

UNIVERZITET CRNE GORE

Prirodno-matematički fakultet Podgorica

Crna Gora
UNIVERZITET CRNE GORE
PRIRODNO-MATEMATIČKI FAKULTET

Broj 900
Podgorica, 14.05. 2014 god.

Vijeću Prirodno-matematičkog fakulteta

IZVJEŠTAJ KOMISIJE O PODOBNOSTI TEME MAGISTARSKOG RADA KANDIDATA MILICE (MARKOVIĆ) MARIĆ

Vijeće Prirodno-matematičkog fakulteta na sjednici održanoj 20.4.2021. godine imenovalo je mentora i Komisiju za ocjenu podobnosti teme za izradu magistarskog rada pod nazivom "Primjena data mining tehnika u telemarketingu u cilju povećanja procenta javljanja korisnika (Reach Rate)", kandidata Milice Marić, u sastavu:

dr Milenko Mosurović, redovni profesor - član

dr Aleksandar Popović, vanredni profesor - član

dr Savo Tomović, vanredni profesor – mentor.

Nakon uvida u podneseni materijal, a u vezi sa članom 24 Pravila studiranja na postdiplomskim studijama, podnosimo sljedeći

IZVJEŠTAJ

Milica Marić, specijalista računarskih nauka, prijavila je temu magistarskog rada pod nazivom "Primjena data mining tehnika u telemarketingu u cilju povećanja procenta javljanja korisnika (Reach Rate)". Tema spada u oblast računarskih nauka za koju je matičan Prirodno-matematički fakultet. Dokumenta podnesena za prijavu teme sadrže: biografiju kandidata, naziv i kratku razradu teme, kao i kratko obrazloženje predmeta istraživanja i strukture rada.

Podaci o kandidatu

Milica Marić je rođena u Nikšiću, 04.12.1985. godine. Osnovnu i srednju školu je završila u Nikšiću. Prirodno-matematički fakultet, odjek Računarske nauke upisuje 2004. U septembru 2008. godine upisuje magistarske studije, na odsjeku za Računarske nauke. Položila je sve ispite predviđene programom studija. Nakon završenih specijalističkih studija, Milica Marić je nastavila praktično usavršavanje u Crnogorskom Telekomu, gdje je radila više od 12 godina. Trenutno je zaposlena u Erste banci.

Aktuelnost teme

Ovaj rad predstavlja istraživanje na temu primjene data mining algoritama u cilju povećanja procenta javljanja korisnika - Reach Rate u telemarketingu. Pristup u ovoj oblasti obično se svodio na selekciju korisnika koji se pozivaju metodom slučajnog uzorka. Upotrebom data mining tehnika pokušava se kreiranje prediktivnog modela koji definiše kada je najbolje zvati korisnika kako bi se povećao procenat javljanja (reach rate). Istraživanje će obuhvatiti operativne baze podataka sa servisne platforme Kron, koju Telesales odjeljenje Crnogorskog Telekoma koristi za pozivanje korisnika. Biće predloženo više tehnika pripreme podataka, odabira atributa i algoritama mašinskog učenja za konstruisanje prediktivnog modela.

Predmet istraživanja rada su efekti implementacije i upotrebe prediktivnih data mining tehnika sa ciljem unapređivanja procedure za Telesales. Telesales je način direktnog marketinga u kojem prodavac nudi potencijalnim korisnicima da kupe određeni proizvod/servis preko telefona. Tradicionalno, potencijalni kupci se identifikuju ručnom analizom dostupnih obilježja – atributa, a mogu biti definisane neke heuristike, kao na primjer analitičari prodaje mogu unaprijediti broj javljanja tako što će označiti korisnike koji su se u prošlosti javljali, pa njih staviti na početak spiska za pozivanje i zvati ih više puta. Ručnom analizom može biti unaprijeđen broj javljanja, ali mogu biti propušteni neki važni atributi i heuristike koje značajno utiču na rezultat. Bez prediktivnog modela, vjerovatnoća da će se ovako odabrani korisnik javiti je 50%. Dakle, predmet istraživanja sastoji se u inženjeringu obilježja ili atributa koji su dostupni u operativnim podacima ili mogu biti izvedeni iz njih a koji imaju značajnu informacionu dobit vezano za učenje prediktivnog modela čija bi primjena značajno povećala vjerovatnoću da će odabrani korisnik pozitivno odgovoriti tokom telesales postupka.

U literaturi je prepoznat pozitivan efekat data mining tehnika, koje sa dobro pripremljenim

podacima o korisnicima i na osnovu raspoloživog istorijskog skupa, mogu sa visokom vjerovatnoćom da predvide da li će se u određenom intervalu, određenog dana korisnik javiti. Nakon primijenjenih data mining tehnika dobiće se odgovor na pitanje „kojeg korisnika zvati i kada”. U literaturi su dostupni rezultati studija iz drugih domena primjene, dok u domenu telekomunikacionih usluga, koliko je nama poznato, nije publikovano nijedno istraživanje i prateći rezultati.

Cilj, struktura i metodologija rada

Telesales Crnogorskog Telekomu napravi prosječno 30.000 svih poziva za mjesec dana. Od toga je uspješno ostvareno približno 55% poziva (mjereno za period od 2015-2019). Sa strane biznis analitičara u odjeljenju Prodaje Crnogorskog Telekomu, ovaj procenat je označen kao nedovoljan. Osnovni motiv istraživanja sastoji se u osmišljavanju i implementaciji naprednih data mining tehnika čija bi primjena garantovala značajno povećanje prethodno navedenog procenta. Drugim riječima, motivacija za ovaj rad je potreba da se simulira Teledales procedura u okviru Crnogorskog Telekomu i da se ispita da li je moguće postići da se za kraće vrijeme javi više korisnika čime bi se poboljšali rezultati Telesales službe u smislu efikasnosti.

Osnovna ideja je da se na osnovu raspoloživih podataka iz operativnih baza Crnogorskog Telekomu, „nauči“ prediktivni model koji može da procijeni vjerovatnoću da će se u datom vremenskom intervalu (dan ili period dana) javiti konkretni korisnik. Na osnovu tih vjerovatnoća sistem bi sortirao korisnike u opadajućem poretку, pa bi Telesales agenti pozivali korisnike datim redoslijedom do određenog praga, kada više nije isplativo pozivati korisnike.

Cilj ovog magistarskog rada je pokazati da se na realnom problemu i realnom skupu podataka, primjenom data mining tehnika može postići veća efikasnost Telesales procedure, odnosno povećanje uspješnih poziva u odnosu na ukupan broj upućenih poziva. Konkretno, sa strane biznis analitičara u odjeljenju Prodaje Crnogorskog Telekomu, zadat je poslovni cilj da se postigne povećanje od 10% na kumulativnom uzorku od 50% populacije. Kumulativni uzorak od 50% populacije predstavlja prvih 50% od ukupnog broja korisnika sortiranih po vjerovatnoći koju vrati model, od veće ka manjoj.

Dodatno, teorijski cilj ovog rada je proučiti ponašanje modernih boosting algoritama poput Xgboosta i GBM na prethodno definisanom problemu.

Zadati problem će biti rješavan u skladu sa procedurem Knowledge Discovery in Databases ili KDD koja će biti prilagođena procedurama i pohranjenim podacima u repozitorijumima Crnogorskog Telekomu.

Ovo podrazumijeva dizajn i implementaciju postupaka za pre-procesiranje, čišćenje, integraciju, transformaciju i selekciju podataka koji će formirati dataset za istraživanje.

Dalje, je neophodno osmisliti data mining korak, što porazumijeva primjenu specifičnih algoritama za rješavanje postavljenog zadatka. Planirano je da se testiraju algoritmi koji su u literaturi prepoznati kao najuspješniji za zadatke sličnog tipa i to: Distributed Random forests (DRF), Gradient Boosting Machine (GBM), Generalized Linear Models (GLM), kao i složeni (heterogeni) ansambl sastavljen od prethodno pomenutih modela i Extreme Gradient Boost (XGB).

Praksa je da se algoritmi obučavaju na 80% a testiraju na 20% cjelokupnog dataseta. Dodatno, u ovom radu biće upotrijebljena funkcija stratifikacija koja kreira uzorak ravnomjerno u odnosu na response varijablu u smislu procenta prisutnosti obje klase. Na ovaj način se garantuje maksimalna objektivnost kada su ostvareni rezultati u pitanju. Iz istih razloga, testiranje će biti implementirano po pravilu unakrsne validacije sa 5 slojeva (5-fold cross validation).

Posljednji korak u KDD metodologiji je ocjenjivanje performanse modela i prezentacija dobijenih šablona i znanja. S tim u vezi, generisani modeli biće procijenjeni na osnovu sljedećih metrika: AUC, TOP_LIFT, f0.5, F1, precision, accuracy [1].

Za implementaciju je izabran R programski jezik, sa pozivanjem H2O 3 platforme za kreiranje, estimaciju i tuniranje hiperparametara modela.

Rad će se sastojati od sljedećih cjelina: Uvod, Pregled dosadašnjih istraživanja, Data mining, Prediktivno modelovanje, Odnos bias-varijansa, Evaluacija (i selekcija) modela, Metode Ansambla, Odabir i inženjering atributa (feature selection and feature engineering), Studija slučaja i Zaključak.

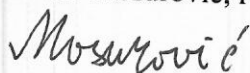
Zaključak

Na osnovu prethodno izloženog smatramo da predložena tema magistarskog rada kandidata Milice Marić ispunjava sve uslove predviđene Pravilima studiranja na postdiplomskim studijama, propisanim od strane Senata Univerziteta Crne Gore. Komisija predlaže Vijeću Prirodno-matematičkog fakulteta da kandidatkinji Milici Marić odobri izradu magistarskog rada pod nazivom "Primjena data mining tehnika u telemarketingu u cilju povećanja procenta javljanja korisnika (Reach Rate)".

U Podgorici, 13. 5. 2021. godine

KOMISIJA

Dr Milenko Mosurović, redovni professor – član



Dr Aleksandar Popović, vandredni professor – član



Dr Savo Tomović, vandredni professor – mentor

