti kupca na prihod narednog vremenskog perioda (tablica 6.28) pokazuje (Adjusted R^2 = 0,2728) da linearna regresija objašnjava 27,28% varijance u podacima. Vrijednost Durbin-Watson koeficijenta iznosi d=2,0281 (1.5 < d < 2.5) i ukazuje da ne postoji linearna autokorelacija u regresijskom modelu. Nulta hipoteza F-testa modela je da ne postoji linearni odnos između varijabli kompleksnosti kupca i ostvarenog prihoda perioda *t+1*. F vrijednost pokazuje značajnu signifikantnost tako da možemo pretpostaviti linearnu relaciju među varijablama modela.

Tablica 6.28 Utjecaj kompleksnosti kupca na profitabilnost

Output	Adjusted R^2	F	Durbin-Watson	Sig		Nezavisne varijable - Kompleksnost kupca				
						KK_1	KK_2	KK_3	KK_4	KK_5
M	0,2728	144,3574	2,0281	4,01E-130	Sig	7,71E-31*	9,87E-17*	0,0044*	3,71E-31*	1,43E-35*
					B	0,1368	0,5480	-0,0177	0,4238	-0,7345

**statistički značajna varijabla uz 5% signifikantnosti*

Sve varijable kompleksnosti su statistički značajni prediktori zavisne varijable modela. Beta koeficijent varijabli KK_1,KK_2 i KK_4 je pozitivan, a kod varijabli KK_3- Broj lokacija isporuke i KK5-Broj transakcija povrata robe, je negativan. Varijable KK_1, KK_2, KK_4 i KK_5 imaju apsolutnu vrijednost koeficijenta B manju od 1 i relativno veliku vrijednost, za razliku od varijable KK_3. Sukladno navedenom, varijabla KK_3 slabo utječe na relativni udio

prihoda od kupca za jedinično povećanje promatrane nezavisne dok ostale varijable ove grupe pokazuju jasan utjecaj na zavisnu varijablu modela.

Analiza zadnjeg regresijskog modela (tablica 6.29) ukazuje na jako nisku vrijednost Adjusted R^2 koeficijenta, tj. da model objašnjava 0,11 % varijance u podacima. Analizom značajnosti svih varijabli koje definiraju ponašanje kod plaćanja kupca, može se zaključiti da su, osim varijable $PP_{1,}$ sve varijable statistički značajne i da postoji njihov značajan utjecaj na predviđanje prometa s kupcem u narednom vremenskom periodu *t+1*.

Tablica 6.29 Utjecaj ponašanja pri plaćanju obveza kupca na profitabilnost

	Nezavisne varijable - Ponašanje u plaćanju kupca						
	Output	Adjusted R^2	F	Durbin-Watson	Sig		
	M	0,0117	4,2464	2,0042	0,0001		
	PP_1	PP_2	PP_3	PP_4	PP_5	PP_6	PP_7
Sig	0,4529	7,09E-06*	3,89E-06*	3,62E-06*	7,57E-06*	3,49E-05*	3,83E-06*
B	-0,0059	0,7946	0,8161	0,8199	0,7979	0,7862	0,8116

**statistički značajna varijabla uz 5% signifikantnosti*

6.5 Interpretacija rezultata

Prva hipoteza istraživanja pretpostavlja da metode neuronske mreže mogu predvidjeti profitabilnost kupaca jednako dobro ili bolje od statističkih metoda logističke i višestruke linearne regresije. U provođenju empirijskog dijela istraživanja, korištene su metode neuronske mreže za predviđanje sve tri zavisne varijable modela uz uporabu NeuroSolutions programskog paketa, Logit metoda za predviđanje R i F varijabli, te višestruka linearna regresija za predviđanje M varijable. Logit metoda primjenjena je upotrebom Weka softvera, a višestruka linearna regresija SPSS programskim paketom. Tablica 6.30 daje sumarne rezultate poređenja ovih metoda.

U predviđanju svih varijabli profitabilnosti pojedinačnog kupca, metode neuronske mreže pokazale su bolje rezultate od odgovarajućih metoda multivarijacijske statističke analize. Kod klasifikacije zavisnih varijabli R i F, neuronska mreža ima veći postotak korektnih predviđanja. U odnosu na logističku regresiju, kod neuronske mreže su manje i

srednje vrijednosti kvadrata greške, te i mnogo bolji omjer predviđanja pozitivnih (1) i negativnih vrijednosti (0) zavisne varijable.

Tablica 6.30 Usporedba rezultata metoda NN i multivarijantne statističke analize

Glavna hipoteza	Varijabla	Metoda	% Correct	% Positive hit rate (1)	% Negative hit rate (0)	MSE	Correlation (r)
1	R	Logit	69,33%	98,18%	9,43%	0,3724	
		NN	71,56%	71,92%	68,57%	0,2076	
	F	Logit	73,31%	46,10%	97,67%	0,3400	
		NN	87,16%	84,66%	90,58%	0,1109	
	M	Linear Regression				0,0002	0,9924
		NN				0,0001	0,9955

Regresijski problem predviđanja zavisne varijable M pokazao je također neuronsku mrežu kao bolju metodu. Korelacijski koeficijent je gotovo jednak, a i srednja vrijednost kvadrata greške je nešto malo manja kod modeliranja problema upotrebom neuronske mreže. Ovim rezultatima je pokazano da metode neuronske mreže jednako dobro (varijabla M) ili bolje (varijable R i F) od metoda multivarijacijske statističke analize predviđaju profitabilnost kupaca. Ovim je dokazana prva hipoteza istraživanja.

Druga hipoteza istraživanja pretpostavlja da politika prema kupcima utječe na profitabilnost kupca. Troškovna i cjenovna komponenta politike prema kupcima obrađene su empirijskim dokazivanjem dvije pomoćne hipoteze ove glavne hipoteze istraživanja. Treća hipoteza istraživanja postavlja u odnos karakteristike kupca i profitabilnost kupca. Kroz tri pomoćne hipoteze empirijskim dokazivanjem nastojalo se potvrditi ovu glavnu hipotezu. Rezultati empirijskog postupka prikazani su u tablicama 6.31 i 6.32.

Tablica 6.31 Rezultati metoda multivarijantne statističke analize u dokazivanju 2. i 3. hipoteze istraživanja

Glavna hipoteza	Pomoćna hipoteza	Input	Adjusted R^2	F	Sig
2	2.1	T	0,9907	41038,3741	0,0000
	2.2	C	0,0843	59,6674	7,23E-37
3	3.1	VK	0,9899	46902,8474	0,0000
	3.2	KK	0,2728	144,3573	4,01E-130
	3.3	PP	0,0117	4,2464	0,0001

Model višestruke linearne regresije pokazao je varijablu T_1-Trošak prodaje i direktnog marketinga kao neznačajnu (tablica 6.32), dok su varijable T_5-trošak pripreme narudžbi od dobavljača i T_7- Trošak skladišnog prostora i kapitala izbačene iz modela kao neznačajni u samom začetku testiranja modela. Ostale varijable ovog modela pokazale su statističku značajnost i pokazale utjecaj na veličinu M komponente profitabilnosti kupca u vremenskom periodu *t+1*. Sveukupni model objašnjava 99% totalne varijabilnosti u veličini profitabilnosti kupca. Kako je $p<0,05$, možemo odbaciti nultu hipotezu modela regresije da model nema snagu interpretacije promatranog sustava. Ovim je dokazana prva pomoćna hipoteza (H2.1) druge glavne hipoteze istraživanja.

Uz značajnu statističku signifikantnost utvrđeno je da vrijednost danog rabata utječe na M komponentu profitabilnosti kupca, pri čemu su varijable individualnog (C_1) i količinskog (C_3) rabata statističke neznačajne, a varijabla teritorijalnog (C_2) rabata je statistički značajna. Količinski rabat (C_3) se odobrava na kraju vremenskog perioda i vezan je za ukupni ostvareni promet. Ovaj zadnji rabat uglavnom i ostvaruju najveći kupci koji po prirodi nastavljaju prometovanje u opsegu sličnom onom iz prethodnog perioda. Sveukupni model objašnjava 8,44% varijance varijable M i statistički je značajan ($p<0,05$). Ovim je dokazana i druga pomoćna hipoteza koja definira da cjenovna politika utječe na profitabilnost kupca.

Tablica 6.32 Značajnost nezavisnih varijabli u regresijskim modelima pri dokazivanju 2. i 3. hipoteze istraživanja

Glavna hipoteza	Pomoćna hipoteza	Varijable						
		Sig						
2	2.1	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**
		0,1759	1,33E-16*	1,18E-09*	1,95E-40*	Excluded	Excluded	4,31E-10*
	2.2	**C1**	C2	**C3**				
		0,1194	0,0028*	0,3792				
3	3.1	**VK1**	**VK2**	**VK3**	**VK4**			
		0,0000*	0,0039*	0,2116	0,0156*			
	3.2	**KK1**	**KK2**	**KK3**	**KK4**	**KK5**		
		7,71E-31*	9,80E-17*	0,0044*	3,71E-31*	1,43E-35*		
	3.3	**PP1**	**PP2**	**PP3**	**PP4**	**PP5**	**PP6**	**PP7**
		0,4529	7,09E-06*	3,83E-06*	3,62E-06*	7,57E-06*	3,49E-05*	3,83E-06*

**statistički značajna varijabla uz 5% signifikantnosti*

Prosudbom dvije pomoćne hipoteze, može se zaključiti da je dokazana druga hipoteza istraživanja, tj. da politika koju poduzeće vodi prema pojedinom kupcu direktno utječe na njegove mjere profitabilnosti.

Varijable veličine kupca pokazale su očekivani utjecaj na zavisnu varijablu modela. Model objašnjava 99% varijance veličine profitabilnosti kupca u periodu *t+1* i statistički je značajan. Ključnim i signifikantnim, s jasnim i snažnim utjecajem na zavisnu varijablu M, pokazale su se varijable VK_1-Veličina prometa s kupcem i VK_2-Bruto marža ostvarena u prometovanju s kupcem. Ostale dvije varijable pokazale su statistički neznačajan utjecaj (VK_3-klasifikacija veličine kupca i VK_4-teritorijalna pripadnost). Ipak, snažan utjecaj varijabli VK_1 i VK_2 koje i čine osnovne značajke veličine kupca u promatranom modelu, dokazuje prvu pomoćnu hipotezu (H3.1) treće glavne hipoteze istraživanja.

Sve varijable koje karakteriziraju kompleksnost kupca (KK1-KK_5) pokazale su jasan utjecaj na veličinu profitabilnosti kupca uz objašnjenje 27,29% njene varijance. Ovim je dokazana pomoćna hipoteza H3.2 koja pretpostavlja utjecaj kompleksnosti kupca na njegovu profitabilnost.

Treća pomoćna hipoteza također je dokazana. Regresijski model pokazuje statističku značajnost pri čemu su, osim varijable PP_1, sve ostale varijable elementata ponašanja u

plaćanju kupaca (PP_2-PP_7), statistički značajne i utječu na predviđanje prometa s kupcem u narednom vremenskom periodu *t+1*.

Prosudbom navedene tri pomoćne hipoteze, može se zaključiti da je dokazana treća hipoteza istraživanja, tj. da karakteristike kupca direktno utječu na mjere profitabilnosti kupca.

7. ZAKLJUČAK

Klasične metode natjecanja i stjecanja prednosti na globalnom tržištu polako postaju stvar prošlosti. Suvremeno poduzeće raspolaže s ogromnom količinom podataka o poslovnom okruženju i vlastitom poslovanju. Ono ih mora iskoristiti na način da postigne kompetitivne prednosti, uspostavi efikasne procese i postigne optimalan poslovni rezultat. Odnos s kupcem čini jedan, ali veoma bitan kotačić u navedenom mozaiku. Poduzeća svih veličina moraju težiti idealu uspostavljanja bliske relacije (1:1) sa svojim kupcima, razumjeti svakoga kupca posebno, te to razumijevanje iskoristiti na način da uvjere kupca kako je bolje da s njima posluju nego s konkurencijom.

Istraživanja na polju profitabilnosti pojedinačnih kupaca pokazala su da svi kupci ne vrijede jednako. Čak i više od toga, istraživanja tekuće profitabilnosti dovela su do zaključaka da se Paretovo pravilo 20/80 (20% kupaca generira 80% profita) drastično mijenja (20% kupaca generira u ekstremnim slučajevima i više do 200-250% profita). Dakle, pojedini kupci nisu uopće profitabilni i ta velika grupa kupaca ima marginu približnu 0 ili negativnu. Prepoznavanje ovih trendova i empirijsko mjerenje uključuje promjenu računovodstvenih metoda i tehnika obračuna troškova.

Izračun tekuće profitabilnosti kompleksan je postupak. Za njegove potrebe neophodno je razviti opći okvir koji će sadržavati sve odgovarajuće ulazne varijable, te i mjere kojima će se iskazivati profitabilnost. Osnovu čine linearne mjere troškova i cijena, a pored njih tu je i niz linearnih i nelinearnih varijabli koje opisuju osobine kupaca i osobine poduzeća čije se poslovanje s kupcima analizira. Poduzeće se promatra u užoj i široj okolini koja svojim karakteristikama utječe na poslovanje. Svaka od navedenih mjera utječe različitim intenzitetom na pripadajuću latentnu varijablu. Latentne varijable (troškovi, cijene, kupac) pak koreliraju međusobno te određenim intenzitetom utječu na mjere profitabilnosti. Izračun koeficijenata korelacije, diferenciranje baze kupaca i jasni pokazatelji pojedinačnih profitabilnosti čine osnovu za donošenja odluka u procesima upravljanja kupcima.

Metode strojnog učenja putem različitih algoritama mogu izvući i apsorbirati skriveno znanje u postojećim podacima. Ovakvo samostalno učenje na osnovi podataka iz prošlosti, može generirati pokazatelje različitih aspekata profitabilnosti kupaca kroz primjenu u problemskim domenama poput klasteriranja, asocijacije ili klasifikacije podataka. Jedna od najvažnijih osobina im je da su, teoretski, sposobne aproksimativno odrediti bilo koju

neprekinutu funkciju koja opisuje realni svijet. Dakle, metode strojnog učenja, kombinirane s nekima od statističkih metoda multivarijantne analize, čine dobar izbor pri odabiru alata za generiranje holističkog pogleda na pitanje tekuće profitabilnosti kupaca i za predviđanje buduće profitabilnosti kupaca.

Na osnovu navedenog, nametnuo se zaključak kako je neophodno provesti istraživanje i dati znanstveni doprinos definiranjem modela za izračun tekuće i predviđanje buduće profitabilnosti kupaca te prikazati rezultate korištenja i usporedbe klasičnih statističkih metoda multivarijacijske analize s metodama umjetne inteligencije. Provedeno empirijsko istraživanje i na njemu nastala disertacija, plod su prepoznavanja ove potrebe i nastojanja da se na ovom polju da odgovarajući znanstveni i pragmatični doprinos.

7.1 Provjera postavljenih hipoteza istraživanja

Provedenim istraživanjem formiran je model, te su nad njim provjerene ispravnosti postavljene tri glavne i pet pomoćnih hipoteza istraživanja.

Prva glavna hipoteza istraživanja postavlja u odnos metode kojima se služilo istraživanje. Ona tvrdi kako neuronske mreže, jedna od metoda strojnog učenja, mogu jednako dobro ili bolje predvidjeti pokazatelje profitabilnosti pojedinačnog kupca od multivarijacijskih statističkih metoda binarne logističke regresije i višestruke linearne regresije. Predviđanje sve tri zavisne varijable modela provedeno je metodom neuronske mreže. Zbog prirode zavisnih varijabli morale su se koristiti različite metode statističke analize. Logistička regresija pokazala se odgovarajućom za binarnu prirodu R (ponavljanje kupnje) i F (učestalost buduće kupnje) varijabli. Višestruka linearne regresija se koristila za predviđanje monetarne vrijednosti buduće kupnje (M varijabla).

Rezultati istraživanja potvrdili su prvu hipotezu. Klasifikacija kupaca u odnosu na buduće ponavljanje kupnje točnije je predviđano NN metodom u odnosu na logit metodu (71,56% : 69,33%). Sveukupna točnost predviđanja dodatno je dobila na značaju ravnomjernijim rasporedom stopa predviđanja pozitivnog (dogoditi će se ponovna kupnja) i negativnog (neće se dogoditi ponovljena kupnja) ishoda zavisne varijable (**1:** 71,92% : 98,18% , **0:** 68,57% : 9,43%). Ova blaga prednost metode strojnog učenja u predviđanje R varijable, dodatno je povećana i utvrđena u predviđanju učestalosti buduće kupnje. Metoda

neuronske mreže postigla je značajno veću točnost u odnosu na logit metodu (87,16% : 73,31%).

Regresijski problem predviđanja monetarne vrijednosti pokazao je također neuronsku mrežu kao bolju metodu. Dok je korelacijski koeficijent gotovo jednak (0,9955 : 0,9924), srednja vrijednost kvadrata greške je ipak značajno manja pri upotrebi neuronske mreže (0,0001 : 0,0036).

Druga hipoteza postavlja u odnos politiku koju poduzeće provodi prema kupcu i njegovu profitabilnost. Glavnom hipoteza tvrdi da postoji direktna ovisnost, a dokazivala se kroz provjeru pomoćnih hipotezi. Slijedeći logiku projektiranog modela koja je politiku prema kupcu definirala troškovnom i cjenovnom komponentom, pomoćne hipoteze su ih postavile u odnos s monetarnom vrijednošću profitabilnosti pojedinačnog kupca. Priroda zavisne varijable definirala je višestruku linearnu regresiju kao metodu kojom se empirijski dokazuju tvrdnje definirane pomoćnim hipotezama.

Prva pomoćna hipoteza u odnos stavlja troškovnu politiku i profitabilnost kupca. Postavljeni regresijski model objašnjava 99% totalne varijabilnosti monetarne vrijednosti profitabilnosti. Tri varijable su se pokazale kao statistički neznačajne, dok su ostale pokazale statističku značajnost i utjecaj na veličinu M u budućem vremenskom periodu (*t+1)* u odnosu na onaj iz kog su podaci o troškovima. Zanimljivo je istaknuti kako je istraživanje pokazalo da je T_1-Trošak prodaje i direktnog marketinga jedna od statistički neznačajnih varijabli. Ova varijabla je u originalnom modelu (Niraj, 2001), preuzetom za potrebe izračuna tekuće profitabilnosti, izuzeta iz izračuna, a ovim istraživanjem je uključena. Kako je $p<0,05$, odbačena je nulta hipotezu modela regresije da model nema snagu interpretacije promatranog sustava, odnosno dokazana je prva pomoćna hipoteza druge glavne hipoteze istraživanja.

U postupku dokazivanja druge pomoćne hipoteze, pokazana je značajna statistička značajnost tvrdnje da vrijednost danog rabata utječe na M komponentu profitabilnosti kupca. Sveukupni model objašnjava 8,44% varijance monetarne vrijednosti buduće kupnje i statistički je značajan ($p<0,05$). Ovim je dokazana i druga pomoćna hipoteza koja definira da cjenovna politika utječe na profitabilnost kupca. Dokazane obje pomoćne hipoteze potvrdile su drugu osnovnu hipotezu istraživanja.

Treća hipoteze postavila je u odnos heterogenost skupa kupaca tvrdeći da ona direktno utječe na veličinu mjera profitabilnosti kupca. Kako su karakteristike kupca grupirane u kategorije veličine, kompleksnosti i ponašanja pri plaćanju obveza, tako su formirane tri pomoćne hipoteze sukladno predstavljenim kategorijama i za potrebe dokazivanja formirani odgovarajući regresijski modeli.

Veličina kupca definirana je financijskim i nefinancijskim elementima. Financijske elemente čine veličina prometa i ostvarena marže, a nefinancijske pripadnost grupi kupaca sukladno veličini i teritorijalna pripadnost. Model objašnjava 99% varijance veličine profitabilnosti kupca i statistički je značajan. Financijske varijable pokazale su očekivani utjecaj na zavisnu varijablu modela, dok su ostale dvije, nefinancijske, pokazale statistički neznačajan utjecaj. Očekivano, snažan utjecaj varijabli VK_1-veličina prometa i VK_2-ostvarena marža, koje i čine osnovne značajke veličine kupca u promatranom modelu, dokazuje prvu pomoćnu hipotezu treće glavne hipoteze istraživanja.

Kompleksnost kupca definira stupanj složenosti poslovnog odnosa s kupcem. Modelirana je varijablama koje kvantificiraju količinu narudžbi od kupca, različitost asortimana isporučenih proizvoda, broj mjesta isporuke te količinu i vrijednost transakcija povrata robe. Sve navedene varijable pokazale su jasan utjecaj na veličinu profitabilnosti kupca uz objašnjenje 27,29% njene varijance čime je dokazana druga pomoćna hipoteza treće glavne hipoteze istraživanja.

Ponašanje pri plaćanju definira način ponašanja kupca u finalnom koraku poslovnog odnosa na relaciji promatranog poduzeća i pojedinačnog kupca. Varijable ove kategorije određuju iznose potraživanja naplaćene na vrijeme, vremensku strukturu naplate potraživanja van ugovorenog vremena plaćanja, nenaplaćena potraživanja i prosječno vrijeme kašnjenja u plaćanju. Regresijski model pokazao je statističku značajnost. Sve varijable, osim varijable PP_1-vrijeme kašnjenja u plaćanju kupca, pokazale su se statistički značajnim u utjecaju na predviđanje prometa s kupcem u narednom vremenskom periodu *t+1*. Sve tri pomoćne hipoteze su dokazane, a time i treća glavna hipoteza istraživanja.

7.2 Ograničenja provedenog istraživanja

Ne postoje idealno vrijeme i uvjeti za provedbu bilo kojeg istraživanja. Svako se odvija u uvjetima specifičnih ograničenja, što je bilo i slučaj s ovim koje je bilo doista kompleksno. Vrijeme u kojem je provedeno i na koje se odnose podaci istraživanja (2008. – 2013.godina), karakterizirano je globalnom svjetskom recesijom, odnosno usporavanjem globalne privredne aktivnosti i pogoršanjem indikatora opće ekonomske klime. U uvjetima recesije karakterističan je odgovor kupaca na nastale okolnosti, a ogleda se promjenom ponašanja u potrošnji. Uobičajeno je smanjeno i mnogo pažljivije trošenje, a u centru istraživanja se nalazi profitabilnost kupaca. Zato je potrebno uzeti u obzir ova obilježja promatranog vremenskog perioda i njihov utjecaj na dobivene rezultate istraživanja.

Podaci za potrebe istraživanja dobiveni su iz poduzeća koje pripada industriji hrane. Prerađivački industrija i u njoj sektor proizvodnje hrane su među najznačajnijim sektorima u Bosni i Hercegovini. Prema podacima Agencije za statistiku BiH, udio industrije hrane u bruto domaćem proizvodu iznosio je 2008.godine 1,82% (prerađivačka industrija – 11,01%), a 2013.godine 1,89% (prerađivačka industrija 10,94%). Vidljivi su pokazatelji rasta u periodu 2008.-2010.godina, ali i stagnacije i laganog pada u periodu 2010.-2013.godine. Neki trendovi u industriji hrane s posebnim naglaskom na zdravu i ekološki uzgojenu hranu, dali su ipak posebnu priliku poduzećima poput ovog koje je ustupilo podatke za istraživanje. Orijentacija no nove proizvode i tržišta pokazala se ključnom u opstanku i nastavku razvoja u uvjetima nepovoljnih uvjeta poslovanja uzrokovanih recesijskim kretanjima globalnog tržišta.

Sljedeće ograničenje s kojim se susrelo istraživanje, odnosilo se na izvor podataka. Podaci su preuzeti iz skladišta podataka promatranog poduzeća. Određene kategorije podataka, a prije svega misli se na strukturu indirektnih troškova o pojedinačnom kupcu, nisu mogli da se preuzmu u izvornom obliku. Bila su neophodni dodatni izračuni uz preuzimanje podataka iz upravljačkog informacijskog sustava.

7.3 Preporuke za buduća istraživanja

Projektirani model i dobiveni empirijski pokazatelji istraživanja ukazuju na nekoliko pravaca mogućih istraživačkih napora koji bi dali dodatne doprinose na polju izučavanja metoda strojnog učenja u predviđanju profitabilnosti kupaca.

Prije svega, korištena je jedna metoda strojnog učenja – metoda neuronskih mreža. Postoji čitav niz drugih metoda koje vjerojatno mogu jednako dobro ili bolje modelirati klasifikacijske i regresijske probleme definirane ovim istraživanjem. Kako je potvrđena uspješnost metode neuronske mreže, bilo bi zanimljivo provjeriti i usporediti ostale metode strojnog učenja nad ovim problemom.

Informacijska osnovica istraživanja modela za predviđanje profitabilnosti preuzeta je iz industrije hrane u Bosni i Hercegovini. Ova industrija ima svoje specifičnosti i pitanje je koliko bi model bio primjenjiv na druge industrijske grane. Izazov budućih istraživanja mogao bi se ogledati u pokušaju adaptacije i primjene modela na druge industrijske grane, ali i na pronalasku izvora podataka u poduzećima s razvijenih i uređenijih tržišta poput onih u razvijenim zemljama EU.

U početku istraživanja naišao sam na određena neodobravanja kolega u pogledu velikog broja nezavisnih varijabli modela. Potaknut podrškom mentora, ohrabren velikom računalnom moći današnjih hardverskih resursa koji stoje na raspolaganju istraživaču te modernim softverskim paketima za obradu podatka, ostao sam na ideji modeliranja problema sa što većim brojem pokazatelja koji direktno ili indirektno utječu na mjere profitabilnosti kupca. Postoji još prostora na ovom polju kako bi se neke nove varijable uključile u definiciju modela, a neke koje su se pokazale neznačajnim isključile iz modela. Prethodno spomenuti globalni uvjeti poslovanja, koji su imali recesijsku notu u slučaju ovog istraživanja, najbolji su primjer. Oni su snažan indikator da bi bilo poželjno u narednim istraživanjima uključiti i pokazatelje poslovne okoline kako bi poduzeće i njegove kupce promatrali u još širem i globalnijem aspektu. Svakako da će, na ovom tragu, istraživanja u novim industrijama predstaviti i neke specifične varijable koje karakteriziraju poslovni odnos s kupcem.

LITERATURA

1. Agencija za statistiku BiH (2005). ***TB-01 Bruto domaći proizvod prema proizvodnom, dohodovnom i rashodnom principu***. www.bhas.ba, [29.1.2016.].
2. Ayres, I. (2007) ***Super Crunchers***.New York: Bantam Dell.
3. Abraham, B. i Ledolter, J. (2005). ***Statistical Methods for Forecasting***. John Willey & Sons Inc.
4. Bahovec, V. i Erjavec, N. (2009). ***Uvod u ekonometrijsku analizu***. Zagreb: Element.
5. Bekhet, H.A. i Eletter, S.F.K. (2014). ***Credit risk assessment model for Jordanian commercial banks: Neuralscoring approach***. Review of Development Finance 4 (2014) 20–28.
6. Berger, P.D. i Nasr, N.I. (1998). ***Customer Lifetime Value: Marketing Models and Applications.*** Journal of Inter- active Marketing, 12 (1), 17-30.
7. Berry, M.J.A. i Linoff, G. (2004). ***Data mining techniques: for marketing, sales and customer relationship managment.*** Second edition, Wiley Publishing.
8. Bishop, C.M. (2008). ***Pattern Recognition and Machine Learning***. Springer.
9. Blattberg, C.R., Kim, B.-D. i Nesslin, A.S. (2008). ***Database Marketing, Analyzing and Managing Customers***. Springer.
10. Bowerman, B.L., O'Connell, R.T. i Murphree, E.S. (2014). ***Business Statistics in Practice***. Seventh Edition. New York: McGraw-Hill.
11. Breiman,L. (2001). ***Statistical Modeling, The Two Cultures***. Statistical Science 2001, Vol. 16, No. 3, 199–231.
12. Brownlee, J. (2014). ***Machine Learning Foundations: Master the Definitions and Concepts***. Dostupno na http://machinelearningmastery.com/ [30. Rujan 2014.].
13. Bult, J.R. i Wansbeek, T.J. (1995). **Optimal selection for direct mail**, Marketing Science, Vol.14, No. 4, 378-394.
14. Campbell, D. i Frei, F. (2004). ***The persistence of customer profitability: Empirical evidence and implications from a financial service firm***. Journal of Service Research, 7(2), 107–123.
15. Chattamwelli, R. (2009). ***Data mining methods***. Alpha Science International ltd.
16. Cheng C.-H. i Chen Y.-S. (2009). ***Classifying the segmentation of customer value via RFM model and RS theory***. Expert Systems with Applications 36 (2009) 4176–4184.

17. Chiu, S. i Tavella, D. (2008). ***Data Mining and Market Intelligence for Optimal Marketing Returns***. First Edition, Elsevier.
18. Chuang, H. i Shen, C. (2008). ***A study on the applications of data mining techniques to enhance customer lifetime value – based on the department store industry***. Proceedings of the 7th International Conference on Machine Learning and Cybernetics, 168-173.
19. Cokins, G. (2001). ***Activity-based cost managment: an executives' guide***. John Wiley & sons, Inc.
20. Cooper, R. i Kaplan, R. (1991). ***Profit Priorities from Activity-Based Costing***. Harward Bussines Review, May-June 1991.
21. Davenport, T.H i Harris, J.G. (2007). ***Competing on analytics***. Harvard Business School Press.
22. Dietterich T.G. (2000). **Ensemble Methods in Machine Learning**. In: Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science, vol 1857. Springer, Berlin, Heidelberg
23. Donio, J., Massari, P. i Passante, G. (2006). ***Customer satisfaction and loyalty in a digital environment***. Journal of Consumer Marketing, 23/7 (2006) str.445-457.
24. Donkers, B., Verhoef, P.C. i de Jong, M. (2007). ***Modeling CLV: A test of competing models in the insurance industry***. Quantitative Marketing and Economics, 5(2), 163–190.
25. Drucker, P. (1954). ***The Practice of Managment,*** Harper Business - Harper Collins Publisher.
26. Duda, R.O. i P.E. Hart (1973). ***Pattern Classification and Scene Analysis***. Wiley.
27. Fader, P.S., Hardie, B.G. S i Lee, K. L. (2005). ***Counting your customers the easy way: an alternative to the Pareto/NBD model***. Marketing Science, 24(2), 275–284.
28. Farris, P.W, Bendle, N.T., Pfeifer, P.E. i Reibstein, D.J. (2006). ***Marketing metrics***. Wharton School Publishing.
29. Fausett, L. (1994). ***Fundamentals of Neural Networks***. New York: Prentice Hall.
30. Fodor, I.K. (2002). ***A survey of dimension reduction techniques***. Center for Applied Scientific Computing, Lawrence Livermore National Laboratory.
31. Fogel, L. J., Owens, A. J. i Walsh, M. J. (1966). ***Artificial Intelligence through Simulated Evolution***. John Wiley & Sons.
32. Foster, G., Gupta, M. i Sjoblom, L. (1996). ***Customer Profitability Analysis: Challenges and New Directions***, Cost Management Spring, 5–17.

33. Frawley, W.J., Piatetsky-Shapiro, G. i Matheus, C.J. (1992). ***Knowledge Discovery in Databases: An Overview***. AI Magazine Volume 13 Number 3.
34. Goldberg, D.E. (1989). ***Genetic algorithms in search, optimisation, and machine learning***. Addison Wesley.
35. Gupta , S. i Lehmann,D.R. (2003). ***Customer as asset***, Journal of Interactive Marketing . 2003, 17(1), 9-24.
36. Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., Sriram, S. (2006). **Modeling Customer Lifetime Value**. Journal of Service Research 9: 139
37. Guyon, I. (2007). ***Introduction to machine learning***, Bootcamp Villanva 2007. http://videolectures.net/bootcamp07_guyon_itml/ [15.4.2014.].
38. Han, J. i Kamber, M. (2001). ***Data Mining: Concepts and Techniques***. Academic Press, An Imprint of Elsevier
39. Hansen, L. i Salamon, P. (1990). ***Neural network ensembles***. IEEE Trans Pattern Analysis and Machine Intelligence, 12, 993-1001.
40. Haykin, S. (1999). ***Neural Networks: A Comprehensive Foundattions***, Second Edition, Prentice Hall.
41. Haykin S. (2009). ***Neural Networks and Learning Machines***. Third Edition, Prentice Hall.
42. Hastie, T., Tibshirani, R. i Friedman, R. (2008). ***The Elements of Statistical Learning: Data Mining, Inference, and Prediction***. Second Edition, Springer.
43. Hebb, D.O. (1949). ***The Organization of Behavior,a Neuropsychological Theory***. New York: John Wiley & Sons, Inc..
44. Howell, R.A. i Soucy, S.R. (1990). ***Customer profitability; as critical as product profitability***. Managment Accounting, rujan, 43-73.
45. Howson, C. (2008). ***Successful Business Inteligence: Secrets to Making BI a Killer App***, McGraw-Hill.
46. Hughes, A. M. (1994). ***Strategic database marketing***. Chicago: Probus Publishing Company.
47. Jackson, J.E. (1991). ***A User's Guide to Principal Components***. New York: John Wiley and Sons.
48. Johnson, R., Wichern, D. (2014). ***Applied Multivariate Statistical Analysis*****.** Sixth Edition, Pearson.

49. Joseph, A., Larrain, M. i Singh, E. (2011). ***Predictive Ability of the Interest Rate Spread Using Neural Networks***. Procedia Computer Science 6 (2011) 207–212.
50. Kaplan, R. (1989-1). ***Kanthal (A)***. HBS Case 9-190-002.
51. Kaplan, R. (1989-2). ***Kanthal (B)***. HBS Case 9-190-003.
52. Kaplan, R. i Norton, D.P. (1996). ***The balanced scorecard— translating strategy into action***. Harvard Business Scholl Publishing Corporation.
53. Kaplan,R. i Atkinson, A. (1998). ***Advanced management accounting***, Third edition, Prentice Hall.
54. Kaplan, R. i Cooper, R. (1998). ***Cost and Effects, Using Integrated Costs Systems to Drive Profitability and Performance***. Harward Bussines School Press.
55. Kaplan, R. i Narayanan, V.G. (2001). **Measuring and managing customer profitability**. Journal of Cost Management, 15(5): 5–15.
56. Kaplan, R., Norton, D.P. (2005). ***The balanced scorecard—measures that drive performance***. Harvard Bussines Review March 2005
57. Kaplan, R. i Anderson, S. (2007). ***Time-Driven Activity-Based Costing***, Harvard Business Scholl Publishing Corporation.
58. Kaplan, R., i Norton, D.P. (2007). ***Using the Balanced Scorecard as a Strategic Management System***. Harvard Business Review July-August 2007
59. Kim,S.-Y., Jung,T.-S., Suh, E.H i Hwang, H.-S. (2006). ***Customer segmentation and strategy development based on customer lifetime value***. A case study; Expert Systems with Applications 31 (2006) 101-107.
60. Klepac, G. i Mršić, L. (2006). ***Poslovna inteligencija kroz poslovne slučajeve***. Biblioteka Poslovna znanja, Lider press i TIM press.
61. Krakhmal, V. (2006). ***Customer profitability analysis in service industries***. BAA Annual Conference, 11-13 Apr 2006, Portsmouth, UK.
62. Krstić B. i Sekulić V. (2007). ***Upravljanje performansama preduzeća***. Ekonomski fakultet, Niš.
63. Kujamaki, I. (2007). ***Assesment of customer profitability in a multinational paper company***. Master's Thesis, Helsinki University of technology.
64. Kumar, A., Rao, V. R. i Soni, H. (1995). ***An empirical comparison of neural networks and logistic regression models***. Marketing Letters, 6(4), 251–263.
65. Kumar, V., Shah, D. i Venkatesan, R. (2006). ***Managing retailer profitability – one customer at a time***. Journal of Retailing 82 (4, 2006) 277-294.
66. Langley, P. (1992). ***Elements of Machine Learning***. Morgan Kaufman Publishers.

67. Lee, C., Rey, T., Mentele, J., i Garver, M. (2009). ***Structured Neural Network Techniques for Modeling Loyalty and Profitability***. Proceedings of SAS User Group International (SUGI 30), 2005.
68. Lee, J.H., Park, S.C. (2005). ***Intelligent profitable customers segmentation system based on business intelligence tools***. Expert Systems with Applications 29, 145-152.
69. Lee, T. H. i Young, S. (2000). ***Forecasting creditworthiness: Logistic vs. artificial neural net***. The Journal of Business Forecasting Methods and Systems, 18(4), 28–30.
70. Lejune, M.A.P.M. (2001). ***Measuring the impact of data mining on churn managment***. Internet Research: Electronic Networking Applications and Policy; Volume 11 Number 5 2001, 375-387.
71. Malthouse, E.C. i Blattberg, R.C. (2005). ***Can we predict customer lifetime value?*** Journal of Interactive Marketing, 19(1), 2–16.
72. Marsland, S (2009). ***Machine Learning, An Algorithmic Perspective***. Chapman and Hall.
73. McCulloch, W.S., Pitts, W. (1943). ***A Logical Calculus of the Ideas Immanent in Nervous Activity.*** Bulletin of mathematical biophysics volume 5, 1943, Chicago.
74. Mitchell, M., (1999). ***An Introduction to Genetic Algorithms***, A Bradford Book The MIT Press , 1999, Cambridge, Massachusetts, USA.
75. Mitchell, T.M. (1997). ***Machine Learning***. McGraw-Hill Science/Engineering/Math.
76. Mitchell, T. (2006). **The Discipline of Machine Learning**. http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf , [21.6.2012.].
77. Mittal, V., Sarkees, M., i Murshed, F. (2008). ***The right way to manage unprofitable customers***. Harvard Business Review, April 2008.
78. Mulhern, F.J. (1999). ***Customer Profitability Analysis: Measurement, Concentration, and Research Directions.*** Journal of Interactive Marketing, 13 (1), 25-40.
79. Ngai, E.W.T (2005). ***Customer relastionship management research (1992-2002): An academic literature review and classification***. Marketing Intelligence, Planning 23 , 582-605.
80. Ngai, E.W.T, Xiu, L. i Chau, D.C.K (2009). ***Application of data mining techniques in customer relationship management: A literature review and classification***. Expert Systems with Application 36 (2009) 2592-2602.
81. Niraj, R., Gupta,M. i Narasimhan, C. (2001). ***Customer profitability in a supply chain***. Journal of Marketing 65(July) 1–16.

82. Paliwal, M. i Kumar, U. (2009). ***Neural networks and statistical techniques: a review of applications***. Expert System Applications 36 (1), 2–17.
83. Panian, Ž. i Klepac, G. (2003). ***Poslovna inteligencija***. Zagreb: Masmedia.
84. Panian, Ž. (2007). ***Poslovna inteligencija, studije slučajeva iz hrvatske prakse***. Zagreb: Narodne Novine.
85. Peck, H. (1997). ***Towards a framework of relationship: a research methodology***. Bedfordshire: Cranfield University.
86. Pourshahriar, H. (2012). ***Correct vs. accurate prediction: A comparison between prediction power of artificial neural networks and logistic regression in psychological researches***. Procedia - Social and Behavioral Sciences 32 (2012) 97 – 103.
87. Principe, J.C., Euliano, N.R. i Lefebvre, W.C. (2010). ***Neural and adaptive systems, Fundamentals Through Simulations***. John Wiley & Sons.
88. Provost, F., Fawcett, T. (2013). ***Data Science for Business***. O'Reilly Media, First Edition.
89. van Raaij, E.M., Vernooij, M. i Van Triest, S. (2003). ***The implementation of customer profitability analysis: A case study***. Industrial Marketing Management, Vol. 32(7), 573-583.
90. Razi , M. A. i Athappilly, K. (2005). ***A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models***. Expert Systems with Applications 29 (2005) 65–74.
91. Rechenberg, I. (1965). ***Cybernetic Solution Path of an Experimental Problem***. Royal Aircraft Establishment, Library Translation 1122, Farnborough.
92. Reinartz,W.J., Kumar, V. (2003). ***The impact of customer relationship characteristics on profitable lifetime duration***. Journal of Marketing, 67, 77–99.
93. Rencher, C.A. i Christensen, F.W. (2012). ***Methods of Multivariate Analysis***. Third Edition, Wiley & Sons.
94. Rochester, N., Holland, J.H., Haibt, L.H. i Duda, W.L. (1956). ***Tests on a cell assembly theory of the action of the brain, using a large digital computer.*** IRE Transactions on Information Theory, vol. IT-2,80-93.
95. Rosenblatt, F. (1958). ***The Perceptron: A probabilistic model for informational storage and organization in the brain***. Psihological Review, vol.65,: 386-408.
96. Rosenblatt, F. (1962). **Principles of neurodynamics; perceptrons and the theory of brain mechanisms**. Washington: Spartan Books.

97. Ross, E.B. (1984). ***Making Money with Proactive Pricing***. Harvard Bussines Review November–December 1984, p. 145.
98. Rozga, A. (2010). **Multivarijacijska statistička analiza**. Ekonomski fakultet Split-Mostar, autorizirana predavanja, travanj 2010.
99. Rust, R.T., Lemon, K.N. i Zeithaml, V.A. (2004). ***Return on marketing: Using customer equity to focus marketing strategy***. Journal of Marketing, 68(1), 109–127.
100. Rust, R.T. , Kumar, V. i Venkatesan, R. (2011). ***Will the frog change into a prince? Predicting future customer profitability.*** International Journal of Research in Marketing 28 (2011) 281–294.
101. Ryals, L. (2003). ***Making customers pay: measuring and managing customer risk and returns***. Journal of Strategic Marketing, 11: 165–175.
102. Shapiro,B., Kasturi, R., Moriarty , R. i Ross, E. (1987). ***Manage Customers for Profits (Not Just Sales)***. Harvard Bussines Review, September-October 1987.
103. Shih, Y.Y. i Liu, C.Y. (2003). **A method for customer lifetime value ranking – combining the analytic hierarchy process and clustering analysis**, Journal of Database Marketing & Customer Strategy Management, Vol. 11 No. 2, 159.
104. Sonnenburg, S., Ratsch, G., Henschel, S., Widmer, C., Behr, J., Zien, A., de Bona, F., Gehl, C., Binder, A. i Franc, V. (2010). ***The SHOGUN Machine Learning Toolbox***. Journal of Machine Learning Research 11 pp. 1799-1802.
105. Stone, B. (1995). ***Successful direct marketing methods***. Lincolnwood, IL: NTC Business Books. pp. 37–57.
106. Storbacka, K. (1997). ***Segmentation based on customer profitability-retrospective analysis of retail bank customer bases***. Journal of Marketing Management 13, 479-492.
107. Turban, E., Aronson,J. i Liang T.-P. (2005). ***Decision support systems and inteligent systems***. Pearson Prentice Hall.
108. van Triest, S. (2005). ***Customer size and customer profitability in non-contractual relationships***. Journal of Business and Industrial Marketing, Volume 20 Issue:3. 148-155.
109. Vapnik, V.N. (1992). ***Principles of risk minimization for learning theory***. Dostupno na: http://papers.nips.cc/paper/506-principles-of-risk-minimization-for-learning-theory.pdf [03.09.2015.].
110. Vapnik, V.N. (1995). ***The Nature of Statistical Learning Theory***. Springer.

111. Vapnik, V.N. (1998). ***Statistical Learning Theory***. Wiley.

112. Vercellis, K. (2009). ***Business Intelligence: Data Mining and Optimization for Decision Making***. John Wiley and Sons.

113. Wang, H.-F., Hong, W.-K. (2006). ***Managing customer profitability in a competitive market by continuous data mining***. Industrial Marekting Management 35 (2006) 715-723.

114. Widrow, B. i Hoff., M.E. (1960). ***Adaptive switching circuits***. New York Institute of Radio Engineers, IRE WESCON convention record

115. Zhang, W., Cao, Q. i Schniederjans, M. J. (2004). ***Neural networks earning per share forecasting models: A comparative analysis of alternative methods***. Decision Sciences, 35(2), 205–237.

TABLICE

Tablica 2.1 Izračun parametara rada genetičkog algoritma 49

Tablica 4.1 Prosječna struktura troškova iz grupe artikala s velikim udjelom u ukupnom prometu 95

Tablica 4.2 Broj dobavljača po kontinentima u 2013.godini 96

Tablica 4.3 Broj kupaca po regijama u 2013.godini 97

Tablica 4.4 Struktura kupaca u 2013.godini 97

Tablica 4.5 Uvoz u BiH 2014.godine.....98

Tablica 4.6 Izvoz iz BiH 2014.godine 100

Tablica 4.7 Države partneri s najvećim prometom u uvozu i izvozu za 2014.godinu 101

Tablica 4.8 Države partneri s najvećim prometom u uvozu i izvozu za 2014.godinu 102

Tablica 5.1 Nezavisne varijable grupe troškova 118

Tablica 5.2 Nezavisne varijable grupe cijena 119

Tablica 5.3 Nezavisne varijable grupe značajki kupaca 120

Tablica 5.4 Zavisne varijable modela 121

Tablica 5.5 Definicija setova podataka za neuronsku mrežu.....127

Tablica 5.6 Rezultat usporedbe različitih arhitektura mreže u predviđanju R varijable 130

Tablica 5.7 Rezultat usporedbe različitih arhitektura mreže u predviđanju F varijable 132

Tablica 5.8 Rezultat usporedbe različitih arhitektura mreže u predviđanju M varijable 134

Tablica 5.9 Usporedni rezultati učenja mreže ponavljanjem testiranja u više ciklusa 136

Tablica 5.10 Usporedni rezultati učenja mreže varijacijom parametara.....137

Tablica 5.11 Usporedni rezultati učenja mreže varijacijom parametara.....138

Tablica 5.12 Usporedni rezultati učenja mreže upotrebom genetičkog algoritma.....140

Tablica 6.1 Varijable troškova 147

Tablica 6.2 Varijable značajki kupca 148

Tablica 6.3 Deskriptivna statistika varijabli VK_3, VK_4, KK_3 i PP_1 149

Tablica 6.4 Varijable cijenovne politike 150

Tablica 6.5 Primjena metoda po hipotezama istraživanja.....152

Tablica 6.6 Broj kupaca po godinama.....153
Tablica 6.7 Količina prodaje u kg.....153
Tablica 6.8 Broj narudžbi po godinama.....154
Tablica 6.9 Prosječna količina po isporuci u kg.....154
Tablica 6.10 Prosječna ukupna naručena količina robe po kupcu u KG.....155
Tablica 6.11 Prosječni broj narudžbi po kupcu.....155
Tablica 6.12 Broj različitih lokacija isporuke po kupcu.....155
Tablica 6.13 Prosječan broj različitih artikala naručenih od pojedinog kupca.....156
Tablica 6.14 Prosječan promet ostvaren po narudžbi za pojedinog kupca.....156
Tablica 6.15 Struktura prometa po godinama.....156
Tablica 6.16 Kategorizacija kupaca prema veličini prihoda.....157
Tablica 6.17 Portfolio kupaca po veličini prihoda.....158
Tablica 6.18 Prvih 12 kupaca iz grupe „Velikih kupaca“.....158
Tablica 6.19 Deskriptivna statistika normaliziranih podataka.....159
Tablica 6.20 Rezultati Logit regresijskog modela u predviđanju R varijable.....161
Tablica 6.21 Rezultati Logit regresijskog modela u predviđanju F varijable.....163
Tablica 6.22 Rezultati modela višestruke linearne regresije u predviđanju M varijable 164
Tablica 6.23 Rezultati najboljih modela neuronske mreže.....165
Tablica 6.24 Rezultati optimiziranih modela neuronske mreže upotrebom genetičkog algoritma.....166
Tablica 6.25 Utjecaj troškova na profitabilnost.....166
Tablica 6.26 Utjecaj cjenovne politike na profitabilnost.....167
Tablica 6.27 Utjecaj veličine kupca na profitabilnost.....168
Tablica 6.28 Utjecaj kompleksnosti kupca na profitabilnost.....169
Tablica 6.29 Utjecaj ponašanja pri plaćanju obveza kupca na profitabilnost.....170
Tablica 6.30 Usporedba rezultata metoda NN i multivarijantne statističke analize..171
Tablica 6.31 Rezultati metoda multivarijantne statističke analize u dokazivanju 2. i 3. hipoteze istraživanja.....172
Tablica 6.32 Značajnost nezavisnih varijabli u regresijskim modelima pri dokazivanju 2. i 3. hipoteze istraživanja.....173

SLIKE

Slika 2.1 Struktura biološkog neurona 25
Slika 2.2 Umjetni neuron 27
Slika 2.3 Grafovi osnovnih aktivacijskih funkcija 28
Slika 2.4 Jednoslojna mreža bez povratnih veza 29
Slika 2.5 Višeslojna mreža bez povratnih veza 30
Slika 2.6 Primjer mreže s povratnom vezom 31
Slika 2.7 Blok dijagram učenja „s učiteljem“ 32
Slika 2.8 Blok dijagram podržanog učenja 33
Slika 2.9 Blok dijagram nenadziranog učenja 34
Slika 2.10 Blok dijagram višeslojne mreža bez povratne veze s jednim neuronom u izlaznom sloju 35
Slika 2.11 Jednostavna arhitektura mreže kompetitivnog algoritma učenja 39
Slika 4.1 Tipična distribucija narudžbi u odnosu na cijenu proizvoda 71
Slika 4.2 Odnos cijena i ukupnih troškova 74
Slika 4.3 Promjene u troškovnoj strukturi 75
Slika 4.4 Informacijski tijek u analizi profitabilnosti kupca 76
Slika 4.5 Kanthal krivulja 78
Slika 4.6 Stobachoff krivulja 79
Slika 4.7 Portfolio analiza baze kupaca 81
Slika 4.8 Tijek podataka i roba u lancu snabdijevanja 92
Slika 4.9 Struktura troškova gotovog proizvoda 95
Slika 5.1 Model istraživanja 104
Slika 5.2 Proces odabira najbolje konfiguracije mreže 128
Slika 5.3 Arhitektura mreže za predviđanje R varijable 131
Slika 5.4 Arhitektura mreže za predviđanje R varijable 133
Slika 5.5 Arhitektura mreže za predviđanje M varijable 135
Slika 6.1 Kategorizacija kupaca prema neto prihodu 157

ŽIVOTOPIS

Osobni podaci

Ime i prezime:	Ivica Ćorić
Ime oca:	Ante
Nacionalnost:	Hrvat
Datum rođenja:	29. siječanj 1968. godine
Mjesto rođenja:	Mostar, Bosna i Hercegovina
Adresa stanovanja:	ak.Ivana Zovke 9, 88000 Mostar, BiH
Državljanstvo:	Hrvatsko i BiH
Maternji jezik:	Hrvatski
Bračno stanje:	oženjen, 3 djece
Obrazovanje:	diplomirani ekonomist
Kontakt telefon:	063 318 424
Mail:	ivica.coric@hera.ba

Obrazovanje

- ... - 2009 Poslijediplomski doktorski studij Ekonomskog fakulteta Sveučilišta u Mostaru i Ekonomskog fakulteta Sveučilišta u Splitu.
- 1993. - 1988. Ekonomski fakultet, Mostar, Bosna i Hercegovina
- 1986. - 1982. *Elektrotehnička škola*, Mostar, Bosna i Hercegovina

Tijek stalnih uposlenja:

- ... – 01.03.2001. Direktor i suvlasnik u poduzeću HERA doo Mostar, BiH
- 28.02.2001 – 16.12.1995. Sistem Analitičar, Optima-OR Mostar, BiH
- 15.12.1995. - 01.05.1995. Sistem Analitičar, Hrvatska Poštanska Banka, Mostar, BiH
- 30.04.1995. - 01.04.1994. Viši savjetnik za izgradnju IS, Ministarstvo Financija HZ HB, Porezna uprava Središnji ured, Mostar, BiH
- 31.03.1994. - 01.06.1993. Programer, Elektroprivreda HZ HB, Mostar,BiH

Vještine

- *Strani jezici:* engleski jezik (tečno u govoru i pismu), francuski jezik (osnove).
- *Tehničke kvalifikacije:*

- Baze podataka: Oracle 8i – 12c, MySQL, MS SQL, MongoDB
- Softwarski razvojni alati: Oracle Designer, Oracle Developer Suite, Oracle Apex
- Programski jezici: PL/SQL, Java, JavaScript
- Operacijski sustavi: Windows OS's, Unix, Linux
- Softwarski paketi i platforme: MS Office, NeuroSolutions, Oracle BI Server EE

- *Ostale sposobnosti:* sposobnost rada u timu i razvijene komunikacijske vještine

POPIS OBJAVLJENIH RADOVA

1. Markić, B., Gašpar, D. i Ćorić, I. (2011). ***Neural Networks in Customer Profitability Analysis***. International Conference „Economic Theory and Practice: Meeting the New Challenges“, Mostar
2. Ćorić, I. (2011). ***Model predviđanja profitabilnosti kupaca***. Zbornik radova Ekonomskog fakulteta u Mostaru - broj XVII, Mostar
3. Markić, B., Gašpar, D. i Ćorić, I. (2012). ***Machine Learning in Customer Profitability Forecasting***. 16th International Research/Expert Conference ”Trends in the Development of Machinery and Associated Technology”, Dubai, 2012
4. Gašpar, D., Ćorić, I. i Mabić, M. (2014). ***Data mining in customer profitability analysis***. Proceedings of International Conference for Entrepreneurship, Innovation and Regional Development ICEIRD 2014, Nicosia
5. Ćorić, I., Gašpar, D. i Mabić, M. (2014). ***Business Inteligence and Production Systems***. Proceedings of 4th International Conference "Vallis Aurea: focus on Regional and Innovation Development " in Požega, Croatia
6. Gašpar, D., Mabić, M. i Ćorić, I. (2015). ***Risk Managment Software – Standpoints of Users***. Proceedings of International Conference "ENTRENOVA 2015" in Kotor, Montenegro.
7. Gašpar, D., Mabić, M. i Ćorić, I. (2015). ***Data Mining and Predicting Student Performance***. Proceedings of 19th International Research/Expert Conference "Trends in the Development of Machinery and Associated Technology“ TMT 2015 in Barcelona, Spain.
8. Gašpar, D., Ćorić, I. i Mabić, M. (2015). ***Software Eases Burden of Risk Managment Process***. Proceedings of 16th EBES Conference - Istanbul, At Istanbul, Turkey.
9. Ćorić, I. (2016). **Comparison of Multivariate Statistical Analysis and Machine Learning Methods in Retailing: Research Framework Proposition**. Proceedings of International Conference "ENTRENOVA 2016" in Rovinj, Croatia.