



ODREĐIVANJE KVALITETA POJEDINAČNIH JELA IZ RECENZIJA RESTORANA  
UPOTREBOM SENTIMENT ANALIZE

DETERMINATION OF QUALITY OF INDIVIDUAL MENU ITEMS FROM  
RESTAURANT REVIEWS USING SENTIMENT ANALYSIS

Marija Joksimović, *Fakultet tehničkih nauka, Novi Sad*

**Oblast – ELEKTROTEHNIKA I RAČUNARSTVO**

**Kratak sadržaj** – U ovom radu predstavljeno je jedno rešenje za izdvajanje delova teksta iz korisničkih recenzija koji predstavljaju nosilac sentimenta za određeno jelo i klasifikaciju tih delova teksta prema osećanju koje je u njima iskazano na pozitivno, negativno ili neutralno korišćenjem metoda mašinskog učenja. Segmenti teksta recenzija koji sadrže pominjanje hrane su formirani na osnovu leksičkih veza između reči i na njih su primenjene određene tehnike pretprocesiranja. Potom je izvršena sentiment analiza upotrebom nekoliko modela mašinskog učenja. Svi podaci su prikupljeni sa sajta Donesi.com i ručno anotirani. Korišćeni modeli su evaluirani.

**Ključne reči:** analiza teksta, klasifikacija teksta, procesiranje prirodnog jezika, sentiment analiza, recenzije restorana

**Abstract** – This paper presents one approach for extraction of parts of restaurant reviews which contain information about opinion of certain menu item and classification of those segments by expressed sentiment as positive, negative or neutral using several machine learning algorithms. Text segments which contain food mentions were generated using lexical relationships between words in reviews and several preprocessing techniques were applied. Afterwards, sentiment analysis was done using several machine learning models. Data was acquired from the website Donesi.com and manually annotated. All used models were evaluated.

**Keywords:** text mining, text classification, natural language processing, sentiment analysis, restaurant reviews

**1. UVOD**

Ljudi prilikom donošenja odluka, posebno kada je u pitanju kupovina određenog proizvoda ili usluge, konsultuju mišljenje drugih ljudi pre nego što donesu svoj konačni sud.

Pre ekspanzije interneta pojedinci su bili ograničeni na porodicu i prijatelje ukoliko im je bila potrebna pomoć prilikom donošenja odluke, a kompanije su bile primorane da sprovedu različite studije i ankete kako bi prikupili podatke o zadovoljstvu svojih klijenata.

**NAPOMENA:**

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Aleksandar Kovačević, vanr. prof.

Danas to više nije slučaj. Informacije od značaja su sadržane u korisničkim recenzijama i komentarima na veb stranicama onih koji proizvod nude i najčešće su javno dostupni. Jedan slučaj prethodno opisanog scenarija jesu recenzije na sajtovima za preporuku restorana.

Većina takvih sajtova korisnicima omogućuje pretragu najbolje rangiranih restorana iz određene kategorije, ali se retko sreće slučaj preporuke određenih stavki jelovnika u okviru restorana.

Tema ovog rada jeste izdvajanje delova teksta iz korisničkih recenzija koji predstavljaju nosilac sentimenta za određeno jelo i klasifikacija tih delova teksta prema osećanju koje je u njima iskazano na pozitivno, negativno ili neutralno korišćenjem metoda mašinskog učenja. Obzirom da srpski jezik spada u leksički složene jezike i pošto je čest je slučaj da se u okviru iste recenzije spominje više jela gde svako od njih može biti opisano u drugačijem svetlu, najveći izazov jeste izdvojiti okolinu oko datog pominjanja hrane tako da obuhvati samo jedno jelo i kontekst u okviru kog se ono spominje.

Čitav proces dodatno otežava činjenica da recenzije korisnika spadaju u kraće, neformalne tekstove, u kojima se često koriste žargonski izrazi i sadrže šum. Analiza sentimenta kao vid klasifikacije teksta je poslednjih nekoliko godina sve češća tema naučnih istraživanja.

Osnovne razlike leže u izboru domena, definisanju osnovne jedinice dokumenta za koju će se određivati sentiment, izboru tehnika pretprocesiranja teksta i algoritama mašinskog učenja koji će se primeniti.

Tekstovi sadržani u domenima mogu biti duži i često formalni tekstovi poput novinskih članaka, književnih radova i blogova, ili kraći, neformalni tekstovi, često puni šuma poput online recenzija i postova sa društvenih mreža.

Skup podataka formiran je od recenzija preuzetih sa sajta Donesi.com [1], jednog od najpopularnijih sajtova za naručivanje hrane preko interneta na teritoriji Srbije, Crne Gore i Bosne i Hercegovine.

Od svih recenzija koje su preuzete, nasumično su izdvojene recenzije pisane na srpskom jeziku koje će ući u test i trening skupove. Delovi teksta recenzija u kojima je iskazano mišljenje o nekom jelu su izdvojeni na osnovu leksičkih veza između reči upotrebom biblioteke za procesiranje prirodnih jezika (eng. *Natural Language Processing - NLP*) specijalizovane za srpski jezik i vršena je sentiment analiza na nivou entiteta.

Korišćena su ukupno četiri algoritma nadgledanog mašinskog učenja za klasifikaciju: Naive Bayes, SVM, Random Forest i logistička regresija.

## 2. METODOLOGIJA

U ovom poglavlju predstavljena je metodologija određivanja delova teksta recenzija u kojima je iskazano mišljenje o nekoj hrani i klasifikacija tih delova prema sentimentu koji nose na pozitivan, negativan ili neutralan. Detaljno će biti predstavljen čitav postupak sa svim koracima i alatima koji su implementirani i korišćeni tokom realizacije.

### 2.1. Prikupljanje podataka

Koliko je autor rada upoznat, ne postoji javno dostupan skup podataka koji sadrži recenzije restorana koje su pisane na srpskom jeziku. Svi neophodni podaci su preuzeti sa Donesi.com. Pošto sajt ne poseduje API za preuzimanje podataka napisan je specijalizovani web crawler u programskom jeziku Python. Za implementaciju crawler-a korišćena je biblioteka Selenium [2], a za parsiranje HTML stranica biblioteka BeautifulSoup4 [3]. Odlučeno je da se dobave svi podaci vezani za područje Srbije. Dobavljeno je ukupno 169486 recenzija za 1658 različita restorana.

### 2.2. Anotacija pominjanja hrane u recenzijama

Neophodno je bilo označiti tačna mesta u tekstu gde se je javila pominjanja hrane. Kako bi bile anotirane samo recenzije koje sadrže barem jedno pominjanje hrane u naslovu ili u konkretnom tekstu napravljen je jednostavan program. Recenzije su učitavane sukcesivno i prikazivane anotatoru koji je označavao da li trenutna recenzija sadrži pominjanje hrane ili ne. Ako je odgovor bio potvrđan, recenzija bi se sačuvala u novu kolekciju. Manuelno je anotirano 10000 recenzija u kojima su anotatori označavali svaku reč ili grupu reči koja predstavlja pominjanje hrane.

### 2.3. Proširivanje pominjanja hrane iskazanim mišljenjem

Tekstovi recenzija su prvo podeljeni u rečenice i svaka rečenica je tokenizovana. Za svaki token ubeležen je i njegov položaj u originalnom tekstu. Iako se za analizu sentimenta nije korišćena cela recenzija, POS tagovanje, izvlačenje povezanih reči i lematizacija su izvršeni u ovom koraku kako bi ostala očuvana leksička veza između reči. Originalni tekst recenzije nije izmenjen, samo su zasebno sačuvani nizovi tokena, njihovih POS tagova i lema.

Korišćena je ReLDI [4-8] biblioteka namenjena za procesiranje tekstova na srpskom, hrvatskom i slovenačkom jeziku. Svi izlazi koje daje ReLDI biblioteka su u JSON formatu i nakon procesiranja su povezani sa odgovarajućim objektima recenzija koji su prethodno sačuvani. Svi tokeni koji pripadaju nekom pominjanju hrane su označeni parsiranjem rezultata anotiranja iz prethodnog koraka. Tabelarni prikaz izlaza procesiranja rečenice pomoću ReLDI prikazan je u tabeli 1.1.

Reči koje predstavljaju hranu ili imena jela su najčešće imenice. Imena hrane ili delova jela se sastoje iz jedne ili više imenskih reči kao što su : pizza, pasta sos, hleb, paradajz itd. Nazivi jela se mogu sastojati i iz drugih vrsta reči koje bliže opisuju sastojke ili poreklo određenog jela,

poput „Bečka šnicla“ ili „piletina u slatko-kiselom Sečuan sosu“. Čest je slučaj da se u nazivima jela javi ime restorana, poput „The Saint Pizza“. Takođe, postoji dosta restorana strane kuhinje u kojima neka ili sva jela iz jelovnika nose ime na izvornom jeziku. Korisnici u recenzijama hranu spominju prema sopstvenom nađenju. Nazivi jela su retko napisani onako kako su navedeni u jelovniku.

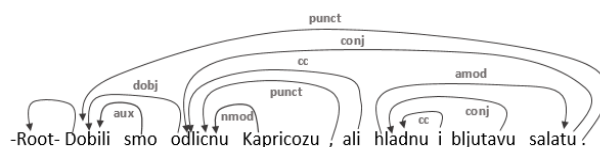
Najčešće se koriste skraćeni ili prevedeni nazivi jela (Quattro Formagi Pizza – Pica 4 sira) i fonetska transkripcija (Quattro Formagi Pizza – Kvatro formadi pica). Čest je i slučaj korišćenja kovanica i gramatički neispravnih oblika reči. Sve ovo za posledicu ima javljanje grešaka prilikom automatskog određivanja vrste reči za svaki token tokom POS tagovanja, lematizacije i formiranja stabla zavisnosti.

Isto tako, korisnici često ne iskazuju mišljenje direktno o nekom jelu, nego sentiment iskazuju kritikom neke komponente tog jela, poput primera „Salata sa groznim, zagorelim pečurkama“. Glavni izazov u ovom koraku jeste definisati algoritam uparivanja i skupa dozvoljenih veza između reči unutar recenzije koje su povezane sa datim pominjanjem hrane tako da se uhvati sentiment vezan samo za jedno jelo.

Tabela 1. *Procesiranje rečenice pomoću ReLDI biblioteke*

ID	Token	POS tag	Lema	Dep-parse
1	Dobili	Vmp-pm	dobiti	0/root
2	smo	Var1p	biti	1/aux
3	odlicnu	Ncfsa	odlicna	1/dobj
4	Kapricozu	Npmsl	Kapricoz	3/nmod
5	,	Z	,	3/punct
6	ali	Cc	ali	3/cc
7	hladnu	Agofsay	hladan	10/amod
8	i	Cc	I	7/cc
9	bljutavu	Agpfsay	bljutav	7/conj
10	salatu	Ncfsa	salata	3/conj
11	.	Z	.	1/punct

Parsiranje zavisnosti se vrši na nivou rečenice. Parser zavisnosti iz biblioteke ReLDI [7] je sposoban da prepozna 39 različitih tipova veza. Svaka veza se sastoji iz dva tokena i naziva tipa veze. Jedan token iz veze predstavlja „glavu“ (eng. *head*), dok je drugi zavisni token u vezi. Jedan token iz rečenice nema svog para, predstavlja koren i označen je vezom „root“. Iako izlaz iz parsera zavisnosti jeste niz objekata veza, rekursivnim obilaskom tog niza se dobija stablo zavisnosti tokena u rečenici. Svi tokeni osim korena se u vezama mogu javiti proizvoljan broj puta. Jedan primer izgleda stabla zavisnosti prikazan je na slici 1.



Slika 1. *Stablo zavisnosti*

Lista tipova veza koje su uzete u obzir je dobijena empirijski. Obično su korišćene veze *amod*, *admod*, i *nsubj*. Na taj način je jedino moguće proširiti pominjanje hrane sa jednom rečju ili frazom koja ga šire opisuje.

Zbog prirode problema odlučeno je da se koristi više veza kako bi bilo moguće obraditi i slučajeve gde je sentiment iskazan za određeni deo jela. Isprobani su različiti skupovi tipova veza i odlučeno je da se u finalnom skupu nađe osam : *advmod* (eng. *adverbial modifier*), *amod* (eng. *adjectival modifier*), *acl* (eng. *adjectival clause*), *nmod* (eng. *nominal modifier*), *nsubj* (eng. *nominal subject*), *neg* (eng. *negation modifier*), *nummod* (eng. *numeric modifier*), *dojb* (eng. *direct object*). Opise tipova veza je moguće pronaći na [10].

Za automatsko izvlačenje segmenata koji sadrže pominjanje hrane i mišljenje vezano za njih je napisan program. Obradeno je svih 10000 anotiranih recenzija. Za svaku recenziju preuzet je skup svih pominjanja hrane. Svako pominjanje hrane je obrađeno zasebno. Inicijalni skup tokena koji predstavlja izlazni segment su činili tokeni datog pominjanja hrane. Provera koji se tokeni iz recenzije odnose na druge tokene segmenta obavljao rekurzivno. Za svaku vezu iz liste zavisnosti provereno je da li postoji veza gde su zadovoljene sledeće stavke :

- Neki token iz trenutnog skupa tokena koji čine segment ima ulogu glave
- Podređeni token se ne nalazi u trenutnom skupu
- Tip veze pripada skupu validnih tipova veza

Ukoliko su sva tri uslova zadovoljena token bi bio dodat u skup tokena za narednu obradu. Kada bi bile ispitane sve veze za trenutni skup tokena, ponovo bi bio pokrenut proces obrade, ali ovaj put ulaz bi bio skup generisan u prethodnoj iteraciji. Čitav proces bi se rekurzivno ponavljao dok se u jednoj iteraciji ni jedan token ne nađe u skupu za narednu obradu.

Nakon toga, skup tokena bi bio pretvoren u listu sortiranu rastuće prema poziciji tokena u tekstu. Iz originalnog teksta recenzije bio bi isečen deo između početka prvog tokena u listi i kraja poslednjeg. To je učinjeno kako bi ostao očuvan smisao. Tokeni i leme svih reči, bez obzira da li su pripadali izlaznom skupu tokena ili su naknadno dodati nakon isecanja, zasebno su sačuvani u novu kolekciju. Sačuvani su samo segmenti gde je pominjanje hrane prošireno barem jednom dodatnom rečju.

## 2.4. Obučavajući i test skup podataka

Inicijalni skup podataka se sastojao iz 10113 segmenata koje je bilo neophodno anotirati prema polaritetu sentimenta koji iskazuju. Kreiran je namenski program za anotiranje čiji je interfejs prikazan na slici 2. Anotatoru bi bio prikazan tekst segmenta i sentiment bi se označavao celobrojnom vrednošću 1 za pozitivan, 2 za negativan i 3 za neutralan sentiment. Izgled alata je predstavljen na slici 2.

```
-----
Obradjuje se : 5d530f3b98a01312309288e2  redni broj: [ 503 ]
Tekst :      Tetso fino i tanko, bas kako treba
[ pozitivno -> 1 | negativno -> 2 | neutralno -> 3 | EXIT -> 0 ]
Oznacite sentiment za prikazani tekst : 1
-----
```

Slika 2. Interfejs alata za anotiranje

Anotirani skup nije bio balansiran. Sadržao je 3466 segmenta sa pozitivnim, 1245 sa negativnim i 5224 sa neutralnim sentimentom. Odlučeno je da se iskoriste svi segmenti sa negativnim sentimentom i da se iz preostala

dva skupa nasumice izabere po 1245 segmenata. Prema tome, kreiran je balansiran skup podataka u kom su se nalazila ukupno 3735 segmenta. Svaki segment je prvo pretprocesiran i potom transformisan u vektor reči kako bi bilo moguće primeniti algoritme mašinskog učenja za klasifikaciju.

## 2.5. Pretprocesiranje segmenata

U segmentima su se često javljali šablonski delovi koji nikada ne nose sentiment, poput nabiranja koliko je kog jela poručeno i po kojoj ceni („2x Kapricioza 32cm“). Ovaj šablon se mogao predstaviti regularnim izrazom i delovi teksta koji su se slagali sa njim su zamenjeni specijalnom oznakom. Postoje i šabloni za koji je polaritet veoma teško odrediti, poput ocenjivanja jela po nestandardizovanoj skali („5/5“ ili „2/10“) koji su takođe zamenjeni. Dobar indikator sentimenta predstavlja polaritet emotikona koji se javlja uz pominjanje hrane. Formirana su dva regularna izraza za zamenu emotikona, jedan za pozitivne i jedan za negativne. Primećeno je da su se u segmentima često javljale reči napisane sa višestrukim ponavljanjem nekog samoglasnika kako bi se dočarao intenzitet iskazanog mišljenja.

Višestruka ponavljanja su ostala zabeležena, ali se broj ponavljanja tog karaktera redukovao na dva. Isto tako, srpski jezik je veoma morfološki složen jezik. Morfološka normalizacija je uključena kako bi se različiti oblici iste reči sveli na istu osnovu.

Primenjene su dve tehnike normalizacije: stemming i lemantizacija. Korišćene su Python implementacije stemera [9] i lemantizera iz ReLDI biblioteke. Morfološka normalizacija je imala veliki doprinos u redukciji dimenzionalnosti problema. Inicijalni skup atributa na kom nisu primenjene tehnike pretprocesiranja brojao je 10358 atributa. Primenom lemantizacije skup atributa je redukovao na 8832 atributa, a stemminga na 8609. Svaki segment je predstavljen kao vektor atributa (eng. *feature vector*). Cilj je predstaviti svaku normalizovanu reč ulaznog segmenta određenom brojčanom vrednošću, kako bi se formirani vektor mogao koristiti kao ulaz ili izlaz modela mašinskog učenja. Korišćene su tri šeme formiranja vektora atributa : binarna vektorizacija, frekvencija terma i tf-idf.

## 3. EKSPERIMENTALNA EVALUACIJA I REZULTATI

Eksperimentalna evaluacija se vrši radi određivanje performansi korišćenih algoritama mašinskog učenja za klasifikaciju teksta prema iskazanom sentimentu. Metrike performansi algoritama korišćene u ovom radu su tačnost, preciznost, odziv i  $F_1$ -mera.

Pojedini parametri svakog od četiri korišćena modela mašinskog učenja su optimizovani. Optimizacija je izvršena kros-validiranom grid-pretragom (eng. *cross-validated grid-search*) po koordinatnoj mreži parametara koji se optimizuju. Upotrebljena je implementacija GridSearchCV [11] iz Python biblioteke Scikit-learn. Vršena je kros validacija iz  $k$  delova (eng. *k-fold cross-validation*). Zasebno su optimizovani parametri za svaku kombinaciju izbora šeme kodiranja atributa, šeme selekcije atributa i tehnike normalizacije i te vrednosti su korišćene u daljoj evaluaciji.



Za Naive Bayes [12] klasifikator je vršena optimizacija parametra  $\alpha$  ( $\alpha$ ) i predstavlja korektivni faktor. Najbolji rezultat tačnosti, preciznosti, odziva i F1 mere je postignut za slučaj gde je korišćen stemming, tf-idf šema kodiranja ulaza i unigrami kao šema selekcije atributa. Postignuti su rezultati od 0,7831, 0,7891, 0,7831 i 0,7816 respektivno za  $\alpha = 0,6$ . Za SVM vršena je optimizacija vrednosti parametra  $C$ . On predstavlja parametar penalizacije greške klasifikacije. Generalno najbolje rezultate klasifikacije je imao SVM [13] klasifikator. Najbolji rezultat tačnosti, preciznosti, odziva i F1 mere je postignut za slučaj gde je korišćena lemantizacija, frekvencija terma kao šema kodiranja ulaza i unigrami kao šema selekcije atributa.

Postignuti su rezultati od 0,8046, 0,8122, 0,8046 i 0,8059 respektivno za  $C = 0,6$ . Za Random Forest [14] klasifikator optimizovani su parametri broja stabala ( $n_{estimators}$ ) i procenat atributa originalnog skupa podataka će preuzeti svaki estimator ( $max\_features$ ). Random Forest je pokazao najlošije rezultate klasifikacije od svih klasifikatora. Najbolji rezultat tačnosti, preciznosti, odziva i F1 mere je postignut za slučaj gde je korišćen stemming, frekvencija terma kao šema kodiranja ulaza i unigrami kao šema selekcije atributa. Postignuti su rezultati od 0,7617, 0,7901, 0,7617 i 0,7614 respektivno za  $n_{estimators} = 70$  i  $max\_features = 0,9$ . Iako se logistička regresija [15] retko koristi za klasifikaciju teksta, u ovom slučaju su dobijeni veoma zadovoljavajući rezultati. Performanse mogu da pariraju onima koji su postignuti korišćenjem SVM-a. Za klasifikator zasnovan na logističkoj regresiji vršena optimizacija parametara  $C$  i  $solver$ . Najbolji rezultati su postignuti upotrebom lemantizacije, tf-idf šeme kodiranja ulaza i unigrama kao šeme selekcije atributa. Postignuti su rezultati od 0,8005, 0,8056, 0,8005 i 0,8021 respektivno za  $solver = lbfgs$  i  $C = 1,0$ .

#### 4. ZAKLJUČAK

U ovom radu predstavljeno je jedno rešenje za izdvajanje delova teksta iz korisničkih recenzija koji predstavljaju nosilac sentimenta za određeno jelo i klasifikacija tih delova teksta prema osećanju koje je u njima iskazano na pozitivno, negativno ili neutralno korišćenjem metoda mašinskog učenja. Skup podataka se sastojao iz recenzija restorana pisanih na srpskom jeziku. Izdvojeno je 10000 recenzija u kojima su manuelno anotirana sva pominjanja hrane. Sve recenzije su tokenizovane i POS-tagovane. Izvršena je lemantizacija i parsiranje zavisnosti. Segmenti recenzija koji predstavljaju nosilac sentimenta za određeno jelo izdvojeni su na osnovu leksičkih veza i skupa pravila koje je svaka veza treba da zadovolji.

Test i trening skup podataka su formirani iz segmenata u kojima su pominjanja hrane proširena barem jednim tokenom. Sentiment svakog segmenta je manuelno anotiran. Tehnike pretprocesiranja korišćene u ovom radu su zamene kvantitativnih i kvalitativnih numeričkih vrednosti i emotikona. Pored lemantizacije, razmatran je i stemming kao tehnika morfološke normalizacije teksta.

Kao šema kodiranja atributa korišćeni su binarna, frekvencija terma i tf-idf. Razmatrane šeme selekcije atributa su unigrami i unigrami kojima su pridruženi bigrami. Za klasifikaciju su korišćena ukupno četiri algoritma nadgledanog mašinskog učenja: Naive Bayes, SVM, Random

Forest i logistička regresija. Rad bi mogao da se proširi u dva aspekta. Prvi se tiče uklanjanja šuma i poboljšavanje tehnika pretprocesiranja, poput ponalaženja i zamene pogrešno napisanih reči validnim oblikom iste i dodavanja rečnika sinonima.

Drugi aspekt tiče se detaljnije optimizacije parametara klasifikatora u cilju postizanja boljih performansi. Generalno, rezultate je moguće unaprediti proširivanjem skupa podataka anotiranjem više recenzija.

#### 5. LITERATURA

- [1] [www.donesi.com](http://www.donesi.com)
- [2] Jason Huggins, *et al*, 2004. Selenium, <https://www.seleniumhq.org>
- [3] Leonard Richardson 2014, BeautifulSoup4 <https://www.crummy.com/software/BeautifulSoup>
- [4] Ljubecic, Nikola, Tomaz Erjavec and Darja Fiser. "Corpus-Based Diacritic Restoration for South Slavic Languages." *LREC* (2016).
- [5] Ljubecic, Nikola and Tomaz Erjavec. "Corpus vs. Lexicon Supervision in Morphosyntactic Tagging: the Case of Slovene." *LREC* (2016).
- [6] Ljubecic, Nikola, Filip Klubicka, Zeljko Agic and Ivo-Pavao Jazbec. "New Inflectional Lexicons and Training Corpora for Improved Morphosyntactic Annotation of Croatian and Serbian." *LREC* (2016).
- [7] Agic, Zeljko and Nikola Ljubecic. "Universal Dependencies for Croatian (that work for Serbian, too)." *BSNLP@RANLP* (2015).
- [8] Fišer, D., Ljubešić, N. & Erjavec, T. Lang Resources & Evaluation (2018). <https://doi.org/10.1007/s10579-018-9425-z>
- [9] Milosevic, Nikola "Stemmer for Serbian Language." *CoRR* abs/1209.4471 (2012): n. pag.
- [10] Universal Dependencies, <https://universaldependencies.org/#language-u>
- [11] GridSearchCV, Scikit-learn [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)
- [12] Leung K. M. (2007). "Naive Bayesian Classifier", Polytechnic University, Department of Computer Science, Finance and Risk Engineering.
- [13] Support Vector Machines (SVM), Statsoft, <http://www.statsoft.com/Textbook/Support-Vector-Machines>
- [14] Breiman L. (2001) "Random Forests, Machine Learning", Vol. 45. Issue 1, pp. 5-32.
- [15] Logistic Regression Scikit-learn, [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

#### Kratka biografija:



**Marija Joksimović** rođena je u Sremskoj Mitrovici 1995. god. Master rad na Fakultetu tehničkih nauka iz oblasti Računarstva i automatike – Sistemi za istraživanje i analizu podataka odbranila je 2019.god. kontakt: mariyajoksimovic@hotmail.com