

UNIVERZITET CRNE GORE

Prirodno-matematički fakultet Podgorica

UNIVERZITET CRNE GORE
PRIRODNO-MATEMATIČKI FAKULTET

Broj 1276
Podgorica, 02.07 2021 god.

Vijeću Prirodno-matematičkog fakulteta;

Predmet: Izvještaj Komisije o pregledu i ocjeni magistarskog rada „*Primjena data mining tehnika u telemarketingu u cilju povećanja procenta javljanja korisnika (Reach Rate)*“ kandidata Milice Marić, specijaliste računarskih nauka.

Na osnovu člana 60 Statuta Univerziteta Crne Gore, a u vezi sa članovima 25 i 28 Pravila studiranja na postdiplomskim studijama, na sjednici Vijeća PMF-a od 13. maja 2021. godine imenovani smo za članove komisije za ocjenu magistarskog rada „*Primjena data mining tehnika u telemarketingu u cilju povećanja procenta javljanja korisnika (Reach Rate)*“ kandidata Milice Marić, specijaliste računarskih nauka.

Milica Marić je predala tekst magistarskog rada 24.06.2021. na uvid javnosti i ocjenu. Nakon uvida u podneseni tekst, a u vezi sa članom 29 Pravila studiranja na postdiplomskim studijama, podnosimo sljedeći

IZVJEŠTAJ

Magistarski rad Milice Marić pod nazivom „*Primjena data mining tehnika u telemarketingu u cilju povećanja procenta javljanja korisnika (Reach Rate)*“ napisan je na 125 strana, sadrži predgovor, izvod rada, izvod rada na engleskom jeziku, sadržaj, prilog i time ispunjava sve zahtjeve propisane članom 27 Pravila studiranja na postdiplomskim studijama.

Rad pripada oblasti računarskih nauka, odnosno uže specijalizovanoj oblasti Data mining. Sastoji se iz deset poglavlja u kojima se detaljno i nadovezano definišu potrebni pojmovi i prikazuju rezultati rada, a pozivajući se na odgovarajuću literaturu.

U uvodu su predstavljeni motivacija i ciljevi kao i definicija problema. U Data mining poglavlju su opisane faze procesa data mining-a kao i oblasti iz kojih se koriste određene tehnike.

„Pregled dosadašnjih istraživanja” je osvrt na dosadasnja istraživanja i stanje u nauci u oblasti primjene data mining algoritama u telemarketingu.

Poglavlje „Prediktivno modelovanje” (*Predictive modeling*) predstavlja teoretske osnove prediktivnog modelovanja, definiciju dva tipa problema: klasifikacioni i regresioni, kao i 2 različita tipa modela: linearni modeli (*linear*) i bazirani na stablima odlučivanja (*decision trees*). Odnos bias-vaijansa (*bias-variance tradeoff*) je opis jedne od glavnih dilema svih onih koji se bave prediktivnom analitikom. U tom poglavlju je opisano šta treba imati u vidu prilikom kreiranja modela da ne dođe do pretjeranog ili nedovoljnog obučavanja modela.

Selekcija i evaluacija modela (Model selection and evaluation) je poglavlje koje opisuje metode na osnovu kojih ocjenjujemo koliko je dobar naš model. U poglavlju Metode Ansambala (*Ensemble methods*) su predložene metode za poboljšanje tačnosti modela koje su našle široku primjenu u nauci podataka (data science).

Poglavlje „Odabir i inženjering atributa (*feature selection and feature engineering*)” opisuje inženjering odabira podataka kao krucijalnu oblast za postizanje dobre tačnosti modela.

Studija slučaja opisuje praktičnu primjenu navedenih metoda. Predstavljeni su podaci, inženjering atributa, metodologija i postignuti rezultati.

Konačno poglavlje „Zaključak” predstavlja sveobuhvatan pregled ciljeva i budućeg rada.

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 kreiran je prediktivni model koji definiše kada je najbolje zvati korisnika kako bi se povećao procenat javljanja (reach rate). Istraživanje je obuhvatilo operativne baze podataka sa

servisne platforme Kron, koju Telesales odjeljenje Crnogorskog Telekoma koristi za pozivanje korisnika. Predloženo je više tehnika pripreme podataka, odabira atributa i algoritama za konstruisanje prediktivnog modela.

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 predikcionog modela, vjerovatnoća da će se ovako odabrani korisnik javiti je 50%. Predmet istraživanja sastojao se u inženjeringu 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.

Kao rezultat istraživanja u okviru ovog magistarskog rada došlo se do zaključka 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 Telekoma, postavljen je 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. U radu je pokazano da je ostvareno poboljšanje od 30% na 50% populacije, uspješno je ostvareno je oko 80% poziva.

Za potrebe eksperimentalnog protokola dizajnirane su i implementarane procedura za pre-procesiranje, čišćenje, integraciju, transformaciju i selekciju podataka, odnosno za formiranje dataseta za istraživanje. Dalje, je neophodno bilo osmisliti data mining korak koji porazumijeva primjenu specifičnih algoritama za rješavanje postavljenog zadatka. Testirani su 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).

U ovom radu je 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 je implementirano po pravilu unakrsne validacije sa 5 slojeva (5-fold cross validation).

Generisani modeli su procijenjeni na osnovu sljedećih metrika: AUC, TOP_LIFT, f0.5, F1, precision, accuracy.

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

ZAKLJUČAK I PREDLOG

Na osnovu prethodno izloženog, Komisija smatra da je kandidat Milica Marić u potpunosti postigla očekivane rezultate i ciljeve definisane u prijavi magistarskog rada.

Komisija predlaže Vijeću Prirodno-matematičkog fakulteta da rad pod nazivom „*Primjena data mining tehnika u telemarketingu u cilju povećanja procenta javljanja korisnika (Reach Rate)*“ kandidata Milice Marić prihvati kao magistarski rad i odobri njegovu javnu usmenu odbranu.

U Podgorici, 2.7.2021.

dr Milenko Mosurović, redovni profesor – član

Mosurović

dr Aleksandar Popović, vanredni profesor - član

Popović

dr Savo Tomović, vanredni profesor – mentor

Tomović