

IZVEŠTAJ O OCENI DOKTORSKE DISERTACIJE

I PODACI O KOMISIJI

1. Datum i organ koji je imenovao komisiju

Dana 26.01.2024. godine, Nastavno naučno veće FEFA Fakulteta

2. Sastav komisije sa naznakom imena i prezimena svakog člana, zvanja, naziv fakulteta, ustanove:

- 1) Prof. dr Goran Pitić, redovni profesor FEFA Fakulteta, Univerzitet Metropolitan u Beogradu, predsednik komisije
- 2) Prof. dr Milan Nedeljković, redovni profesor FEFA Fakulteta, Univerzitet Metropolitan u Beogradu, mentor
- 3) Prof. dr Branko Urošević, redovni profesor Računarskog fakulteta, Univerzitet Union Beograd, član komisije

II PODACI O KANDIDATU

1. Ime, ime jednog roditelja, prezime:
Klod (Sergej) Kolaro

2. Datum rođenja, opština, država:
18.01.1967., Beograd, Republika Srbija

3. Naziv fakulteta, naziv studijskog programa osnovnih i diplomskih akademskih studija – master i stečeni naziva

Diplomske akademske studije: Građevinski fakultet, Univerziteta u Beogradu, (1986-1993), *Master inženjer građevinarstva za konstrukcije*, opšti uspeh: 7.52 i ocenom 10 na diplomskom ispitu.

4. Godina upisa na doktorske studije i naziv studijskog programa doktorskih studija

Upisan 2019. godine na doktorske studije FEFA fakulteta, Univerziteta Metropolitan u Beogradu; studijski program: Poslovna ekonomija

III NASLOV DOKTORSKE DISERTACIJE

Klasifikacioni algoritmi mašinskog učenja i njihova primena u finansijama

IV PREGLED DOKTORSKE DISERTACIJE

Navesti kratak sadržaj sa naznakom broja strana, poglavlja, slika, šema, grafikona i sl.

Doktorska disertacija Kloda Kolaroa je zasnovana na sedam poglavlja posvećenih pitanju analize klasifikacionih algoritama mašinskog učenja i njihove primene u finansijama. Rad je ukupnog obima od 153 strana, i sadrži 24 tabele i 48 grafikona. Pri izradi disertacije korišćeno je ukupno 106 referenci. Pored osnovnog teksta, rad sadrži ključnu dokumentacijsku informaciju, sadržaj, kao i sažetak na srpskom i engleskom jeziku. Nakon osnovnog teksta dat je prikaz literature, kao i tri priloga: Struktura podataka korišćenih u empirijskoj analizi, Dodatni rezultati empirijske analize i Delimičan prikaz razvijenih softverskih algoritama.

Kratak sadržaj disertacije je sledeći:

1. Digitalna transformacija
 - Nauka o podacima i veštačka inteligencija
 - Mašinsko učenje
2. Primena *ML* u finansijama i pregled literature
 - Linearna regresija
 - Logistička regresija
 - Opšti linearni modeli, GLM
 - Gausova diskriminantna analiza, GDA
 - Klasične statističke metode
3. Modeli mašinskog učenja
 - Naivni Bajes, NB
 - K - najbližih suseda, K-NN
 - Metoda potpornih vektora, SVM
 - Klasifikacioni modeli na bazi stabla odlučivanja
4. Evaluacija performansi klasifikacionih modela
5. Pretprocesiranje i transformacija podataka
 - Skaliranje podataka
 - Nedostajući podaci
 - Ekstremne vrednosti
 - Multikolinearnost
 - Nebalansirani podaci
6. Rezultat empirijskog istraživanja
 - Opis uzorka
 - Regresioni modeli

Modeli na bazi stabla odlučivanja

Naivni Bajes

K-NN

SVM

Optimalni ML modeli

Uporedna analiza performansi modela ML i Altman Z-score

7. Zaključak

Reference

Prilog

Struktura podataka

Rezultat empirijske analize

Delimičan prikaz razvijenih softverskih algoritama

Izjava o autorstvu

Izjava o istovetnosti štampane i elektronske verzije doktorskog rada

Izjava o korišćenju

V VREDNOVANJE POJEDINIH DELOVA DOKTORSKE DISERTACIJE

I Digitalna transformacija

U prvom delu disertacije, kandidat je predstavio pojam digitalne transformacije – sveobuhvatne promene u poslovanju inspirisane novim tehnologijama, koja je ključna kako bi kompanije očuvale svoju konkurentnost. Ova strateška odluka dovodi do značajnih promena u kompanijskom lancu vrednosti, kojima se kreiraju novi poslovni modeli, povećava operativna efikasnost, a razvojem novih personalizovanih digitalnih oblika komunikacije kreira superiorno korisničko iskustvo. Kandidat se potom fokusirao na interne aspekte digitalne transformacije, čiji je prvenstveni cilj kvalitetno i efikasno odlučivanje, na osnovu analize relevantnih podataka. Kandidat je predstavio osnove nauke o podacima i mašinskog učenja, kao i tipologiju modela mašinskog učenja kao podlogu čitaocu za nastavak istraživanja sprovedenog u disertaciji.

Pozitivno se ocenjuje.

II Primena ML u finansijama i pregled literature

U drugom poglavlju disertacije kandidat je dao detaljan pregled literature koja analizira primenu različitih modela mašinskog učenja u finansijama, sa posebnim naglaskom na modeliranje kreditnih rizika, što je predmet kasnijeg empirijskog istraživanja sprovedenog u disertaciji. Na taj način čitalac dobija

dovoljno predznanja o trenutno aktuelnim pravcima u literaturi relevantnoj za oblast istraživanja u disertaciji. Pored toga, kandidat u ovom poglavlju daje detaljan prikaz matematičkih osnova u korišćenju različitih modela nadgledanog mašinskog učenja (linearna regresija, logistička regresija, regularizacioni pristup regresiji, opšti linearni model i Gausova diskriminatorska analiza), kao i geometrijskih osnova implicitnih u analiziranim pristupima. Kandidat je takođe prikazao i druge pristupe koji se koriste u literaturi koja analizira kreditne rizike. Na taj način, čitalac dobija sveobuhvatan pregled metodologije koja se koristi i puno razumevanje osnova svake od analiziranih metoda što olakšava njihovu primenu i takođe se može koristiti u dodatne edukativne svrhe.

Pozitivno se ocenjuje.

III Modeli mašinskog učenja

U trećem poglavlju teze kandidat je predstavio dodatne, novije modele mašinskog učenja koji su prisutni u relevantnoj empirijskoj literaturi. Prvo je prikazan generativni Naivni Bajes model, zajedno sa svojim prednostima i ograničenjima. Fokus je potom bio na neparametričkim modelima baziranim na bliskosti (razdaljini), koji grupišu uzorke različitih klasa prema njihovoj sličnosti koja se meri njihovom međusobnom razdaljinom: K-NN (k najbližih suseda) i Metod potpornih vektora (SVM). Konačno, analizirana je klasa klasifikacionih modela na bazi stabla odlučivanja, čija primena je u ogromnom porastu u toku poslednje dve decenije. Kao i u prethodnom poglavlju, kandidat je za svaki pojedinačni model dao detaljan prikaz matematičkih osnova modela, dajući doprinos ograničenoj literaturi na srpskom jeziku koja daje prikaz relevantnih modela. Na taj način, čitalac dobija sveobuhvatan pregled metodologije koja se koristi i puno razumevanje osnova svake od analiziranih metoda što olakšava njihovu primenu i takođe se može koristiti u dodatne edukativne svrhe.

Pozitivno se ocenjuje.

IV Evaluacija performansi klasifikacionih modela

U četvrtom poglavlju doktorske disertacije kandidat je dao prikaz kriterijuma koji se koriste u procesu evaluacije performanse klasifikacionih modela. Detaljno su prikazani koncepti Matrice konfuzije, Prostora ispod ROC krive, Kappa indeks, F1 score i Brier score. Ovo poglavlje kompletira teorijske postavke potrebne za sprovođenje empirijskih istraživanja korišćenjem različitih modela mašinskog učenja opisanih u poglavljima II i III.

Pozitivno se ocenjuje.

V Pretprocesiranje i transformacija podataka

Pretprocesiranje i transformisanje podataka predstavljaju neophodan i često ključan element u empirijskoj primeni modela mašinskog učenja. Peto poglavlje doktorske disertacije stoga daje prikaz osnovnih pristupa i metoda koji se koriste za pretprocesiranje i transformaciju podataka i koja su takođe korišćena u empirijskoj analizi u disertaciji. U poglavlju su analizirani načini skaliranja podataka (svođenja na uporedive vrednosti), načini aproksimacije nedostajućih vrednosti, načini rada sa ekstremnim vrednostima, slučajevi postojanja nebalansiranih podataka, kao i metode koje se mogu koristiti za prevazilaženje problema multikolinearnosti. Na taj način, čitalac dobija sveobuhvatan pregled osnovnih izazova sa kojima se istraživač susreće na početku rada sa podacima i primarnih pristupa koji se mogu koristiti za rešavanje ovih izazova.

Pozitivno se ocenjuje.

VI Rezultat empirijskog istraživanja

Šesto poglavlje doktorske disertacije prikazuje rezultate empirijskog istraživanja koje je sprovedeno u cilju merenja performansi modela koji su analizirani u prethodnim poglavljima. Performanse su testirane na dva nezavisna skupa podataka koji odgovaraju najčešćoj primeni modela mašinskog učenja u finansijama – na problemu predviđanja nenaplativih potraživanja prema kompanijama i predviđanja nenaplativih potraživanja prema pojedincima. U poglavlju je detaljno prikaz skup podataka koji se koriste u analizi, proces pripreme i transformacije podataka i dobijeni rezultati treniranja modela. Rezultati sprovedene analize na oba skupa podataka upućuju na isti zaključak da klasifikacioni modeli na bazi stabla odlučivanja pokazuju najbolje rezultate, a unutar njih klasa gradient boosting modela (GBM) je imala najbolje rezultate na oba skupa podataka. U empirijskoj analizi, kandidat je takođe uporedio performanse preporučenog modela mašinskog učenja (GBM) u odnosu na industrijske ad-hoc standarde (Altmanov Z-score) i pokazao značajna unapređenja koja model mašinskog učenja postiže.

Pozitivno se ocenjuje.

U zaključnim napomenama, kandidat je naveo ključne zaključke rada, uz sumirane rezultate istraživanja, isticanje naučnog doprinosa, navođenje mogućih ograničenja istraživanja na temu rada, i predloga smernica za buduća istraživanja.

Pozitivno se ocenjuje.

Literatura

Spisak korišćene literature sadrži 106 naslova. Korišćena je savremena domaća i inostrana literatura iz oblasti mašinskog učenja, statističke analize i finansija, posebno u vezi teme doktorske disertacije. Većinu korišćenih dela čine publikacije objavljene u visoko kotiranim međunarodnim i domaćim naučnim časopisima. Izvori literature su na adekvatan način citirani u tekstu disertacije.

Pozitivno se ocenjuje.

VII SPISAK NAUČNIH I STRUČNIH RADOVA KOJI SU OBJAVLJENI NA OSNOVU REZULTATA ISTRAŽIVANJA U OKVIRU RADA NA DOKTORSKOJ DISERTACIJI

VII ZAKLJUČCI ODNOSNO REZULTATI ISTRAŽIVANJA

Na osnovu predstavljenih rezultata istraživanja u okviru doktorske disertacije, došlo se do sledećih zaključaka i rezultata:

- A) U teorijskom delu istraživanja (poglavljima II-IV) kandidat je došao do zaključka da primena nadgledanih metoda mašinskog učenja u finansijama omogućava značajna unapređenja empirijskog okvira, pre svega u pogledu postizanja boljeg kvaliteta predikcije ciljane promenljive,

odnosno rešavanja takozvanog y problema u empirijskoj literaturi. Pored kvantifikacije asocijativnosti između pojedinačnih prediktora (ekonomskih promenljivih) i ciljane (zavisne) promenljive, modeli mašinskog učenja, u proseku, daju bolju preciznost predikcije zavisne promenljive uzimajući direktno u obzir postojanje nelinearnih veza, kao i sinergetskih efekata prediktora na vrednost zavisne promenljive. Postupkom regularizacije nadgledanih modela mašinskog učenja može se dodatno povećati tačnosti predikcije, kao i mogućnost generalizacije samih modela. U radu se konstatuje da i minimalna povećanja tačnosti predikcije donose značajnu vrednost kompanijama. Pored toga, kandidat ističe i nenadgledane modele mašinskog učenja kojima se rešavaju takozvani x problemi u empirijskoj literaturi koji pored samog kvaliteta predikcije omogućavaju da se na osnovu uočenih sličnosti i razlika mogu bolje razumeti podaci koji su predmet istraživanja. U ovom delu istraživanja kandidat je dao detaljan prikaz matematičkih osnova u korišćenju različitih modela mašinskog učenja, kao i geometrijskih osnova implicitnih u analiziranim pristupima, pružajući na jednom mestu putu transparentnost u razumevanju mehanizama koji su u osnovi svakog od analiziranih algoritama.

- B) U prvom delu empirijskog istraživanja (poglavlje V) kandidat je istakao različit značaj koji skaliranje podataka ima na performanse pojedinačnih algoritama mašinskog učenja, kao i ograničenja i pristrasnost u rezultatima modela mašinskog učenja koja se mogu javiti u slučaju nedostajućih podataka, prisustva ekstremnih vrednosti u podacima i/ili slučaja nebalansiranih podataka u klasifikacionim problemima. U ovom delu su takođe predstavljene metode koje se mogu koristiti za prevazilaženje svakog od navedenih pojedinačnih problema. Pored toga, rezultati istraživanja ističu i problem multikolinearnosti koji se često javlja u empirijskim istraživanjima i kako se metoda osnovnih komponenti može koristiti u rešavanju ovog problema.
- C) U drugom delu empirijskog istraživanja (poglavlje VI) došlo se do zaključka da algoritmi mašinskog učenja na bazi stabla odlučivanja ostvaruje najbolje performanse merene preciznošću predikcije na oba reprezentativna skupa analiziranih podataka. Modeli na bazi stabla odlučivanja su na drugom skupu podataka, sa 64 promenljive, ostvarili rast vrednost AUC za 12%, dok su u slučaju manjeg, prvog skupa podataka, sa 24 promenljive, ovi modeli postigli bolju vrednost AUC u odnosu na regresione modele za približno 5%. Razlika u performansama modela mašinskog učenja na bazi stabla odlučivanja i industrijskih benčmark modela (Zscore) je još veća. Razlog boljih performansi modela na bazi stabla odlučivanja je u činjenici da u slučaju kada se uzorci različitih klasa ne mogu razdvojiti linearnom graničnom linijom (ili sa hiperravni u višedimenzionalnom prostoru), algoritmi na bazi stabla su u stanju da bolje prepoznaju složene šablone između podataka i razgraničavaju podatke različitih klasa u više sekcija. Unutar klase modela na bazi stabla odlučivanja, došlo se do zaključka da gradient boosting modeli (GTB) ostvaruju najbolje performanse i prezentovana su jasna objašnjenja izvora njihovog unapređenja performansi.

Pozitivno se ocenjuje.

VIII OCENA NAČINA PRIKAZA I TUMAČENJA REZULTATA ISTRAŽIVANJA

Rezultati istraživanja doktorske disertacije kandidata Kloda Kolaroa prezentovani su na sistematičan način u vidu teksta, tabela i grafičkih prikaza i detaljno su obrazloženi. Disertacija je napisana jasno. Stil

pisanja je zadovoljavajući. Izbor vokabulara, gramatika i interpunkcija su adekvatni. Obrada, analiza i interpretacija rezultata u okviru istraživanja urađena je u skladu sa relevantnim statističkim metodama koje odgovaraju cilju istraživanja, i rezultati su predstavljeni jasno, pregledno i iscrpno, praćeni adekvatnim i logičnim objašnjenjima. Tabele i grafikoni su adekvatno označeni. Obuhvat i područja kojima se bavi disertacija, vrlo su značajni za oblast poslovne ekonomije.

Komisija pozitivno ocenjuje način prikaza i tumačenja rezultata istraživanja.

IX KONAČNA OCENA DOKTORSKE DISERTACIJE

1. Da li je disertacija napisana u skladu sa obrazloženjem navedenim u prijavi teme

Doktorska disertacija je napisana u potpunosti u skladu sa obrazloženjem navedenim u prijavi teme.

2. Da li disertacija sadrži sve bitne elemente

Doktorska disertacija sa svojim naslovom, sadržajem, ključnom dokumentacijskom informacijom, sažetkom, rezultatima sekundarnog i primarnog (empirijskog) istraživanja, interpretiranjem rezultata, zaključcima, i literaturom sadrži sve bitne elemente.

3. Po čemu je disertacija originalni doprinos nauci

Nakon razmatranja doktorske disertacije kandidata Kloda Kolaroa Komisija donosi zaključak da disertacija ima originalan doprinos nauci u oblasti poslovne ekonomije, koji se ogleda u produblivanju znanja u oblasti primene savremenih klasifikacionih modela mašinskog učenja u finansijama. Primarni doprinos doktorske disertacije ogleda se u: i) sistematizaciji i objedinjavanju relevantne literature koja se bavi primenom modela mašinskog učenja u finansijama; ii) detaljnom teorijskom prikazu modela mašinskog učenja i njihovih karakteristika (prednosti i nedostataka), u kontekstu procene kreditnih rizika; iii) detaljnom opisu procesa razvoja modela mašinskog učenja i uporedne analize njihovih performansi na dva nezavisna reprezentativna skupa podataka. Sekundarni doprinos doktorske disertacije ogleda se u: i) analiza uticaja neuravnoteženih podataka (minimalno prisustvo uzoraka jedne klase) na performanse modela mašinskog učenja; ii) analizi uticaja primene heterogenih klasifikatora (ensembling modela) na tačnost dobijene predikcije.

Doktorska disertacija, po prvi put u literaturi na srpskom jeziku, daje detaljan teorijski prikaz najčešće korišćenih modela mašinskog učenja u finansijama, odgovarajućih optimizacionih metoda, kao i sistematizaciju relevantne literature. Zainteresovana naučna i stručna javnost, na taj način, na jednom mestu može pronaći sve relevantne informacije potrebne za sprovođenje empirijskih istraživanja iz oblasti mašinskog učenja u finansijama.

Rezultati istraživanja prikazani u disertaciji su originalni doprinos autora.

4. Nedostaci disertacije i njihov uticaj na rezultat istraživanja

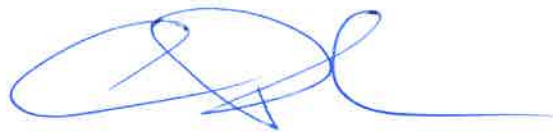
Doktorska disertacija nema nedostataka koji bi bitno uticali na konačan rezultat istraživanja.

X PREDLOG

Na osnovu ukupne ocene disertacije kandidata Kloda Kolaroa, Komisija predlaže da se doktorska disertacija pod nazivom „Klasifikacioni algoritmi mašinskog učenja i njihova primena u finansijama“ prihvati, a kandidatu odobri odbrana.

NAVISTI IME I ZVANJE ČLANOVA KOMISIJE
POTPISI ČLANOVA KOMISIJE

Beograd, mart 2024.



Prof. dr Goran Pitić, redovni profesor
Predsednik komisije
FEFA Fakulteta, Univerzitet Metropolitan u Beogradu



Prof. dr Milan Nedeljković, redovni profesor
Član komisije
FEFA Fakulteta, Univerzitet Metropolitan u Beogradu



Prof. dr Branko Urošević, redovni profesor
Član komisije
Računarskog fakulteta, Univerzitet Union Beograd

NAPOMENA: Član komisije koji ne želi da potpiše izveštaj jer se ne slaže sa mišljenjem većine članova komisije, dužan je da unese u izveštaj obrazloženje odnosno razloge zbog kojih ne želi da potpiše izveštaj.