

**PRIMENA TEHNOLOGIJA RAČUNARSTVA VISOKIH PERFORMANSI U
ISTRAŽIVANJU BEZBEDNOSTI SAOBRAĆAJA****APPLICATION OF TECHNOLOGIES OF HIGH PERFORMANCE COMPUTING IN
RESEARCH OF ROAD TRAFFIC SAFETY**Jovan Vunić, *Fakultet tehničkih nauka, Novi Sad***Oblast – ELEKTROTEHNIKA I RAČUNARSTVO**

Kratak sadržaj – Ovaj rad opisuje postupak primene tehnologija računarstva visokih performansi u analizi faktora koji utiču na stepen ozbiljnosti saobraćajnih nesreća. U radu su navedeni upotrebljene tehnologije i algoritmi mašinskog učenja, skupovi podataka sa kojima je rađeno i primenjena metodologija uz prikazane rezultate istraživanja. Rešenja su implementirana uz pomoć programskog jezika R i sistema za distribuiranu obradu podataka Apache Spark.

Ključne reči: Saobraćajne nesreće, Računarstvo visokih performansi, Mašinsko učenje, Apache Spark

Abstract – The paper presents how technologies of high performance computing are applied in an analysis of factors affecting the severity level of road traffic accidents. The paper specifies the employed machine learning technologies and algorithms, used data sets and applied methodology with presented results of the research. The solutions are implemented using R programming language and Apache Spark engine for distributed data processing.

Keywords: Road traffic accidents, High performance computing, Machine learning, Apache Spark

1. UVOD

Saobraćajne nesreće predstavljaju fenomen koji utiče ne samo na zdravstvene aspekte jedne države, već i na oblikovanje ekonomske slike društva. Bezbednost u saobraćaju predstavlja oblast koja pripada različitim naučnim krugovima, krećući se od sociologije, preko medicine, fizike i mehanike, sve do ekonomije i prava. Glavni cilj date oblasti je obezbeđivanje najvišeg mogućeg stepena sigurnosti učesnika u saobraćaju. Prema podacima *Svetske zdravstvene organizacije* [1], saobraćajne nezgode zauzimaju prvo mesto na lestvici uzroka smrtnih slučajeva među populacijom između 5 i 29 godina starosti.

Visok broj saobraćajnih nesreća često se vezuje za ekonomsko stanje države, što je usko povezano sa kvalitetom infrastrukture saobraćajnica. Sa druge strane, sve češće se javljaju sumnje i tvrdnje o povezanosti pojačanog događanja saobraćajnih nesreća sa socijalnim stanjem društva, specifičnim delovima kolovoza poput

raskrsnica ili delova gde je prisutna železnička pruga, ali i prirodnim fenomenima, poput različitih vremenskih uslova, godišnjih doba i toga koji period dana je u pitanju.

Mašinsko učenje, u kombinaciji sa tehnologijama računarstva visokih performansi, pokazalo se kao grana nauke kojom se, uz pomoć odgovarajućih tehnologija, može ispitati veza između saobraćajnih nesreća sa težim posledicama po učesnike i potencijalnih uzroka takvih ishoda. U ovu svrhu korišćeni su algoritmi bazirani na stablima odlučivanja, čije osobine omogućavaju sagledavanje šire slike u kontekstu važnosti različitih atributa u odnosu na ponašanje obučenog modela, kao i algoritam *logistička regresija*. Sa obzirom na količinu korišćenih podataka, svi procesi koji se tiču mašinskog učenja obavljani su uz pomoć odgovarajućih tehnologija računarstva visokih performansi.

2. PRETHODNA ISTRAŽIVANJA

U radu [2] diskutuje se o mogućnosti predviđanja saobraćajnih nesreća u realnom vremenu uz oslonac na različite grupe atributa, gde postoji uticaj svakoga od njih na mogućnost izazivanja nezgode u okvirima drumskog saobraćaja. Korišćeni su podaci koji se odnose na saobraćajne nesreće registrovane na teritoriji Sjedinjenih Američkih Država [2,3]. Rešenja su formirana upotrebom modela zasnovanog na dubokoj neuronskoj mreži. Kao uticajni faktori uzete su u obzir različite grupacije parametara, poput saobraćajnih elemenata, vremenskih uslova, interesnih tačaka i datumsko-vremenskih atributa. Mera F1 je upotrebljena za evaluaciju formiranih prediktivnih modela. Na osnovu predviđanja vršenog za šest odabranih gradova i geografske regione razumno velikih površina u toku kratkih vremenskih intervala, došlo se do zaključka da podaci o stanju toka saobraćaja u realnom vremenu, uz informacije o interesnim tačkama u blizini vozila, imaju uticaj na predviđanje događanja saobraćajnih nesreća u realnom vremenu.

Rad [4] bavi se evaluacijom performansi različitih modela mašinskog učenja, formiranih na osnovu algoritama *stablo odlučivanja* (engl. *Decision Tree*) i *nasumične šume* (engl. *Random Forests*) u cilju rešavanja problema predviđanja saobraćajnih nezgoda na teritoriji države Kalifornije, koja ulazi u sastav Sjedinjenih Američkih Država. Rađeno je sa skupom podataka koji obuhvata nesreće zabeležene širom Sjedinjenih Američkih Država [2,3]. Glavni cilj je bilo ispitivanje uticaja vremenskih uslova na izazivanje saobraćajne nesreće. Kao glavna mera evaluacije performansi formiranih modela

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Vladimir Ivančević.

upotrebljena je mera F1. Na osnovu dobijenih rezultata izveden je zaključak da postoji čvrsta veza između vremenskih uslova i mogućnosti događanja saobraćajnih nesreća.

3. SKUPOVI PODATAKA

Za potrebe ovog rada korišćeni su podaci sakupljeni na teritorijama dve države, Republike Srbije i Sjedinjenih Američkih Država. Podaci vezani za saobraćajne nesreće registrovane na teritoriji Republike Srbije odnose se na nesreće podeljene prema policijskim upravama i objavljeni su na Portalu otvorenih podataka Republike Srbije [5]. Dati podaci obuhvataju period između 2015. i 2021. godine. Skup podataka koji se odnosi na saobraćajne nezgode zabeležene na teritoriji Sjedinjenih Američkih Država objavljen je na sajtu Kaggle i obuhvata period između 2016. i 2020. godine [2,3].

U oba skupa podataka prisutan je atribut koji ukazuje na stepen ozbiljnosti saobraćajne nesreće. Kod podataka vezanih za teritoriju Republike Srbije dato obeležje nosi naziv *vrsta nesreće*, dok se u skupu podataka za teritoriju Sjedinjenih Američkih Država vodi pod imenom *ozbiljnost nesreće* (engl. *Severity*). Na osnovu datih atributa u oba skupa podataka formirano je novo, ciljno obeležje kategorijske prirode, nazvano *stepen ozbiljnosti nesreće* (engl. *Severity level*). Ciljno obeležje zauzima vrednosti *Sa materijalnom štetom* i *Sa fizičkom štetom* kod podataka vezanih za teritoriju Republike Srbije, odnosno *Low to Medium* i *Medium to High* kod podataka vezanih za teritoriju Sjedinjenih Američkih Država. Između ciljnih obeležja oba skupa podataka može da se povuče analogija, zato što oba ukazuju na nivo ozbiljnosti posledica konkretne saobraćajne nezgode na njene učesnike. Količina podataka vezanih za teritoriju Republike Srbije znatno je manja u odnosu na količinu podataka vezanih za teritoriju Sjedinjenih Američkih Država, gde je inicijalni odnos uzoraka 2906610 prema 195285 u korist Sjedinjenih Američkih Država. Broj sadržanih atributa takođe ide u korist Sjedinjenih Američkih Država, pošto se u skupu podataka koji se odnosi na njihovu teritoriju nalazi 47 različitih obeležja, naspram devet u okviru skupa podataka koji je vezan za nesreće registrovane u okviru Republike Srbije.

4. METODOLOGIJA

Za formiranje rešenja upotrebljena su četiri različita algoritma mašinskog učenja [6]: *logistička regresija* (engl. *Logistic regression*), *stablo odlučivanja* (engl. *Decision tree*), *nasumične šume* (engl. *Random forests*) i *stabla pojačana gradijentom* (engl. *Gradient Boosted Trees – GBT*). Poslednja tri su usko vezana za stabla odlučivanja. Za svaki od navedenih algoritama formirano je više modela za različite vrednosti odgovarajućih hiperparametara i odgovarajuće grupe prediktorskih obeležja. Grupe prediktorskih obeležja za podatke zabeležene na teritoriji Republike Srbije obuhvataju geolokacione, administrativne i sve dostupne prediktorske attribute, dok za Sjedinjene Američke Države postoji ukupno pet različitih grupa prediktorskih obeležja:

- geolokaciona obeležja,
- administrativna obeležja,
- meteorološka obeležja,

- unija geolokacionih, administrativnih i datumsko-vremenskih obeležja i
- svi dostupni prediktorski atributi.

Za *logističku regresiju* odabran je hiperparametar koji ukazuje na maksimalni broj iteracija tokom procesa obučavanja modela. Formirani su modeli za 50, 150, 300, 450 i 600 iteracija. Tokom formiranja modela zasnovanih na *stablu odlučivanja* podešavana je vrednost hiperparametra koji se odnosi na maksimalnu dubinu stabla koja može biti dostignuta tokom procesa obučavanja. Dubina stabla uzimana je u vrednostima od 5, 10, 15, 20, 25 i 30 nivoa. Algoritmom *nasumične šume* formirani su modeli tokom čijeg obučavanja se menjao broj stabala potrebnih za formiranje *nasumičnih šuma*. Modeli su formirani na osnovu *nasumičnih šuma* sastavljenih od 5, 100, 200, 300 i 400 stabala odlučivanja. Slično kao kod *logističke regresije*, za formiranje modela uz pomoć algoritma *stabla pojačana gradijentom* korišćene su različite vrednosti hiperparametra koji ukazuje na maksimalni broj iteracija tokom procesa obučavanja. Za različite modele obučavanje je trajalo u dužini od 5, 50, 100, 150 ili 200 iteracija. Postupak pretprocesiranja podataka izvršen je za podatke iz oba skupa i uključivao je odstranjivanje uzoraka kod kojih barem jedno od prediktorskih obeležja poseduje nedostajuću vrednost, kao i formiranje dodatnih datumsko-vremenskih prediktorskih atributa koji se odnose na period dana, naziv dana u sedmici i godišnje doba u okviru koga je registrovana saobraćajna nesreća.

Nad oba korišćena skupa podataka izvršeno je balansiranje podataka na osnovu ciljnog obeležja putem algoritma *ROSE* (engl. *Random Over-Sampling Examples*) [7], usled pojave visokog stepena disbalansa podataka. Obučavajući i testni skup formirani su za oba balansirana skupa podataka u razmeri 75% prema 25% polaznog skupa podataka. Procesi obučavanja i testiranja obučanih modela obavljani su uz pomoć sistema za distribuiranu obradu velikih količina podataka *Apache Spark*, kojoj se pristupilo putem programske biblioteke *Sparklyr*, namenjene radu sa programskim jezikom R.

U evaluaciji performansi obučanih modela korišćene su sledeće metrike [8]:

- tačnost (engl. *Accuracy*),
- odziv (engl. *Recall*),
- preciznost (engl. *Precision*) i
- mera F1 (engl. *F1 measure*).

Navedene metrike računane su na osnovu matrice konfuzije formirane za svaki obučeni model.

5. REZULTATI

Formiranjem modela za sve moguće kombinacije grupa prediktorskih obeležja i vrednosti odgovarajućih hiperparametara postignut je broj od 168 formiranih modela, od toga 63 modela za podatke prikupljene na teritoriji Republike Srbije i 105 modela za podatke prikupljene na teritoriji Sjedinjenih Američkih Država. Grupe prediktorskih obeležja označavane su skraćenicama navođenim u zagradama, prema sledećoj podeli: geolokacioni atributi – (L), administrativni atributi – (A), meteorološki atributi – (V), svi dostupni prediktorski atributi – (ALL). Dodatno je uvedena grupa

prediktorskih obeležja označena sa (D) i ona predstavlja datumsko-vremenske atribute.

Pokazalo se da maksimalni broj iteracija koji može da bude dostignut tokom procesa obučavanja ne pravi uticaj na modele formirane postupkom *logističke regresije*, gde je razlika između vrednosti evaluacionih metrika za modele kojima je bilo dozvoljeno samo 50 iteracija i one kojima je bilo dozvoljeno 600 iteracija neznatna ili, u određenom broju slučajeva, nepostojeća. Geološki podaci su uzeti u obzir u gotovo identičnoj meri kao podaci administrativne prirode, poput detaljnog opisa saobraćajne nesreće ili prisustva saobraćajnih elemenata u blizini događanja udesa. Na osnovu vremenskih uslova modeli formirani *logističkom regresijom* nisu uspeli da precizno odrede ozbiljnost saobraćajne nesreće i uzeti su u obzir u vidno manjoj meri u odnosu na geološke i administrativne atribute. Kombinacijom te dve grupe atributa i datumsko-vremenskih atributa postignuti su bolji rezultati, da bi dostigli vrhunac performansi upotrebom svih dostupnih prediktorskih obeležja.

Modeli obučavani algoritmom *stabla odlučivanja* formirani su za različite maksimalne dubine koje stablo može da dostigne u toku procesa obučavanja. Zapaženo je da različite dubine stabla imaju različit uticaj na ponašanje modela, pri čemu se vrednost mere F1 povećavala kako se dubina stabla kretala ka maksimalnoj zadatoj vrednosti. Geološki i administrativni atributi su uzimani u meri sličnoj kao što je to slučaj kod modela formiranih *logističkom regresijom*. Razlika je napravljena pri upotrebi atributa vezanih za vremenske uslove, gde *stablo odlučivanja* postiže znatno bolje rezultate u odnosu na *logističku regresiju*. Upotreba svih dostupnih prediktorskih obeležja i u ovom slučaju dovodi do najboljih i najpreciznijih modela.

Formiranje modela još jednim algoritmom zasnovanim na stablima odlučivanja, *nasumičnim šumama*, rezultiralo je ishodom veoma sličnim onom koji je zabeležen kod modela formiranih postupkom *logističke regresije*. Performanse formiranih modela se nisu međusobno razlikovale u bitnoj meri, iako se inicijalni broj upotrebljenih stabala bitno razlikovao od krajnje vrednosti upotrebljene za dati hiperparametar. Sami rezultati su bolji u odnosu na *logističku regresiju* ukoliko se posmatra američki skup podataka, a lošiji ukoliko se posmatra srpski skup podataka. Kao što je to slučaj sa *stablom odlučivanja*, *nasumične šume* su pokazale bolje performanse za meteorološke atribute u odnosu na *logističku regresiju*, ali se najbolji rezultati i dalje postižu upotrebom svih dostupnih prediktorskih obeležja.

Modeli formirani uz pomoć algoritma *stabla pojačana gradijentom* pokazali su se kao najbolji i zabeležili su najbolje vrednosti evaluacionih metrika. U pitanju su modeli čije se performanse međusobno vidljivo razlikuju u odnosu na podatak o tome koliko je maksimalno iteracija bilo dozvoljeno u okviru procesa obučavanja. Za većinu grupa prediktorskih obeležja ovi modeli ne postižu rezultate koji odstupaju od rezultata koje postižu modeli obučeni prethodno navedenim algoritmima, ali prave razliku kada se radi o upotrebi svih dostupnih prediktorskih obeležja, pogotovo kada je reč o podacima prikupljenim na teritoriji Sjedinjenih Američkih Država. U tabelama 1. i 2. prikazani su, redom, rezultati evaluacije

modela i vrednosti evaluacionih metrika za podatke prikupljene na teritoriji Sjedinjenih Američkih Država i Republike Srbije, pri čemu su u tabelama korišćene sledeće oznake:

- Grupa – naziv grupe prediktorskih atributa,
- LogR – logistička regresija,
- DT – stablo odlučivanja,
- RF – nasumične šume i
- GBT – stabla pojačana gradijentom.

Metrike evaluacije performansi modela označene su sledećim abrevijacijama: tačnost – A, odziv – R, preciznost – P, mera F1 – F1. U tabelama su prikazani najbolji rezultati za konkretne kombinacije grupa prediktorskih obeležja i algoritama mašinskog učenja.

Tabela 1. Rezultati za podatke iz Sjedinjenih Američkih Država

Grupa	LogR	DT	RF	GBT
(L)	A: 0,66017 R: 0,68314 P: 0,65452 F1: 0,66852	A: 0,65457 R: 0,63169 P: 0,66356 F1: 0,64723	A: 0,65588 R: 0,62008 P: 0,66952 F1: 0,64385	A: 0,66254 R: 0,68464 P: 0,65706 F1: 0,67057
(A)	A: 0,67361 R: 0,71321 P: 0,66218 F1: 0,68675	A: 0,67769 R: 0,66861 P: 0,68244 F1: 0,67545	A: 0,66761 R: 0,63146 P: 0,68227 F1: 0,65588	A: 0,67826 R: 0,67219 P: 0,6819 F1: 0,67701
(V)	A: 0,5566 R: 0,5769 P: 0,55594 F1: 0,56623	A: 0,60895 R: 0,67344 P: 0,59786 F1: 0,6334	A: 0,61459 R: 0,66893 P: 0,60473 F1: 0,63521	A: 0,62768 R: 0,66492 P: 0,62023 F1: 0,6418
(L, A, D)	A: 0,70354 R: 0,71912 P: 0,69871 F1: 0,70877	A: 0,70558 R: 0,70557 P: 0,70694 F1: 0,70625	A: 0,70413 R: 0,71188 P: 0,70235 F1: 0,70708	A: 0,7248 R: 0,72775 P: 0,72477 F1: 0,72626
(ALL)	A: 0,70468 R: 0,72207 P: 0,69911 F1: 0,7104	A: 0,73657 R: 0,75337 P: 0,7301 F1: 0,74155	A: 0,73368 R: 0,75355 P: 0,72596 F1: 0,7395	A: 0,7676 R: 0,79136 P: 0,75655 F1: 0,77356

Tabela 2. Rezultati za podatke iz Republike Srbije

Grupa	LogR	DT	RF	GBT
(L)	A: 0,63559 R: 0,60322 P: 0,6468 F1: 0,62425	A: 0,63485 R: 0,62163 P: 0,64025 F1: 0,6308	A: 0,63191 R: 0,58277 P: 0,64821 F1: 0,61375	A: 0,64674 R: 0,58281 P: 0,67021 F1: 0,62346
(A)	A: 0,68578 R: 0,56743 P: 0,74561 F1: 0,64443	A: 0,6858 R: 0,56768 P: 0,74549 F1: 0,64455	A: 0,658 R: 0,38009 P: 0,86053 F1: 0,52728	A: 0,66303 R: 0,65362 P: 0,66782 F1: 0,66064
(ALL)	A: 0,7318 R: 0,73502 P: 0,73173 F1: 0,73337	A: 0,73239 R: 0,76287 P: 0,72036 F1: 0,74101	A: 0,70745 R: 0,6942 P: 0,71466 F1: 0,70428	A: 0,74584 R: 0,72839 P: 0,75617 F1: 0,74202

Nad modelima koji su zasnovani na stablima odlučivanja primenjen je postupak određivanja važnosti atributa, čime se ispitalo koji atributi i njihove konkretne vrednosti imaju uticaj na dalje ponašanje obučenog modela.

Modeli formirani putem algoritma *stabla odlučivanja* nad podacima za Republiku Srbiju bitnim atributima smatraju geološke podatke, poput geografskih koordinata mesta događanja nesreće ili naziva opštine, policijske uprave ili savezne države u okviru čije teritorije se nesreća dogodila. Manji uticaj sa strane imaju faktori koji se tiču konkretnog opisa nezgode ili prisustva određenih saobraćajnih elemenata. Modeli formirani nad podacima za Sjedinjene Američke Države dodatno obraćaju pažnju na različite administrativne i datumsko-vremenske

podatke, poput naziva godišnjeg doba ili prisustva semaforizovanog pešačkog prelaza u blizini događanja nesreće.

Dato ponašanje dele i modeli formirani uz pomoć algoritama *nasumične šume* i *stabla pojačana gradijentom*, sa napomenom da modeli bazirani na *stablima pojačanim gradijentom* uzimaju u obzir meteorološke podatke u većoj meri u odnosu na modele zasnovane na preostalim algoritmima. Pokazalo se da je za određivanje ozbiljnosti nesreće uziman u obzir i podatak o tome da li je postojao sudar i, ako jeste, da li je u pitanju sudar sa parkiranim vozilom ili vozilom u kretanju. Dodatno je utvrđeno da značajan uticaj na ponašanje modela ima podatak o dužini puta na koju je nesreća direktno uticala, što se manifestuje, na primer, izazivanjem zastoja u saobraćaju.

6. ZAKLJUČAK

Ovim radom je predstavljena analiza različitih faktora koji mogu da utiču na ozbiljnost saobraćajne nesreće i samim tim na bezbednost učesnika u saobraćaju. Istraživanje je sprovedeno nad većim količinama podataka prikupljenim u toku dužeg vremenskog perioda. Iz navedenog razloga su upotrebljene tehnologije iz oblasti računarstva visokih performansi, na osnovu kojih su dalje sprovedeni procesi vezani za oblast mašinskog učenja. Prediktivni modeli formirani su na osnovu različitih algoritama baziranih na stablima – *stablo odlučivanja*, *nasumične šume*, *stabla pojačana gradijentom* – i regresionog algoritma *logistička regresija*.

Iz dobijenih rezultata se može zaključiti više stvari vezanih za uticaj različitih fenomenoloških grupa na mogućnost izazivanja saobraćajnih nezgoda sa teškim posledicama po učesnike. Pre svega treba naglasiti da se na osnovu evaluacije formiranih modela može zaključiti da nijedna zasebna grupa parametara, poput geolokacionih podataka ili informacija administrativne prirode, ne može služiti kao apsolutno pouzdani oslonac za određivanje mogućnosti događanja nesreće sa posledicama težeg oblika. Jasno je da postoji određeni uticaj lokacije vozila-učesnika, strukture dela kolovoza na kome se nesreća odvija, kao i različitih datumsko-vremenskih odrednica, ali on sam po sebi nije dovoljan za tačno i sigurno određivanje težine same nesreće, čime postaje jasno da se bezbednost u saobraćaju ne može poboljšati posmatranjem samo jedne grupe parametara. Sa druge strane, ukoliko se prethodno navedene grupe parametara ujedine i formiraju se modeli na osnovu takve unije, rezultati se poboljšavaju i dolazi se do mogućnosti nešto sigurnijeg određivanja stepena ozbiljnosti udesa.

Na osnovu podataka koji su korišćeni i modela koji su formirani zaključeno je da se tokom procesa obučavanja određeni obučavajući parametri smatraju bitnijima od ostalih, pa se tako geolokacioni podaci i jedan deo administrativnih podataka ističu u odnosu na, na primer, različite meteorološke podatke, poput vlažnosti vazduha ili brzine vetra.

7. LITERATURA

- [1] World Health Organization, “Road traffic injuries”, World Health Organization (WHO), Jun 2021. Dostupno na adresi: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries> (pristupljeno u avgustu 2021.)
- [2] S. Moosavi, M. H. Samavatian, S. Parthasarathy, R. Teodorescu & R. Ramnath, “Accident risk prediction based on heterogeneous sparse data: New dataset and insights”, Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2019.
- [3] S. Moosavi, M. H. Samavatian, S. Parthasarathy & R. Ramnath, “A Countrywide Traffic Accident Dataset”, arXiv preprint arXiv:1906.05409, 2019.
- [4] C. Parra, C. Ponce & S. F. Rodrigo, “Evaluating the Performance of Explainable Machine Learning Models in Traffic Accidents Prediction in California”, 2020 39th International Conference of the Chilean Computer Science Society (SCCC), pp. 1-8, doi: 10.1109/SCCC51225.2020.9281196, 2020.
- [5] “Подаци о саобраћајним незгодима по ПОЛИЦИЈСКИМ УПРАВАМА и ОПШТИНАМА (СН) - Отворени подаци”, Data.gov.rs, 2021. Dostupno na adresi: <https://data.gov.rs/sr/datasets/podatsi-o-saobratshajnim-nezgodama-po-politsijskim-upravama-i-opshtinama/> (pristupljeno u avgustu 2021)
- [6] S. Shalev-Shwartz & S. Ben-David, “Understanding Machine Learning: From Theory to Algorithms”, Cambridge: Cambridge University Press. doi:10.1017/CBO9781107298019, 2014.
- [7] N. Lunardon, G. Menardi & N. Torelli, “ROSE: a Package for Binary Imbalanced Learning”, The R Journal, vol. 6, no. 1, p. 79, 2014. Available: 10.32614/rj-2014-008.
- [8] A. Zheng, “Evaluating Machine Learning Models”, O’Reilly Media, Inc., Sep. 2015.

Kratka biografija:



Jovan Vunić rođen je u Novom Sadu 1997. godine. Završio je gimnaziju „Jovan Jovanović Zmaj” u Novom Sadu, prirodno-matematički smer. Fakultet tehničkih nauka, studentski program Računarstvo i automatika, upisao je 2016. godine. Nakon završenih osnovnih akademskih studija, upisao je master akademske studije na istom studentskom programu.