

Broj 01/ 275
Podgorica, 06.02.2020. godine

UNIVERZITET CRNE GORE
-Odboru za doktorske studije i Senatu-

Predmet: Materijal za sjednicu Odbora i Senata

Poštovani,

U skladu sa članom 35. Pravila doktorskih studija, dostavljamo Vam materijal za narednu sjednicu Odbora za doktorske studije, odnosno Senata Univerziteta Crne Gore i to:

-Obrazac D1 (Ocjena podobnosti doktorske teze i kandidata) sa propratnom dokumentacijom za kandidata mr Sunčicu Rogić.



DEKAN
Prof. dr Nikola Milović

**UNIVERZITET CRNE GORE
EKONOMSKI FAKULTET PODGORICA
DOKTORSKE STUDIJE**

Br. 01/ 28

Podgorica, 06.02.2020.god.

Na osnovu čl. 64. Statuta Univerziteta Crne Gore, člana 35. Pravila doktorskih studija, Vijeće Ekonomskog fakulteta je na sjednici održanoj 06.02.2020.godine donijelo

ODLUKU

1. Usvaja se Izvještaj Komisije za ocjenu podobnosti doktorske teze i kandidata **mr Sunčice Rogić**.
2. Prihvata se kao podobna za izradu doktorska teza pod nazivom „**Prediktivni modeli odlučivanja u direktnom marketingu bazirani na Support Vector Machine metodu**“ i kandidat **mr Sunčica Rogić**.
3. Predlog dostaviti Odboru za doktorske studije i Senatu Univerziteta Crne Gore koji su nadležni za donošenje konačne Odluke.

OBRAZLOŽENJE

Vijeće Ekonomskog fakulteta je na sjednici održanoj 06.02.2020.godine razmatralo Izvještaj Komisije za ocjenu podobnosti doktorske teze „**Prediktivni modeli odlučivanja u direktnom marketingu bazirani na Support Vector Machine metodu**“ i kandidata **mr Sunčice Rogić**.

Na osnovu izloženog odlučeno je kao u dispozitivu.



DEKAN

[Signature]
Prof.dr Nikola Milović

DOSTAVLJENO:

- a/a
- referentu doktorskih studija,
- Centru za doktorske studije,
- Senatu UCG.

[Handwritten mark]

OCJENA PODOBNOSTI DOKTORSKE TEZE I KANDIDATA

OPŠTI PODACI O DOKTORANDU	
Titula, ime i prezime	Mr Sunčica Rogić
Fakultet	Ekonomski fakultet, UCG
Studijski program	Doktorske studije Ekonomije
Broj indeksa	2/18
Podaci o magistarskom radu	Naziv: „Uticaj sponzorstva na percepciju brenda“ Naučna oblast: Marketing Institucija na kojoj su završene magistarske studije: Ekonomski fakultet, UCG Godina završetka: 2018 Srednja ocjena: 10.00
NASLOV PREDLOŽENE TEME	
Na službenom jeziku	Prediktivni modeli odlučivanja u direktnom marketingu bazirani na <i>Support Vector Machine</i> metodu
Na engleskom jeziku	Predictive decision support models in direct marketing based on Support Vector Machine method
Datum prihvatanja teme i kandidata na sjednici Vijeća organizacione jedinice	06.02.2020.
Naučna oblast doktorske disertacije	Poslovna inteligencija u marketingu – prediktivna analitika poslovnih podataka u marketingu
Za navedenu oblast matični su sljedeći fakulteti	
Ekonomski fakultet	
A. IZVJEŠTAJ SA JAVNE ODBRANE POLAZNIH ISTRAŽIVANJA DOKTORSKE DISERTACIJE	
Javna prezentacija i odbrana polaznih istraživanja organizovana je 4. decembra 2019. godine na Ekonomskom fakultetu Univerziteta Crne Gore, u Podgorici. Doktorand je u uvodnom izlaganju predstavio motive za odabir teme, obrazloženje teme, ciljeve istraživanja i formulisane hipoteze, materijal i metode koji će biti korišteni u istraživanju, kao i očekivani naučni doprinos. Pored toga, predstavljeni su dosadašnji rezultati istraživanja i osnovni zaključci rada „<i>Customer value prediction in direct marketing using hybrid Support Vector Machine Rule Extraction method</i>“, koji je predstavljen na <i>European Conference on Advances in Databases and Information Systems (ADBIS 2019)</i> na Bledu. Nakon uvodnog izlaganja kandidata, članovi komisije su dali sugestije za unaprijeđenje strukture prijave rada, kao i izmjenu naslova i potencijalno kvantifikovanje hipoteza. U toku diskusije, članovi Komisije su postavili pitanja kandidatu, nakon čega je Komisija zaključila da je kandidat uspješno odbranio polazna istraživanja doktorske disertacije.	
B. OCJENA PODOBNOSTI TEME DOKTORSKE DISERTACIJE	

B1. Obrazloženje teme

Organizacije danas posluju u uslovima jake konkurencije i u nestabilnom poslovnom okruženju gdje se ponašanje kupaca stalno mijenja i teško ga je predvidjeti. Na tako konkurentnom i zahtjevnom tržištu direktni marketing je postao neophodna komponenta poslovanja koja obezbjeđuje efikasniju promociju proizvoda i razvijanje strateški jačih odnosa sa klijentima (Chen, Chiu & Chang 2005). Cilj metoda direktnog marketinga je povećanje odgovora kupaca na kampanju što podrazumijeva stvaranje efikasne interakcije sa kupcem. Za razliku od opšteg marketinga, čija je strategija orijentisana na proizvode, kod direktnog marketinga fokus je na kupcu. Da bi se razumjele potrebe potrošača i predvidjelo kupovno ponašanje koriste se podaci o njima koji se čuvaju u bazama podataka (data-driven pristup). Ovo se postiže identifikovanjem karakteristika kupaca koje omogućavaju da se predvidi vjerovatnoća odgovora na marketinšku kampanju, tržišna vrijednost kupca i analiziraju njihove potrebe (Bose, Chen 2009).

Direktni marketing je sve češće primjenjivana strategija u različitim oblastima poslovanja. U razvijenim zemljama evidentan je stalni rast troškova direktnog marketinga, budžeta koji se opredjeljuje za njega i broja zaposlenih u ovoj oblasti a samim tim i procentualno učešće ove poslovne aktivnosti u ukupnom bruto domaćem proizvodu (Direct Marketing Association Statistical Fact Book-DMA 2018). Zbog sve većeg ekonomskog značaja raste popularnost ove teme u akademskim i praktičnim istraživanjima.

Uporedo sa rapidnim rastom upotrebe Interneta dolazi do porasta elektronske tj. on-line trgovine. Zahvaljujući internet tehnologijama povećava se agilnost, selektivnost i interakcija u odnosima sa kupcima. Sa druge strane, zbog velikog broja sajtova za online trgovinu kupci imaju veće mogućnosti izbora. Usled efikasnosti internet tehnologija i mogućnosti izuzetno brze pretrage proizvoda, online tržište postaje izuzetno dinamično. U takvim uslovima konkurencija za internet preduzeća postaje sve jača što dovodi do značajnih promjena njihovog stava prema upravljanju odnosa sa klijentima. Osim toga, broj kupaca se u uslovima online trgovine značajno povećava (ne uzimaju se u obzir samo registrovani kupci već i nepoznati-neregistrovani posjetioći sajtova koji obavljaju transakcije). Takođe, dolazi do uvećanja broja karakteristika kupaca koje se evidentiraju u bazama podataka. Naime, klasične karakteristike kupaca kao što su socio-demografske i transakcione dopunjuju se karakteristikama online ponašanja kao što su broj posjeta sajtu, broj klikova, ključne riječi za pretragu, vrijeme provedeno na strani nekog proizvoda i slično (Kim & Yum, 2011).

U tom kontekstu metode direktnog marketinga postaju sve više inspirativne za naučna istraživanja i u novije vrijeme se poklapaju sa istraživanjima u oblasti *big data* tj. *data mining* metoda (Olson, 2007; Chen et al., 2015; Olson & Chae, 2012). Analitika velikih podataka, kvantitativne metode za podršku odlučivanju i moćni algoritmi za ekstrakciju obrazaca ponašanja potrošača u oblasti marketinga postaju predmet velikog interesovanja ne samo prakse već i naučne zajednice (Lessmann, Coussement, De Bock & Haupt 2018). Implementacija markentiške *big data* analitike dovodi do značajnog povećanja rentabilnosti i produktivnosti firme (Germann et al. 2013; Tambe 2014).

Uzimajući u obzir aktuelnost problematike istaknutu u prethodnoj literaturi, u ovom radu biće predloženi novi, efikasniji modeli odlučivanja u direktnom marketingu bazirani na data mining metodama, u prvom redu na Support Vector Machine (SVM) metodu. Modeli će biti testirani na podacima iz on-line direktnog marketinga (gdje se ponuda upućuje preko društvenih mreža), što ne umanjuje opštost definisanih modela i njihovu primjenu u drugim oblicima direktnog marketinga (kada je ponuda poslata poštom, e-mailom ili na neki drugi način), jer će se modeli razlikovati samo dijelom u skupu prediktora, dok predložena metodologija ostaje ista. U daljem tekstu su navedeni problemi kod već postojećih modela u literaturi i objašnjeno na koji način predloženi modeli povećavaju efikasnost procesa odlučivanja.

Glavni ciljevi direktnog marketinga su identifikovanje tj. targetiranje kupaca iz postojeće baze koji će najvjerovatnije odgovoriti na određenu markentišku kampanju, predviđanje nepoznatih kupaca koji bi mogli odgovoriti na kampanju, kao i analiza potreba i ponašanja kupaca u cilju jačanja odnosa sa njima i njihove lojalnosti (Z.-Y.Chen et al., 2014).

Metode za targetiranje kupaca mogu se podijeliti u dvije grupe: segmentacione i *scoring* metode (Jonker et al. 2004; Kaymak 2001).

Segmentacione metode dijele kupce u grupe (segmente) koristeći odgovarajuće eksplanatorne varijable tako da pripadnici segmenta budu što je moguće više homogeni po pitanju očekivanog odgovora na direktnu kampanju. Ponuda se šalje kupcima iz segmenata koji imaju najveću vjerovatnoću odgovora. Kupci se najčešće segmentiraju po kodiranim atributima njihovog kupovnog ponašanja kao što su Recency (datum poslednje transakcije), Frequency (broj obavljenih transakcija) i Monetary (ukupan iznos od transakcija) i ta metoda je poznata kao RFM segmentacija kupaca (Hughes, 1994; Hughes, 2005; McCarty & Hastak, 2007). Ove tehnike uglavnom zahtijevaju subjektivnu procjenu načina kodiranja RFM atributa i optimalnog broja segmenata. Iz tog razloga *data mining* metode kao što su *k-means* klasterizacija ili klasterizacija pomoću neuronskih mreža, mogu dati objektivnije rezultate kod RFM segmentacije kupaca (Cheng & Chen, 2009; Hosseini, Maleki & Gholamian, 2010; Sarvari, Ustundag & Takci, 2016; Khalili-Damghani, Abdi & Abolmakarem, 2018). **Modeli prediktivne segmentacije** dodatno omogućavaju predviđanje pripadnosti nepoznatog kupca određenom segmentu na osnovu njihovih karakteristika (Cheng & Chen, 2009; Hosseini, Maleki & Gholamian, 2010; Sarvari, Ustundag & Takci, 2016; Khalili-Damghani, Abdi & Abolmakarem, 2018).

Kod *scoring* pristupa svakom pojedinačnom kupcu se dodjeljuje odgovarajući *scor* koji se dobija na osnovu predviđene vjerovatnoće odgovora na ponudu i/ili predviđenog profita koji će biti ostvaren. Kupci se zatim uređuju na osnovu *scora* i oni koji imaju najveći *scor* se targetiraju za slanje ponude.

Većina istraživanja kod *scoring* pristupa uzima u obzir samo predikciju vjerovatnoće odgovora kupca na kampanju (tzv. **customer response modeli**). S obzirom da je cilj prepoznati kupce kao moguće respondente ili ne-respondente na odgovarajuću markentišku kampanju *customer response* modeliranje spada u binarni klasifikacioni problem. Kao eksplanatorne varijable se najčešće

koriste karakteristike kupaca i podaci o kupovnom ponašanju tj. transakcijama. Zavisna varijabla je binarna varijabla koja sadrži podatak da li je kupac u prošlosti odgovorio na kampanju ili ne. Response modeli na osnovu ovih podataka predviđaju vjerovatnoću odgovora na buduću kampanju. Za ovo predviđanje najčešće se primjenjuju *data mining* metodi logističke regresije, vještačke neuronske mreže (ANN) i drvo odlučivanja (DT) (Bose & Chen 2009; Han, Lu & Leung, 2012; Guido, Prete, Miraglia, De Mare 2011; Coussement, Van den Bossche & De Bock, 2014; Kang, Cho & MacLachlan, 2012).

Međutim, visoka stopa odgovora ne znači nužno i visok profit (Kim, Lee & Cho, 2008). Kod *scoring* pristupa je veoma značajno uključiti i predikciju profitabilnosti respondenata (kupaca koji će vjerovatno odgovoriti na kampanju) (tzv. **profit-maximization modeli**) kako bi se identifikovale kategorije visoko profitabilnih i nisko profitabilnih respondenata (Cui, Wong & Wan, 2015; Otter, Van der Scheer & Wansbeek, 2006).

Glavni problem modela prediktivne segmentacije je nebalansiranost klasa (*class imbalance*) jer je segment najvrednijih kupaca po pravilu najmanji a najvažnije je predvidjeti kupce koji pripadaju tom segmentu. Slično, kod *customer response modela*, stopa odgovora je obično mala u odnosu na cjelokupni skup podataka o kupcima (obično manja od 5%), pa se javlja problem minorne klase. Ovaj problem kod većine prediktivnih klasifikacionih metoda dovodi do pristrasnosti prema malim klasama i najčešće do njihove pogrešne klasifikacije (*misclassification*) (Kim, Chae & Olson, 2012; Miguéis, Camanho & Borges, 2017).

Različiti su pristupi u rješavanju ovog problema u literaturi. Najviše se primjenjuju metodi *under-sampling* (uzimanje podskupova veće klase najčešće slučajnim - *random* izborom) ili *over-sampling* (generisanje sintetičkih primjera kao dopuna manjoj klasi). Glavni nedostatak *under-sampling* metode je što se se iz analize isključuje ogroman broj podataka o nerespondentima pa uzeti podskup moguće ne reprezentuje dobro čitavu klasu, dok *over-sampling* sintetički generiše respondente koji mogu biti nerealni i ne odgovarati njihovim stvarnim karakteristikama.

Međutim, u literaturi je potvrđeno da u slučaju nebalansiranih i linearno neseparabilnih klasa najbolje klasifikacione performanse ima SVM metod i da se može koristiti kao predprocesor koji prečišćava podatke (separira klase) za druge klasifikatore (Farquard & Bose, 2012; Martens et al., 2006; Diederich, 2008; Barakat and Bradley, 2010; Martens et al., 2008; Kaščelan et al., 2014). Farquard & Bose (2012) su potvrdili da SVM može uspješno ukloniti preklapanje klasa i dopuniti malu klasu novim relevantnim primjerima. Oni su takođe utvrdili da je ovaj metod efikasniji u balansiranju klasa nego *under-sampling* i *over-sampling* metode, kao i da primjena standardnih klasifikatora na SVM outputu daje najbolje rezultate. Međutim, SVM klasifikator je „crna kutija“ tj. ne generiše model koji se može interpretirati. Taj nedostatak se može riješiti hibridnim pristupom, gdje se za ekstrakciju pravila SVM klasifikacije koristi DT metod (Barakat and Bradley, 2010; Kaščelan et al., 2014).

Glavni problem kod modela za predikciju profitabilnosti je broj visoko-profitabilnih respondenata koji je značajno manji od ostalih. Usled ovog problema javlja se asimetričnost (zakrivljenost) distribucije zavisne varijable koja može uzrokovati nekonzistentnost standardnih

regresionih metoda (linearnih ili polinomnih). Postoji veliki broj istraživanja koji za predikciju profitabilnosti koriste neku vrstu parametarskog regresionog metoda, kao što su generalizovani linearni modeli, *ridge* ili *quantile* regresija, koje su u stanju da se izbore sa asimetričnom zavisnom varijablom (Malthouse, 1999; Zhang, 2009). Međutim, preduslovi koje zahtijevaju ovi modeli često nisu zadovoljeni (multikolinearnost regresora, u prvom redu RFM atributa) a prisutan je i problem specifikacije funkcionalne forme, posebno za interaktivne članove kojih može biti veliki broj za veći broj regresora. U literaturi je potvrđeno da je SVM regresija robustan metod za modeliranje nelinearnih zavisnosti sa asimetričnom distribucijom zavisne varijable (Christmann 2004; Basak et al., 2007).

Imajući u vidu sve njegove navedene prednosti, u ovom radu je za prevazilaženje problema minorne klase, kod modela prediktivne segmentacije i *customer response* modela, predložen SVM metod. Primjenom ovog metoda kao predprocesora podataka, povećava se broj tačno targetiranih najvrijednijih kupaca i respondenata u postojećoj bazi, što dovodi do smanjenja nepotrebnih troškova u kampanji i u krajnjem do većih mogućnosti za ostvarenje profita od kampanje. Hibridnim SVM-Rule Extraction (SVM-RF) metodom kojim se na osnovu SVM izlaza generiše DT model, dobijaju se eksplicitna pravila o kupcima, ponudi koju preferiraju i kupovnom ponašanju. Na osnovu ovih pravila je moguće uspostaviti efikasniju interakciju sa kupcem, prilagoditi mu ponudu i vjerovatnije zadržati njegovu lojalnost. Na ovaj način predloženi modeli odlučivanja ne podržavaju samo selekciju kupaca za kampanju već i jačanje odnosa sa njima.

Na osnovu njegovih, u literaturi prepoznatih mogućnosti, za prevazilaženje problema nelinearnih zavisnosti i asimetričnosti distribucije profita kod modela za predikciju profitabilnosti respondenata, predložen je SVM regresioni metod. Pomoću SVM regresionog modela mogu se efikasno selektovati najprofitabilniji respondenti u postojećoj bazi i predvidjeti očekivana profitabilnost za potencijalne respondente (nepoznate kupce) na osnovu njihovih karakteristika. Tako dolazi do značajnih ušteda zbog neracionalnog slanja ponuda i povećanja prihoda ako postojeći ili potencijalni visoko profitabilni kupci odgovore. Senzitivnom analizom se mogu utvrditi nelinearne zavisnosti prediktora i profitabilnosti tj. opisati profil najprofitabilnijih kupaca, koji mogu, iako ih je malo, činiti najveći dio prihoda od kampanje. Dakle, SVM bazirani *profit-maximization* model omogućava izbor specifičnih aktivnosti prilagođenih profilu najprofitabilnijih respondenata i time jača strateške odnose sa najvažnijim segmentom kupaca.

Predloženi modeli povećaju efikasnost procesa odlučivanja u direktnom marketingu, usled čega se očekuje profitabilnija kampanja, bolje interakcija sa najvažnijim kupcima, lakše kreiranje specifične ponude i veće zadržavanje kupaca.

B2. Cilj i hipoteze

Procesi odlučivanja u direktnom marketingu zahtijevaju efikasne modele za selekciju i profilisanje kupaca, kako bi se maksimizirao ostvareni profit od kampanje i poboljšali odnosi sa kupcima.

Kvalitet podataka sa kojima marketari raspolazu u bazi i izbor efikasnog metoda su ključni za uspješne modele odlučivanja. S obzirom na veliki broj podataka u markentiškim bazama sve više

se koriste *data mining* metode za realizaciju efikasnih modela. Ovo posebno važi za on-line kampanje za koje je broj podataka još veći, jer se u bazi evidentiraju svi on-line pristupi korisnika sa pripadajućim pokazateljima kao što su broj klikova, vrijeme provedeno na sajtu i slično. Svi ti podaci su jako važni kod targetiranja on-line kupaca jer govore o jednom novom modalitetu kupovnog ponašanja. Na primjer, dok je kod klasičnog direktnog marketinga jako važan datum zadnje trgovine, kod on-line targetiranja je podjednako važno kada je kupac zadnji put ulazio na sajt, koliko se zadržao, koje proizvode je razmatrao i slično.

Osnovni cilj ovog istraživanja je da se definišu efikasni modeli odlučivanja u direktnom marketingu zasnovani na data mining metodama, pri čemu, kako je već ranije istaknuto, SVM ima najvažniju ulogu (zbog svojih izvanrednih mogućnosti kod nelinearno separabilnih i nebalansiranih klasa, kao i kod nelinearnih zavisnosti i asimetrične distribucije zavisne varijable u slučaju regresije).

Ovaj cilj biće ostvaren kroz sledeće pomoćne ciljeve:

1. Definisane prediktivnog RFM segmentacionog modela koji prevazilazi problem nebalansiranosti klasa (tj. problem najmanjeg segmenta najvrijednijih kupaca).
2. Definisane *customer response* modela koji prevazilazi problem minorne klase (tj. problem izuzetno malog broja respondenata u odnosu na nerespondentne).
3. Definisane modela za predikciju profitabilnosti respondenata koji prevazilazi problem asimetrične (*skewed*) distribucije profitabilnosti (tj. problem malog broja visoko profitabilnih respondenata).
4. Komparacija i unapređenje performansi primijenjenog SVM-RE metoda se *ensemble* metodama

Kada je riječ o prvom cilju, problem klasične RFM segmentacije kupaca je što korisnik ručno definiše na koliko segmenata će kupci biti podijeljeni i koji segmenti će biti targetirani u direktnoj kampanji. U literaturi je potvrđeno da data mining klasterizacija automatski identifikuje homogene grupe kupaca (klasterne) (Cheng and Chen, 2009) a broj klastera se kod ove metode određuje na osnovu indikatora o performansama klasterizacije kao što je Davies-Bouldin (DB) indeks (Khalili-Damghani, Abdi & Abolmakarem, 2018; Rogić and Kaščelan, 2019), čime se obezbjeđuje objektivna segmentacija kupaca. Kod klasterizacije u istraživanju autora Cheng and Chen (2009), RFM atributi su uniformno kodirani tako da 20% najrecentnijih kupaca dobija *score* 5 (najčešće 20%, mada ovaj procenat korisnik može ručno mijenjati), sledećih 20% *score* 4 i tako dalje. Na isti način su kodirani i atributi za frekvenciju trgovanja i monetarnu vrijednost kupca. Međutim, s obzirom da klasterizacija operiše sa numeričkim atributima, ovi atributi, koji su numerički po definiciji, se mogu koristiti bez kodiranja, čime se eliminiše subjektivnost i smanjuje gubitak značajnih informacija (tj. finih razlika između kupaca po vrijednosti ovih atributa) (Rogić and Kaščelan, 2019).

Kada su kupci podijeljeni u segmente, sledeći korak prediktivnog segmentacionog modela je

predviđanje pripadnosti kupca odgovarajućem segmentu, na osnovu njegovih karakteristika i podataka o ponudi u kampanji (podaci o proizvodima, popustu i slično). Cheng and Chen (2009) su kao prediktore uključili i RFM atribut što dovodi do povećanja tačnosti predikcije (jer su segmenti definisani na bazi ovih atributa). Međutim, njima se može apsorbovati uticaj ostalih prediktora i izgubiti važne informacije potrebne za targetiranje novih kupaca (Rogić and Kaščelan, 2019).

S obzirom da je klaster najvrijednijih kupaca po pravilu i najmanji, većina prediktivnih metoda dovodi do pogrešne klasifikacije takvih kupaca (problem minorne klase). Pogrešna klasifikacija minorne klase obično dovodi do niskih vrijednosti (koje nekada iznose i nula) za prediktivne pokazatelje ove klase kao što su *class recall* (procenat tačno klasifikovanih aktuelnih primjera klase) i *class precision* (procenat tačno predviđenih primjera klase).

Kao što je već naglašeno, SVM je u literaturi potvrđen kao klasifikacioni metod koji se uspješno bori sa ovim problemom. Sa druge strane, za interpretaciju SVM klasifikacije potreban je metod koji ekstrahuje pravila iz SVM izlaza. Najčešće se u tu svrhu primjenjuje DT metod koji generiše sveobuhvatni skup jednostavnih if-then pravila. Kombinovanje SVM i DT metoda na ovaj način dobija se hibridni SVM-RE metod koji povećava tačnost klasifikacije minorne klase i generiše eksplicitna pravila klasifikacije (Kaščelan et al., 2014; Rogić and Kaščelan, 2019). Imajući u vidu ove, u prethodnoj literaturi potvrđene, mogućnosti SVM-RE metoda, u cilju povećanja efikasnosti RFM segmentacionog modela, definisane su hipoteze H1 i H2 (sa odgovarajućim podhipotezama):

H1: Primjenom data mining klasterizacije povećava se efikasnost RFM segmentacije kupaca i izbjegava subjektivnost pri izboru broja segmenata.

H1.1: Primjenom k-means klasterizacije na nekodiranim RFM atributima i Davies-Bouldin (DB) indeksa za izbor broja klastera, automatski se realizuje segmentacija kupaca sa maksimalnom homogenošću unutar klastera, maksimalnom heterogenošću između različitih klastera i optimalnim brojem klastera

H2: Primjenom hibridnog SVM-RE metoda povećava se efikasnost targetiranja i predikcije najvrijednijih kupaca kod RFM segmentacionog metoda, čime se smanjuju nepotrebni troškovi kampanje, povećavaju ukupni prihodi i na osnovu ekstrahovanih pravila formira profil segmenta najvrijednijih kupaca koji omogućava efikasniju interakciju sa njima.

H2.1: Hibridni SVM-RE metod, koristeći karakteristike kupaca i podatke o proizvodima, predviđa pripadnost kupca RFM klasteru najvrijednijih kupaca (minorna klasa) sa *class precision* i *class recall* većim od 50% tj. prevazilazi problem pogrešne klasifikacije minorne klase.

H2.2: SVM metod, predprocesiranjem podataka o kupovnim transakcijama tj. eliminisanjem preklapanja i nebalansiranosti klasa (klastera kupaca), povećava prediktivne performanse DT metoda.

H2.3: DT interpretira SVM model, tj. ekstrahuje pravila iz SVM izlaza sa visokim stepenom povjerenja tj. sa *fidelity* većim od 80%.

Što se tiče drugog definisanog cilja, kod *customer response* modela je ekstremno izražen problem minorne klase jer je procenat korisnika koji odgovore na kampanju obično niži od 5%. Kod većine klasifikatora ovaj problem dovodi do izuzetno niskih vrijednosti (često su ove vrijednosti nula) za senzitivnost (procenat tačno klasifikovanih aktuelnih primjera pozitivne klase – u ovom slučaju pozitivna klasa je klasa respondenata tj. minorna) i specifičnost (procenat tačno predviđenih primjera pozitivne klase). Takođe, pokazatelj prediktivnih performansi *Area Under the ROC Curve* – AUC (kriva koja prikazuje odnos između tačno klasifikovanih aktuelnih primjera i pogrešno predviđenih primjera pozitivne klase) može imati vrijednosti koje su jednake ili jedva nešto veće od 0.5 što znači da se model malo razlikuje od *random* pogađanja.

Uzimajući u obzir prethodno obrazložene mogućnosti SVM-RE metoda, da bi se prevazišao ovaj problem, kao i u slučaju hipoteze H2, biće testirane njegove mogućnosti da balansira klase i poveća prediktivne performanse, u smislu smanjenja pogrešne klasifikacije tako minorne klase. *Customer response* model kao prediktore obično uključuje karakteristike kupaca i RFM attribute. Međutim, uključivanjem Web metrika (on-line ponašanja kupaca) kao prediktora može unaprijediti identifikovanje najvjerovatnijih respondenata (Kim & Yum, 2011). Uzimajući ovo u obzir, za povećanje efikasnosti *customer response* modela biće testirana hipoteza H3 (tj. odgovarajuća podhipoteza).

H3: Primjenom hibridnog SVM-RE metoda povećava se efikasnost targetiranja i predikcije kupaca koji će najvjerovatnije odgovoriti na direktnu kampanju, čime se postiže ušteda nepotrebnih troškova, povećanje ukupnog profita ostvarenog u kampanji i formiranje profila respondenata na osnovu ekstrahovanih pravila koji omogućava efikasniju interakciju sa njima.

H3.1: Hibridni SVM-RE metod, koristeći karakteristike kupaca, podatke o proizvodima, podatke o kupovnom ponašanju (RFM attribute) i Web metrike (u slučaju on-line kampanje) predviđa vjerovatnoću odgovora na kampanju (*binary choice* 1/0) sa *sensitivity* i *specificity* većim od 50% i AUC značajno većim od 0.5 tj. prevazilazi problem pogrešne klasifikacije minorne klase.

S obzirom na multikolinearnost RFM atributa kao i zakrivljenu (*skewed*) distribuciju profitabilnosti kao zavisne varijable, a imajući u vidu mogućnosti SVM regresije potvrđene u prethodnim istraživanjima, **za realizaciju trećeg cilja** biće testirana hipoteza H4 (tj. odgovarajuća podhipoteza):

H4: Primjenom SVM regresije povećava se efikasnost targetiranja i predikcije najprofitabilnijih respondenata, čime se postiže povećanje ukupnog profita od direktne kampanje i formiranje profila najprofitabilnijih respondenata kako bi se ostvarila efikasnija interakcija sa ovom najvažnijom kategorijom kupaca za kompaniju.

H4.1: SVM regresija, koristeći karakteristike kupaca, podatke o proizvodima, RFM attribute i Web metrike (u slučaju on-line kampanje), predviđa profitabilnost kupca sa dobrim prediktivnim

performansama (Root Mean Squared Error – *RMSE* tj. greškom manjom od 10% prosječnog iznosa profita i *R-kvadrat* tj. koeficijentom determinacije većim od 0.5 dobijenih kros-validacijom) tj. prevazilazi probleme asimetričnosti (*skewed*) distribucije zavisne varijable i multikolinearnosti nezavisnih varijabli.

Na kraju kao **četvrti postavljeni cilj**, biće napravljena komparacija primijenjenog SVM-RE metoda sa *ensemble* metodama (Bootstrap Aggregating - Bagging, Adaptive Boosting - AdaBoost i Random Forest) koje su u literaturi prepoznate kao efikasne u povećanju prediktivnih performansi slabih klasifikatora (Dietterich, 2002; Zhang & Ma, 2012; Migueis et al. 2017). Osnovna ideja ovih metoda je da se generiše više klasifikatora na random podskupovima podataka (model je sa vraćanjem što znači da isti podatak može biti uključen kod sledećeg uzorkovanja). Rezultati se na kraju agregiraju najčešće tako što modeli glasaju i uzima se onaj rezultat za koji je glasalo najviše modela. Kod nebalansiranih klasa, iz veće klase se slučajno bira broj primjera jednak manjoj klasi i tako pokušava riješiti problem nebalansiranosti. S obzirom da se podaci koji ulaze u veću klasu biraju *random*, ove metode smanjuju problem gubitka informacija koje mogu biti važne za diferencijaciju između klasa. U literaturi je već potvrđeno da daju bolje rezultate od *under-sampling* i *over-sampling* metoda (Galar et al. 2012; Migueis et al. 2017). Farquad & Bose (2012) su potvrdili da *Random Forest* metoda primijenjena poslije SVM predprocesiranja daje značajno bolje rezultate nego prije SVM procesiranja, odnosno ako se primijeni samostalno. Kang et al. (2012) su pokazali da se primjenom *ensemble* metoda povećavaju prediktivne performanse kod *customer response* modela.

Dakle, postavlja se pitanje da li SVM-RE metod daje bolje rezultate kod balansiranja klasa od *ensemble* metoda, u domenu razmatrane problematike direktnog marketinga, i može li se rezultat dobijen SVM-RE metodom popraviti i koliko kombinovanjem sa *ensemble* pristupom (s obzirom na neizvjesnost rezultata koji se mogu dobiti, za četvrti cilj nisu definisane hipoteze, već je formulacija data u vidu istraživačkog pitanja).

B3. Metode i plan istraživanja

Kao osnovni empirijski materijal u radu će biti korištena dva skupa podataka koji su preuzeti od kompanije Sport Vision Montenegro (dio regionalnog Sport Vision sistema – vodećeg distributera sportske opreme na Balkanu). Pomoću ova dva skupa podataka biće obavljeno testiranje definisanih hipoteza (tj. podhipoteza).

Prvi skup koji se koristi kod prediktivne segmentacije (hipoteze H1 i H2), čine podaci o kupovnim transakcijama u okviru direktnih kampanja za period od početka septembra 2018. do kraja januara 2019. godine (ukupno 1606 kupaca). Podaci su preuzeti direktno iz baze podataka kompanije i pripremljene za analizu prečišćavanjem podataka i izračunavanjem izvedenih atributa (RFM atributa) na nivou kupca. Kategoričke varijable su transformisane u *dummy* varijable a numeričke normalizovane [0,1] rang transformacijom.

U sličnim istraživanjima, Cheng and Chen (2009) su za empirijsko testiranje predložene metode koristili jedan skup podataka sa 401 instancom, dok su Khalili-Damghani, Abdi & Abolmakarem (2018) eksperimentalno testirali svoj metod na dva skupa podataka koji sadrže 476 i 1000

instanci. Za testiranje SVM metode kao predprocesora koji balansira podatke, Farquard & Bose, (2012) su koristili jedan skup podataka koji sadrži 5822 instance. Rogić and Kaščelan (2019) su za predikciju vrijednosti kupaca pomoću SVM-RE metoda koristili isti skup podataka i postigli dobru prediktivnu tačnost. Uzimajući u obzir iskustva iz prethodnih istraživanja, kao i činjenicu da veći skup instanci za obučavanje modela može samo povećati prediktivne performanse, veličina skupa za empirijsko testiranje metoda za prediktivnu segmentaciju se može smatrati dovoljnom.

Drugi skup čine podaci za period od oktobra 2018. do aprila 2019. godine, o kampanjama (sa preko 100 000 on-line pristupa), koje su komunicirane putem on-line medija *Facebook*, *Messenger* i *Instagram*, što, osim ova tri medija uključuje i *Facebook Audience Network*¹ (na ovom skupu podataka biće testirane hipoteze H3 i H4). Za analitiku, metriku i prikupljanje podataka o potrošačima i kampanjama korištene su baze podataka dobijene iz *Google Analytics* alata i *Facebook Ads Manager*-a. Svaka kampanja sadrži podatke o: periodu trajanja kampanje, broju ostvarenih sesija, prosječnom trajanju sesije, *bounce rate*², prihodu, broju transakcija, kategoriji kupljenih proizvoda, regionu odakle se pristupa sajtu, operativnom sistemu i tipu uređaja sa kojeg se pristupa. Svaki od podataka zajedno sa jedinstvenim identifikatorom posjetioca sajta (*Client ID*)³ predstavlja jednu cjelinu podataka. Sve cjeline podataka su spojene pomoću SQL upita, na osnovu *Client ID*-a i kampanja direktnog marketinga, na koje su potrošači odgovorili u navedenom periodu. Konačni rezultat je jedinstvena baza podataka o kampanjama i transakcijama potrošača, koji su „kliknuli“ na neku od kampanja direktnog marketinga za polugodišnji period, kao i njihovih karakteristika.

Ukupno ova baza će sadržati preko 100 000 on-line pristupa za period od 6 mjeseci, od kojih će za testiranje biti korišćen podskup čija će veličina zavisi od jačine servera i vremena potrebnog za obučavanje SVM modela. Potvrda validnosti takvog testnog skupa može se naći u prethodnim istraživanjima. Na primjer, Olson and Chae (2012) su za predikciju odgovora kupca na kampanju kao testne skupove uzeli samo narudžbe iz posljednja 4 mjeseca i skupove podataka od preko sto hiljada instanci sveli na 20 000. Kim and Yum (2011) su kod testiranja sistema za on-line preporuku proizvoda, koristili jedan eksperimentalni skup podataka koji obuhvata pristupe korisnika u periodu od 50 dana, sa ukupno 2 465 instanci. Za testiranje *customer response* modela baziranog na SVM predikciji, Kim, Chae & Olson (2012) su koristili tri skupa podataka sa malom, srednjom i velikom stopom odgovora. Sva tri skupa su sadržala približno po 100 000 podataka ali su random izabrali 33% i 50% predstavnika nerespondenata (*undersampling*). Za predikciju profitabilnosti kupaca Kim, Lee & Cho, 2008 su koristili skup od ukupno 101 532 podataka iz kojeg su izdvojili 4 000 kupaca od kojih je ostvarena zarada (respondenata).

¹ *Facebook Audience Network* je mreža za oglašavanje za mobilne aplikacije, namijenjena kupcima koji pristupaju sa mobilnih web sajtova i aplikacija koje nisu *Facebook*, proširujući svoj doseg (*reach*) izvan platforme. Iako se reklama plasira na medijima koji nisu u vlasništvu *Facebook*-a, platforma za kreiranje i distribuciju pripada *Facebook*-u.

² *Bounce rate* predstavlja procenat posetilaca određenog web sajta, koji su napustili sajt nakon pregleda samo jedne stranice.

³ *Client ID* predstavlja nasumično generisani string (od strane *Google Analytics*-a), koji se čuva u kolačićima (*cookies*) *Internet browser*-a, tako da se naredne posjete tom sajtu mogu povezati sa istim korisnikom.

Za testiranje hipoteza biće korišten metod ***k-means* klasterizacije**, klasifikacioni **SVM i DT metodi** tj. hibridni **SVM-RE metod**, kao i **regresioni SVM metod**.

***K-means* klasterizacija** (MacQueen, 1967), je tehnika koja, za odabranu vrijednost k , identifikuje k klastera objekata, koji su bazirani na blizini objekata centru grupe, pri čemu je centar definisan kao prosjek od n -dimenzionalnih vektora atributa u okviru svakog klastera (EMC Education Services, 2015). Dakle, *k-means* predstavlja tehniku koja bez nadzora (*unsupervised*) klasifikuje potrošače u određeni broj klastera, pri čemu su potrošači u okviru klastera slični, a klasteri među sobom različiti. Za definisanje optimalnog broja klastera, koristi se *Davies-Bouldin* indeks. Ovaj indeks mjeri euklidsko rastojanje od centara, unutar i između klastera. Na nivou svakog klastera uzima se maksimalni koeficijent rasipanja unutar klastera i mjera odvojenosti od ostalih klastera. Zatim se izračunava prosjek za sve klasterne. Niže apsolutne vrijednosti DB indeksa ukazuju na bolji kvalitet klasterizacije.

Uzimajući u obzir da ovaj metod obezbjeđuje automatsku klasterizaciju i određivanje optimalnog broja klastera na osnovu DB indeksa, odgovarajući je za objektivnu segmentaciju kupaca.

Za linearno neseparabilne klase, Vapnik (2010) je predložio **SVM metod** koji preslikava podatke (koje posmatra kao n -dimenzionalne vektore) iz originalnog prostora u prostor veće dimenzije (*feature space*), gdje je klase moguće separirati pomoću hiperravni. Pronalaženje takve hiperravni se svodi na minimizaciju rastojanja između njenog krajnjeg položaja (tako da praznina između klasa tj. margina bude što veća) i najbližih tačaka (*support vector-a*). Umjesto eksplicitne funkcije preslikavanja u prostor veće dimenzije, koristi se kernel funkcija koja omogućava izračunavanje skalarnog proizvoda vektora u originalnom prostoru (kernel trik). Maksimizacija margine u prostoru veće dimenzije se svodi na optimizacioni problem kvadratnog konveksnog programiranja u originalnom prostoru, uz upotrebu kernel funkcije. Mogu se upotrijebiti različite kernel funkcije ali je često najefikasnija i najviše se koristi *Radial Basis Function* (RBF) (Sanderson et al., 2010). Obučavanje SVM klasifikatora se realizuje izborom optimanih vrijednosti parametra γ za RBF kernel, i parametra C koji predstavlja granicu za marginu tj. prazan prostor između klasa. Izbor manjih vrijednosti za parametar C smanjuje over-fitting i povećava generalnost SVM modela tj. njegove prediktivne performanse.

Prednost SVM metoda u odnosu na neuronske mreže kod nelinearnih problema, potiče od geometrijske interpretacije i činjenice da je u suštini SVM metoda problem konveksnog matematičkog programiranja koji garantuje globalni minimum. Naime, neuronske mreže imaju empirijsku minimizaciju greške, tj. greška se koriguje svaki put kada neki primjer prođe kroz mrežu, što može dovesti do zapadanja u lokalni minimum. Za razliku od njih SVM ima strukturnu minimizaciju (minimizacija rastojanja između margine i *support* vektora) koja se svodi na minimum kvadratne konveksne forme uz odgovarajuće uslove, tako da SVM uvijek pronalazi globalni minimum. Modeli neuronske mreže su često previše prilagođeni podacima za obučavanje (*overfitting*) i imaju lošije rezultate na nepoznatom skupu podataka. Za razliku od njih, SVM ima izvanrednu moć generalizacije zahvaljujući mogućnosti širenja margine. Osim toga, neuronska mreža zahtijeva podešavanje velikog broja parametara, dok složenost modela u velikoj mjeri zavisi od broja prediktora. Složenost SVM modela zavisi isključivo od broja support

vektora, pa njegovo obučavanje zahtijeva manje kompjuterskog vremena i prostora bez obzira na broj prediktora.

S obzirom na problematiku istraživanja (linearno neseeparabilne i nebalansirane kalse respondenata i nerespndenata, kao i zakrivljenost profitabilnosti respondenata i nelinearne veze sa prediktorima), i uzimajući u obzir navedene prednosti u odnosu na neuronske mreže, primjena SVM metoda je opravdana.

Pravila iz SVM izlaza biće izvedene pomoću klasifikacionog **DT metoda** koji dijeli polazni skup podataka po vrijednostima atributa tako da podskupovi sadrže što više primjera jedne klase. Kriterijum po kojem se obavlja podjela (mjera za kvalitet podjele) može biti *information gain*, *gain index*, *gini index* ili tačnost cijelog deveta - *accuracy of the whole tree*. U toku induktivne podjele formira se model u vidu drveta. Putanje od korjena ka listovima definišu *if-then* klasifikaciona pravila u terminima prediktivnih atributa.

Obučavanje prediktivnih modela biće realizovano primjenom *k-fold* kros-validacije. Ovaj postupak sa stratifikovanim uzorkovanjem podrazumijeva da se polazni skup podataka dijeli na k podskupova, pri čemu se vodi računa da procenti zastupljenosti klasa u podskupovima odgovaraju procentima zastupljenosti klasa u čitavom skupu podataka. Zatim se k-1 podskupova koristi za obučavanje modela (*training set*) a jedan od podskupova za validaciju tj. testiranje kako taj model radi na nepoznatom skupu podataka (*test set*). Postupak se ponavlja k puta tako da svaki od k podskupova bude test set. Na svakoj iteraciji se izračunavaju parametri za klasifikacione performanse (*accuracy rate*, *class precision* i *class recall*) i na kraju nalazi njihova prosječna vrijednost.

Optimizacija parametara za SVM i DT modele biće realizovano *grid-search* tehnikom, koja predstavlja osnovni metod optimizacije parametara. Sa ovom tehnikom, gradi se model za svaku moguću kombinaciju svih ponuđenih vrednosti parametara, koristeći k-fold kros-validaciju. Procjenjujući svaki model pojedinačno bira se arhitektura koja daje najbolje rezultate.

Na kraju, postupak za realizaciju hibridnog **SVM-RE metoda** sastoji se od sledećih koraka:

1. Obučavanje SVM modela (kombinovanjem grid-search-a i kros-validacije) i generisanje SVM izlaza
2. Zamjena aktuelnog *class label-a* u polaznom skupu podataka sa *class label-om* predviđenim od strane SVMa (predprocesiranje podataka)
3. Generisanje DT modela sa SVM *class label-om* (RE iz SVM modela)

Osnovna ideja SVM regresionog metoda (SVR) je da se N podataka za obučavanje modela koji se sastoje od n regresora posmatraju kao N vektora u n-dimenzionalnom prostoru (input space). U slučaju da je veza između regresora i zavisne varijable nelinearna vektori se preslikavaju u prostor veće dimenzije (feature space) gdje je moguće naći optimalnu hiperravan koja linearno modelira ovu vezu. SVR, sa jedne strane teži da minimizuje grešku u ocjeni zavisne varijable, dok sa druge strane nastoji da model u prostoru veće dimenzije bude što ravniji kako bi se povećala njegova generalnost tj. tačnost predviđanja na nepoznatom skupu podataka. Da bi model bio

ravniji potrebno je minimizovati intezitet vektora normalnosti hiperravni ali tako da odstupanja zavisne varijable dobijenih modelom od njenih aktuelnih vrijednosti budu najviše epsilon. Drugim riječima, prilikom minimizacije, greške manje od epsilon se ne uzimaju u obzir (epsilon intensive – loss function), već se nastoji da odstupanja ne budu veća od ovako zadate granice. Dakle, SVR pokušava da riješi konveksni optimization problem.

Nesenzitivna epsilon zona (koja sadrži tačke za koje je greška u ocjeni manja od epsilon) može se malo proširiti uvođenjem dozvoljene devijacije za epsilon. Elastičnost epsilon zone kontroliše parametar C kojim se postiže kompromis između ravnosti modela i epsilon tačnosti. Veće vrijednosti parametra C dozvoljavaju da model postane neravaniji tj. oblikuje se prema podacima u training skupu (overfitting), smanjujući na taj način njegovu generalnost (pri čemu devijacije epsilon ostaju male tj. postiže se epsilon tačnost). Manje vrijednosti parametra C smanjuju broj support vector-a tj. složenost modela i povećavaju njegovu ravnost a sami tim i generalnost (pri čemu se dopuštaju veće devijacije epsilon tj. smanjuje se epsilon tačnost).

Obučavanje SVR modela sastoji se od izbora optimalne kombinacije parametara C, epsilon i γ (iz RBF kernela). Veće vrijednosti parametra C i manje vrijednosti parametra epsilon dovode do manje greške u ocjeni zavisne varijable na podacima za obučavanje modela, ali smanjuju generalnost modela tj. njegovu prediktivnu moć na nepoznatom skupu podataka. Parametar γ zavisi od distribucije podataka za obučavanje modela.

Za izbor SVR parametara (C, epsilon i γ) biće primijenjena grid-search tehnika u kombinaciji sa k-fold kros-validacijom.

S obzirom da je SVR neparametarski metod (koji ne pati od problema standardnih regresionih modela) sa navedenim mogućnostima generalizacije, ovaj metod može postići dobru predikciju profitabilnosti respondenata i njegova primjena je opravdana.

Za komparaciju SVM-RE metoda sa **ensemble** pristupom biće primijenjene **Bagging**, **AdaBoost** i **Random Forest** metode. **Bagging** poboljšava stabilnost i tačnost algoritama mašinskog učenja koji se koriste u statističkoj klasifikaciji i regresiji. Iako se obično primenjuje na DT metodama, može se koristiti sa bilo kojom drugom metodom. Sastoji se od dva dijela - *bootstrapping* i agregacija. *Bootstrapping* je tehnika uzorkovanja koja podrazumijeva da se podskupovi skupa podataka za obučavanje slučajno biraju sa ponavljanjem k puta. Zatim se na svakom od ovih k podskupova generiše po jedan model. Smisao uzorkovanja sa ponavljanjem je učiniti ponovno uzorkovanje zaista slučajnim. Ako se izvrši bez ponavljanja, izvučeni uzorci će zavisiti od prethodnih i stoga neće biti slučajni. Predviđanja iz gornjih k modela sakupljaju se da bi napravili konačno kombinovano predviđanje tj. agregacija rezultata. Agregacija rezultata može da se izvrši na osnovu predviđanja ili verovatnoće predviđanja koja su napravili *bootstrap* pojedinačni modeli. Bagging pomaže u smanjenju varijanse i izbjegava prekomerno podešavanje modela prema podacima na kojima se obučava- *overfitting*. Optimizacija bagging modela podrazumijeva izbor optimalnog broja modela k. **AdaBoost** klasifikator kombinuje više slabih klasifikatora da bi se dobio jaki klasifikator. Jedan klasifikator može loše klasifikovati primjere, ali ako kombinujemo više klasifikatora sa odabirom različitog skupa podataka za obučavanje pri

svakoj iteraciji i dodeljivanjem odgovarajuće težine podacima za konačno glasanje, možemo imati dobru ocjenu tačnosti za ukupni klasifikator. AdaBoost funkcioniše tako što pridaje veće težine instancama koje su pogrešno klasifikovane, a manje onima koji su već dobro obrađeni. Optimizacija podrazumijeva izbor optimalnog broja iteracija. **Random Forest** metod kombinuje *bootstrapping* skupa podataka za obučavanje, random izbor prediktora i DT metod. Ovaj metod generiše „šumu“ DT modela za svaki slučajno izabrani podskup podataka. Optimizacija podrazumijeva izbor optimalnog broja DT modela.

Komparacijom predloženog SVM-RE metoda sa gore navedenim *ensemble* metodama, ukoliko se pokaže njegova superiornost, može se naglasiti efikasnost modela koji su predmet istraživanja i pojačati naučni doprinos disertacije što opravdava uporebu ovih metoda.

Izrada doktorske disertacije biće sprovedena kroz tri faze, i to:

1. U prvoj fazi je prikupljena literatura iz oblasti istraživanja direktnog marketinga, primjene data mining metoda u direktnom marketingu, kao i radovi iz oblasti SVM metode, RFM klasifikacije, DT modela i drugih relevantnih tema za ovo istraživanje.
2. U drugoj fazi biće prikupljeni i pripremljeni podaci iz poslovne baze kompanije koja se bavi distribucijom sportske opreme, na čijem primjeru će biti testiran model. Podaci o kupcima, proizvodima i odgovorima na on-line direktnu kampanju, nakon ekstrakcije iz baze, biće formatirani i pripremljeni za obradu, tj. treću fazu istraživanja.
3. U trećoj fazi biće testirane predložene metode tj. odgovarajuće hipoteze na pripremljenim podacima, kreiranjem odgovarajućih procesa za obučavanje i testiranje modela u Rapid Miner-u. Rezultati će zatim biti sumirani i diskutovani, nakon čega slijedi konačna forma teksta doktorske disertacije.

Teorijsku podlogu za definisanje hipoteza i dizajniranje istraživanja predstavljaju naučni radovi, magistarske i doktorske disertacije i knjige iz oblasti direktnog marketinga, marketinga na društvenim mrežama, data mining metoda, RFM klasterizacije, SVM modela, DT metoda i drugih relevantnih oblasti. Među značajnim časopisima za ovu temu, čija će arhiva biti detaljno istražena, su: *Expert Systems with Applications*, *Journal of Business Research*, *Intelligent Automation & Soft Computing*, *Decision Support Systems*, *Service Business*, *Journal of Interactive Marketing*, *Journal of Direct Marketing*, *Journal of Consumer Marketing*, *Journal of Product & Brand Management*, *Journal of Direct, Data and Digital Marketing Practice* i drugi.

Prikupljanje literature relevantne za istraživanje počelo je u novembru 2018. godine. Prikupljena literatura korišćena je za pripremu i sprovođenje polaznih istraživanja. Rad na temu „*Customer value prediction in direct marketing using hybrid Support Vector Machine Rule Extraction method*“ prezentovan je na *MADEISD Workshop-u – Modern Approaches in Data Engineering and Information System Design*, koji se organizovao u sklopu *European Conference on Advances in Databases and Information Systems (ADBIS 2019)* na Bledu, u Sloveniji od 8. do 11. septembra 2019. godine. Rad je objavljen kao poglavlje u Springerovoj monografiji *Communications in Computer and Information Science (CCIS) book series* (dostavljen u prilogu). Rad je proglašen za najbolji rad u okviru

MADEISD Workshop-a (potvrda u prilogu).

B4. Naučni doprinos

U ovom istraživanju biće definisani efikasni prediktivni modeli odlučivanja u direktnom marketingu zasnovani na data mining metodama, koji prevazilaze nedostatke postojećih modela nastalih zbog problema minorne klase i asimetrične distribucije profitabilnosti respondenata, što predstavlja i glavni naučni doprinos.

Kao rješenje za pomenute nedostatke u ovom radu je predložen SVM metod. Ovaj metod je u prethodnim istraživanjima potvrđen kao efikasan u slučaju nebalansiranosti i linearne neseparabilnosti klasa kao i kod neregularne distribucije zavisne varijable i nelinearnih zavisnosti u slučaju regresije. Međutim, po prvi put se u ovom istraživanju i na ovaj način primjenjuje na probleme iz oblasti direktnog marketinga.

SVM se koristi kao predprocesor koji prečišćava podatke tj. separira i balansira klase. Za ekstrakciju pravila prediktivne klasifikacije primjenjuje se DT metod na izlazu koji generiše SVM.

Primjenom ovog hibridnog SVM-RE metoda povećavaju se prediktivne performanse modela odlučivanja u direktnom marketingu, što dovodi do smanjenja nepotrebnih troškova i povećanja ukupnih prihoda od kampanje. Naime, povećava se tačnost klasifikacije za minornu klasu (*class recall*) i na taj način omogućava tačno targetiranje većeg broja kupaca čija je vjerovatnoća odgovora na direktnu kampanju najveća, a samim tim raste i vjerovatnoća za ostvarenje većeg prihoda od kampanje. Takođe, povećanjem prediktivne tačnosti modela za minornu klasu (*class precision*) smanjuje se broj pogrešno predviđenih respondenata, čime se smanjuju nepotrebni troškovi kampanje koji bi nastali zbog realizacije direktne kampanje prema kupcima koji vjerovatno neće odgovoriti.

Generisanjem eksplicitnih pravila povećava se efikasnost interakcije sa najvažnijim segmentima kupaca: sa onim koji su skoro odgovorili na kampanju, koji najčešće odgovaraju i koji najviše potroše, kao i sa onim koji će najviše vjerovatno odgovoriti na kampanju, što rezultira većim stepenom povjerenja i lojalnosti ovih kupaca i jačanjem strateških odnosa sa njima. Naime, veoma je važno da marketari imaju eksplicitni opis profila kupca koji visoko vjerovatno odgovara na kampanju (preko njegovih karakteristika, kupovnog ponašanja, on-line ponašanja i ostalih dostupnih podataka u bazi), da bi mogli efikasno da prepoznaju potencijalne respondente, prilagode im ponudu i tako povećaju efikasnost interakcije sa njima.

Predikcija profitabilnosti respondenata je poznati model odlučivanja u direktnom marketingu, koji se u prethodnim istraživanjima realizovao uglavnom putem standardnih ili specijalnih parametarskih regresionih modela koji se mogu izabrati sa problemom asimetrične distribucije profitabilnosti respondenata. Međutim, kod ovog modela se pojavljuju i problemi multikolinearnosti regresora (u prvom redu RFM atributa), nelinearne zavisnosti između regresora i profitabilnosti, kao i problem specifikacije odgovarajuće funkcionalne forme regresionog modela. SVM regresija je neparametarski regresioni metod (ne zahtijeva specifikaciju funkcionalne forme) koji uspješno modelira nelinearne veze i zahvaljujući mogućnostima

generalizacije (adekvatnim podešavanjem parametara) ima dobre prediktivne performanse, što je u literaturi potvrđeno. U osnovi SVM regresije je rješavanje optimizacionog problema, pa ovaj metod nema probleme koji su karakteristični za metodu najmanjih kvadrata.

Zbog navedenih prednosti u radu je za povećanje efikasnosti modela za selekciju kupaca na osnovu njihove očekivane profitabilnosti predložen SVM regresioni metod. On povećava efikasnost modela tako što zahvaljujući izvanrednim mogućnostima generalizacije bez obzira na sve pomenute probleme, povećava tačnost predikcije visoko profitabilnih respondenata.

Upućivanjem direktne ponude takvim kupcima, iako ih je po pravilu jako malo, može se ostvariti veći prihod od on-line kampanje, nego kampanjom prema većem broju recentnih i frekventnih kupaca. Često je profitabilnost kupca obrnuto proporcionalna broju odgovora na kampanju, tj. visoko profitabilni kupci kupuju rjeđe, ali troše više. Jasno je da je tačna predikcija takvih kupaca od velikog značaja za kompaniju (jedan tačno predviđeni kupac može donijeti ogroman profit za kampanju, dok jedan pogrešno predviđeni visoko profitabilni kupac, prema kome neće biti upućena ponuda, može značajno smanjiti prihod od kampanje). Senzitivnom analizom se mogu eksplicitno utvrditi karakteristike takvih kupaca i njihovo kupovno ponašanje, što je marketarima važno za targetiranje novih kupaca i uspostavljanje kvalitetnih odnosa sa ovom kategorijom najvažnijih kupaca.

Prema našim saznanjima, u ovom trenutku, postoji manji broj istraživanja koja uključuju podatke iz *online* direktnih marketing kampanja plasiranih putem društvenih mreža, uprkos činjenici da je ovo danas dominantan vid direktnog marketinga u praksi. Ovaj rad u tom smislu predstavlja dopunu postojeće teorije, jer se predloženi modeli mogu primijeniti i na podacima iz *online* direktnih kampanja, uključivanjem odgovarajućih Web metrika kao prediktora. Studija slučaja na kojoj se testiraju modeli se upravo odnosi na online direktne kampanje preko društvenih mreža, gdje su pored standardnih prediktora uključene i Web metrike preuzete sa *Google Analytics* alata i *Facebook Ads Manager-a* (broj pristupa sajtu, prosječno vrijeme provedeno na sajtu i slično), kao i dodatne karakteristike kupaca vezane za tehnologiju (tip uređaja sa kojeg se pristupa, operativni sistem i slično) i lokaciju koju koriste za pristup. Kod on-line marketinga ovi parametri su podjednako važni, ako ne i važniji za selekciju kupaca od standardnih (kao što su demografske karakteristike kupaca i kupovno ponašanje). Na osnovu dostupne literature, po prvi put se u ovom radu definiše koncept modela odlučivanja i za *online* direktni marketing preko društvenih mreža, uz povećanje efikasnosti putem SVM metoda.

Pored naučnog doprinosa rad ima aplikativni, praktični i društveni značaj.

Naime, biće razvijeni prediktivni modeli u vidu gotovih operatora u RapidMiner alatu, koje kompanije mogu direktno implementirati i koristiti kao podršku pri donošenju odluka u direktnim kampanjama. Međutim, treba napomenuti da upotreba prediktivnih modela u direktnom marketingu uvijek zahtijeva sofisticiranog marketing analitičara koji poznaje *data mining* metode.

Koristi od ovakvog sistema nemaju samo kompanije, čiji će poruke i ponude biti plasirane najzainteresovanijim kupcima i na taj način napraviti uštede i ostvariti veći prihod od kampanje,

već i kupci, kojima će biti ponuđeni proizvodi ili usluge koje su direktno vezane za njihove interese i potrebe. Dakle, ovako osmišljen sistem za podršku odlučivanju u direktnom marketingu doprinosi i kompaniji i kupcima na različite načine, počev od materijalnih koristi, pa sve do kvalitetnijih odnosa između njih.

S obzirom da se u Crnoj Gori tek pojavljuju kompanije koje koriste *online* direktne kampanje preko društvenih mreža, rezultati ovog istraživanja ih mogu podstaći da uzmu u razmatranje sve mogućnosti i prednosti ove vrste direktne marketing aktivnosti u kombinaciji sa podrškom u donošenju odluka pomoću metoda mašinskog učenja.

B5. Finansijska i organizaciona izvodljivost istraživanja

Komisija smatra da Ekonomski fakultet u Podgorici i Univerzitet Crne Gore mogu obezbijediti finansijske i organizacione uslove za realizaciju istraživanja neophodnog za izradu doktorske disertacije.

Mišljenje i prijedlog komisije

Prema mišljenju Komisije doktorska disertacija kandidatkinje Sunčice Rogić pod naslovom „**Prediktivni modeli odlučivanja u direktnom marketingu bazirani na *Support Vector Machine* metodu**“ ima veoma aktuelnu problematiku i predstavlja originalano i vrijedno naučno istraživanje koje će proširiti teoriju prediktivnog modeliranja procesa odlučivanja u direktnom marketingu novim efikasnim modelima, doprinijeti efikasnosti procesa odlučivanja iz ovog domena u preduzećima i ukazati na prednosti koje donosi kombinovanje metoda poslovne inteligencije i direktnog marketinga. Predloženi sadržaj i metode su adekvatni a plan istraživanja izvodljiv. Kandidatkinja je zadovoljila sve neophodne kriterijume i uspješno odbranila polazna istraživanja. Imajući sve ovo u vidu, **Komisija jednoglasno predlaže Vijeću Ekonomskog fakulteta i Senatu Univerziteta Crne Gore da prihvate pozitivan izvještaj Komisije i odobre dalju izradu doktorske disertacije.**

Prijedlog izmjene naslova

Komisija je predložila kandidatu promjenu naslova iz „**Konceptualni model sistema poslovne inteligencije u direktnom marketingu baziran na *Support Vector Machine* metodu**“ u „**Prediktivni modeli odlučivanja u direktnom marketingu bazirani na *Support Vector Machine* metodu**“ sa obrazloženjem da je sistem poslovne inteligencije širi pojam koji obuhvata i komponente koje nisu razmatrane u radu (data warehouse i OLAP). Kandidat je prihvatio promjenu naslova pa je to sada konačni naslov doktorske disertacije.

Prijedlog promjene mentora i/ili imenovanje drugog mentora

Nema predloga

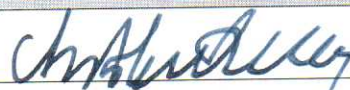
Planirana odbrana doktorske disertacije

2021. godina

Izdvojeno mišljenje

Nema izdvojenog mišljenja

Ime i prezime

Napomena		
(popuniti po potrebi)		
ZAKLJUČAK		
Predložena tema po svom sadržaju odgovara nivou doktorskih studija.	DA	NE
Tema je originalan naučno-istraživački rad koji odgovara međunarodnim kriterijumima kvaliteta disertacije.	DA	NE
Kandidat može na osnovu sopstvenog akademskog kvaliteta i stečenog znanja da uz adekvatno mentorsko vođenje realizuje postavljeni cilj i dokaže hipoteze.	DA	NE
Komisija za ocjenu podobnosti teme i kandidata		
Prof dr Ivan Luković, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, Srbija – predsjednik Komisije		
Prof dr Ljiljana Kaščelan, Ekonomski fakultet, Univerzitet Crne Gore, Crna Gora - mentor		
Prof dr Boban Melović, Ekonomski fakultet, Univerzitet Crne Gore, Crna Gora - član		
U Podgorici, 19.12. 2019. godine	  DEKAN	

PRILOG

PITANJA KOMISIJE ZA OCJENU PODOBNOSTI DOKTORSKE TEZE I KANDIDATA	
Prof dr Ivan Luković, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, Srbija –pedsjednik Komisije	Da li je termin „sistem poslovne inteligencije“ u naslovu preširok s obzirom da ovaj sistem obuhvata komponente koje nisu razmatrane u radu?
	Da li je moguće hipoteze preciznije kvantifikovati?
	U čemu je prednost <i>Support Vector Machine</i> metoda u odnosu na neuronske mreže kod nelinearnog prediktivnog modeliranja?
	Koliko bi <i>ensemble</i> metode mogle da poboljšaju prediktivne performanse modela, posebno senzitivnost i specifičnost minorne klase?
Prof dr Ljiljana Kaščelan, Ekonomski fakultet, Univerzitet Crne Gore, Crna Gora - mentor	Kakva je uloga <i>Google Analytics</i> alata u <i>online</i> direktnom marketingu?
	Na koji način se podaci preuzimaju iz <i>Google Analytics</i> baze podataka i kako se pripremaju za analizu?
	Koliko je procentualno od planiranih podataka preuzeto iz ove baze podataka i kad se očekuje da će skup podataka biti kompletiran i spreman za analizu.
Prof dr Boban Melović, Ekonomski fakultet, Univerzitet Crne Gore, Crna Gora	Da li je adekvatna upotreba termina „konceptualni model“u kontekstu predložene teme?
	Da li je fokus istraživanja na selekciji kupaca kod online direktnih kampanja ili se modeli odlučivanja mogu primijeniti u direktnom marketingu uopšte? Da li se osim selekcije razmatra i interakcija sa kupcima u cilju jačanje odnosa sa njima?
	Da li između naslova rada i obrazloženja teme postoji terminološka nepreciznost? Naime, u naslovu se koristi termin „direktni marketing“ a hipoteze, ciljevi i doprinos su fokusirane na online direktni marketing. Shodno tome potrebno je prilagoditi navedene elemente prijave naslovu.
PITANJA PUBLIKE DATA U PISANOJ FORMI	
(Ime i prezime)	Nije bilo pitanja publike u pisanoj formi.
(Ime i prezime)	
(Ime i prezime)	

ZNAČAJNI KOMENTARI

Na osnovu člana 32 stav 1 tačka 14 Statuta Univerziteta Crne Gore, u vezi sa članom 34 Pravila doktorskih studija, Senat Univerziteta Crne Gore, u postupku razmatranja prijedloga Vijeća Ekonomskog fakulteta i na prijedlog Centra za doktorske studije, na sjednici održanoj 11.11.2019. godine, donio je sljedeću

O D L U K U

Imenuje se Komisija za ocjenu podobnosti doktorske teze i kandidatkinje mr Sunčice Rogić, u sastavu:

1. Dr Ljiljana Kaščelan, redovni profesor Ekonomskog fakulteta Univerziteta Crne Gore
2. Dr Boban Melović, vanredni profesor Ekonomskog fakulteta Univerziteta Crne Gore
3. Dr Ivan Luković, redovni profesor Fakulteta tehničkih nauka Univerziteta u Novom Sadu

II

Zadatak Komisije je da, u roku od 45 dana od dana javnog izlaganja studenta podnese Vijeću Ekonomskog fakulteta i Senatu izvještaj o ocjeni podobnosti doktorske teze i kandidata.

III

Odluka stupa na snagu danom donošenja.

Broj: 03- 1884/4
Podgorica, 11.11.2019. godine



PREDSJEDNIK SENATA

Prof. dr Danilo Nikolić, rektor



UNIVERZITET CRNE GORE
EKONOMSKI FAKULTET
EKONOMIJA

Broj dosijea: 2/2018

Na osnovu člana 165 Zakona o opštem upravnom postupku ("Službeni list RCG" br. 60/03) i službene evidencije, a po zahtjevu Rogić Branko Sunčica, izdaje se

POTVRDA O STUDIRANJU

Student **Rogić Branko Sunčica**, rođena **23-10-1992** godine u mjestu **Podgorica**, opština **Podgorica**, Republika **Crna Gora**, upisana je studijske **2018/2019** godine, u **I** godinu studija, kao student koji se **samofinansira** na **akademske doktorske studije**, studijski program **EKONOMIJA**, koji realizuje **EKONOMSKI FAKULTET** - Podgorica Univerziteta Crne Gore u trajanju od **3 (tri)** godine sa obimom **180** ECTS kredita.

Studijske **2018/2019** godine prijavila je *da sluša* **5** predmeta sa **40.00** (četrdeset) ECTS kredita.

Po prvi put iz **I (prve)** godine, prijavila je *da sluša* **5** predmeta sa **40.00** (četrdeset) ECTS kredita, što iznosi **66.67%** od ukupnog broja ECTS kredita u **I** godinu.

Saglasno Statutu Univerziteta Crne Gore, **Rogić Branko Sunčica** je po prvi put prijavila *da sluša* **2/3**, odnosno **66,67% (šezdesetšest 67/100 %)**, od ukupnog broja ECTS kredita sa **I** godine i studijske **2018/2019** ima **status redovnog studenta** koji se **samofinansira**.

Uvjerjenje se izdaje na osnovu službene evidencije, a u svrhu ostvarivanja prava na: (dječji dodatak, porodičnu penziju, invalidski dodatak, zdravstvenu legitimaciju, povlašćenu vožnju za gradski saobraćaj, studentski dom, studentski kredit, stipendiju, regulisanje vojne obaveze i slično).



Broj: 20 / 01 -
Podgorica, 23.01.2020 godine

SEKRETAR,

Vuković Jajkova

Na osnovu člana 165 stava 1 Zakona o opštem upravnom postupku ("Službeni list RCG", broj 60/03.), člana 115 stava 2 Zakona o visokom obrazovanju ("Službeni list CG", broj 44/14.) i službene evidencije, a po zahtjevu studenta Rogić Branko Sunčica, izdaje se

UVJERENJE O POLOŽENIM ISPITIMA

Student **Rogić Branko Sunčica**, rođena **23-10-1992** godine u mjestu **Podgorica**, opština **Podgorica**, Republika **Crna Gora**, upisana je studijske **2018/2019** godine, u **I** godinu studija, kao student koji se **samofinansira** na **doktorske akademske studije**, studijski program **EKONOMIJA**, koji realizuje **EKONOMSKI FAKULTET - Podgorica** Univerziteta Crne Gore u trajanju od **3 (tri)** godine sa obimom **180 ECTS kredita**.

Student je položio ispite iz sljedećih predmeta:

Redni broj	Semestar	Naziv predmeta	Ocjena	Uspjeh	Broj ECTS kredita
1.	1	EKONOMETRIJSKI METODI I MODELI	"A"	(odličan)	8.00
2.	1	MAKROEKONOMIJA	"A"	(odličan)	8.00
3.	1	METODE EKONOMSKIH ISTRAŽIVANJA	"A"	(odličan)	6.00
4.	1	METRIKA MARKETINGA	"A"	(odličan)	8.00
5.	2	KOMPJUTERSKE METODE ZA ANALIZU PODATAKA U EKONOMIJ	"A"	(odličan)	10.00

Zaključno sa rednim brojem **5**.

Ostvareni uspjeh u toku dosadašnjih studija je:

- srednja ocjena položenih ispita **"A" (10.00)**
- ukupan broj osvojenih ECTS kredita **40.00** ili **66.67%**
- indeks uspjeha **6.67**.

Uvjerjenje se izdaje na osnovu službene evidencije, a u svrhu ostvarivanja prava na: (dječji dodatak, porodičnu penziju, invalidski dodatak, zdravstvenu legitimaciju, povlašćenu vožnju za gradski saobraćaj, studentski dom, studentski kredit, stipendiju, regulisanje vojne obaveze i slično).

Broj: 20 / 01 -
Podgorica, 23.01.2020 godine



SEKRETAR,

Muković Jajkova