

Postupci za automatsku detekciju lažnih vijesti

Sliško, Mario

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka / Sveučilište u Rijeci**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:195:488456>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-12**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Informatics and Digital Technologies - INFORI Repository](#)



Sveučilište u Rijeci – Fakultet informatike i digitalnih tehnologija

Informacijski i komunikacijski sustavi

Mario Sliško

Postupci za automatsku detekciju lažnih vijesti

Diplomski rad

Mentor: Izv. Prof. dr. sc. Ana Meštrović

Rijeka, prosinac 2022.

Rijeka, 2.6.2022.

Zadatak za diplomski rad

Pristupnik: Mario Sliško

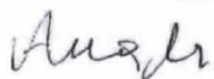
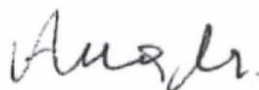
Naziv diplomskog rada: Postupci za automatsku detekciju lažnih vijesti

Naziv diplomskog rada na eng. jeziku: Algorithms for automatic detection of fake news

Sadržaj zadatka: Zadatak diplomskog rada je istražiti i opisati postupke za automatsku detekciju lažnih vijesti. Potrebno je dati pregled postojećih anotiranih skupova podataka koji su prilagođeni za učenje modela za automatsku detekciju lažnih vijesti. U praktičnom dijelu zadatka potrebno je odabrati neke od algoritama strojnog učenja te ih trenirati na nekoliko dostupnih skupova podataka. U radu je potrebno usporediti rezultate koje postižu različiti algoritmi.

Mentor:
Prof. dr. sc. Ana Meštrović

Voditeljica za diplomske radove:
Prof. dr. sc. Ana Meštrović



Zadatak preuzet: 2.6.2022.



(potpis pristupnika)

Sažetak

Tema ovog diplomskog rada je primjena postupaka za automatsku detekciju lažnih vijesti. Rad se sastoji od teorijskog dijela, u kojem se opisuje problem koji predstavlja širenje lažnih vijesti, pružen je pregled na ostale radove koji se bave sličnom tematikom te su opisani alati i algoritmi koji će biti korišteni u radu. Drugi dio rada se odnosi na eksperiment u kojem je opisan postupak razvoja modela za detekciju lažnih vijesti. Opisani su skupovi podataka koji će se koristiti, prikazano je korištenje metoda za pripremu podataka i zatim korišteni algoritmi strojnog učenja. Na kraju rada su prikazani dobiveni rezultati, uspoređeni su sa rezultatima iz drugih radova te je donesen zaključak.

Ključne riječi: *dezinformacije, lažne vijesti, anotirani skup podataka, obrada prirodnog jezika, algoritmi strojnog učenja, točnost, preciznost, odziv, F1, matrica konfuzije*

Abstract

The topic of this master's thesis is the application of algorithms for automatic detection of fake news. The thesis consists of a theoretical part, which describes the problem of the spread of fake news, provides an overview of other papers dealing with a similar topic, and describes the tools and algorithms that will be used in the thesis. The second part of the thesis refers to an experiment in which the process of developing a model for detecting fake news is described. The data sets that will be used are described, all the data preparation methods used and then the machine learning algorithms used are shown. At the end of the work, the obtained results are presented, they are compared with the results from other papers, and a conclusion is drawn.

Keywords: Disinformation, fake news, labelled dataset, natural language processing, machine learning algorithms, accuracy, precision, recall, F1, confusion matrix

Sadržaj

1. Uvod.....	1
2. Problem dezinformacija i pregled dosadašnjih istraživanja	3
2.1. Dezinformacije	3
2.1.1. Primjer širenja dezinformacija: predsjednički izbori u SAD-u	4
2.2. Pregled dosadašnjih istraživanja	7
2.2.1. Pregled postojećih skupova podataka s anotacijama vezanim uz dezinformacije ...	9
2.3. Opis metodologija	12
2.3.1. Obrada prirodnog jezika.....	12
2.3.2. Strojno učenje.....	13
2.4. Opis evaluacije modela	23
3. Eksperiment.....	25
3.1. Python biblioteke.....	25
3.2. Skupovi podataka	25
3.3. Priprema podataka.....	28
3.4. Algoritmi strojnog učenja.....	31
4. Rezultati	34
4.1. Usporedba rezultata s ostalim radovima	36
4.2. Matrice konfuzije za logističku regresiju	38
4.3. Matrice konfuzije za stablo odluka	39
4.4. Matrice konfuzije za slučajne šume	40
4.5. Matrice konfuzije za SVM	41
4.6. Matrice konfuzije za KNN	42
4.7. Matrice konfuzije za Navini Bayes	43
4.8. Matrice konfuzije za pasivno agresivni.....	44
5. Zaključak.....	45
6. Reference.....	47
7. Popis slika	51
8. Popis tablica	52

1. Uvod

Dezinformacije predstavljaju značajan problem s kojima se svakodnevno susreće većina svjetske populacije. To su vijesti koje su nastale korištenjem lažnih informacija prezentiranih kao istinite. Razvoj interneta i društvenih mreža omogućio je stvaranje i rapidno širenje dezinformacija. Svaki korisnik interneta, pa s time i društvenih mreža, ima mogućnost kreiranja takvih vijesti zbog niza različitih razloga [1]. U većini slučajeva društvene mreže ne provjeravaju korisnike kao ni njihove objave. Neki od njih mogu kreirati laže profile s kojih će objavljivati dezinformacije i time imati potpunu slobodu i sigurnost koji ne bi imali objavljujući dezinformacije sa potvrđenih profila na kojima se jasno vidi njihovo pravo lice i kako izgledaju. Veliki broj web stranica se predstavljaju kao vjerodostojni novinski portali te dijele dezinformacije kako bi narušili reputaciju osobe ili udruge. Također postoji financijski aspekt koji je jedan od glavnih razloga kreiranja web stranica za širenje dezinformacija. Svakim klikom na link, web stranica zaradi određenu svotu novca. Iako jedan klik neće donijeti veliko bogatstvo, dezinformacije mogu privući veliku zainteresiranost javnosti, zbog svoje „šokantne“ prirode te time navesti ljude da ih podijele i pridodaju zaradi stranice.

Korištenje umjetne inteligencije, preciznije strojnog učenja, jedan je od načina detektiranja dezinformacija. Obrada prirodnih jezika (engl. Natural language processing, NLP u daljnjem tekstu) je grana umjetne inteligencije koja omogućuje računalima razumijevanje, analizu i upotrebu ljudskog jezika [2]. Pomoću raznih algoritama obrade prirodnih jezika, riječi se pretvaraju u brojeve (vektore) kako bi računalo moglo praviti predviđanja korištenjem algoritama strojnog učenja. Postoji više algoritama strojnog učenja koji se mogu koristiti za detekciju dezinformacija. Algoritmi koji se koriste u svrhu ovog rada su logistička regresija, stablo odluke (engl. *Decision tree*), slučajne šume (engl. Random Forest), stroj potpornih vektora (engl. *Support Vector Machine, SVM u daljnjem tekstu*), K-najbliži susjed (engl. *K-Nearest Neighbor, KNN u daljnjem tekstu*), Naivni Bayes (engl. *Naive Bayes*) i pasivno agresivni algoritam (engl. *passive aggressive*). Algoritmi koji pokazuju najbolje rezultate mogu se koristiti u modelima za uspješnu detekciju dezinformacija i time pomoći u detekciji i sprječavanju njihovog širenja.

Motivacija za ovaj rad proizlazi iz potrebe za sprječavanjem širenja dezinformacija. S obzirom na trenutnu situaciju u svijetu i na događaje prošlih nekoliko godina, nikada nije bilo bitnije razlikovati istinu od laži. Rat između Rusije i Ukrajine, pandemija Covid-19, Brexit

(odvajanje Ujedinjenog Kraljevstva iz Europske Unije) imali su značajan utjecaj na živote velikog broja ljudi. Širenje dezinformacija stvara dodatne probleme koji nisu potrebni u takvim situacijama, radi zarade ili drugih razloga, ne uzimajući u obzir posljedice koje mogu nastati zbog takvih postupaka. Jedan primjer je širenje vijesti o pridruživanju Turske Europskoj Uniji, prilikom glasanja za Brexit. Prikazivale su se slike na kojima se vidi prosječna plaća u Turskoj i Engleskoj [3]. Cilj je bio uvjeriti stanovnike Ujedinjenog Kraljevstva kako će većina poslodavaca zaposliti Turske doseljenike, a njih otpustiti. Prošlo je 2 godine od Brexit-a a Turska još uvijek nije članica Europske Unije.

Tema ovog diplomskog rada je detekcija dezinformacija, odnosno lažnih vijesti, korištenjem algoritama strojnog učenja. Prvi dio rada se odnosi na teorijski dio (općenito o dezinformacijama, usporedba s drugim istraživanjima i usporedba postojećih skupova podataka) dok se u drugom dijelu predstavlja praktični dio zadatka (korišteni skupovi podataka i algoritmi) i prikazuju dobiveni rezultati. Na kraju se navode dobiveni zaključci i spominju moguće nadogradnje projekta. Također se želi saznati koliko je uspješna detekcija dezinformacija primjenom različitih algoritama strojnog učenja.

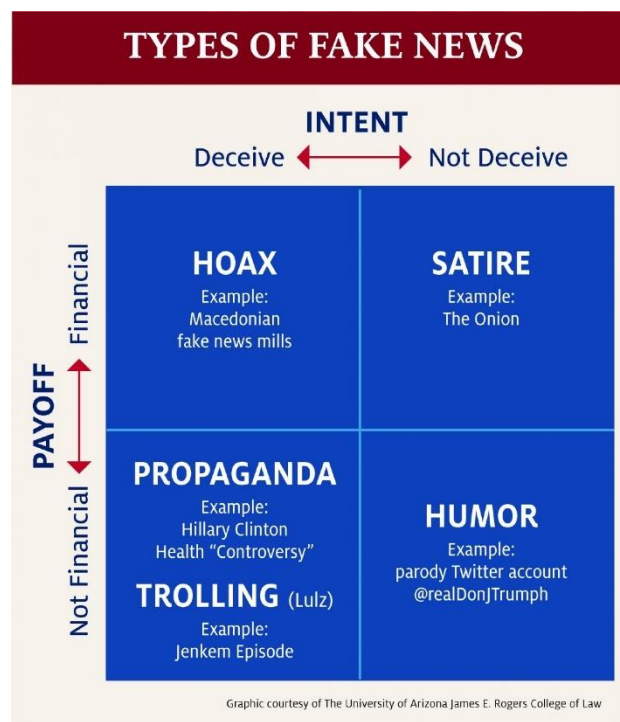
2. Problem dezinformacija i pregled dosadašnjih istraživanja

Ovaj dio rada obuhvaća opis i primjere uporabe dezinformacija, usporedbu sa provedenim istraživanjima i opis metodologija koje se koriste u radu.

2.1. Dezinformacije

Dezinformacije predstavljaju sve netočne ili zavaravajuće informacije. Takve vijesti nemaju valjane izvore i činjenice koje bi mogle potvrditi njihovu pouzdanost. Iako je pojam „dezinformacija“ populariziran zadnjih nekoliko godina, od početka modernog novinarstva postoje vijesti koje bi se mogle svrstati u lažne, odnosno netočne. 1835. godine, The New York Sun objavljuje nekolicinu članaka pod nazivom „Great Moon Hoax“ u kojima tvrde da postoji izvanzemaljska civilizacija na mjesecu [4]. Iako su napisani u obliku satire sa ciljem ismijavanja ranijih, ozbiljnijih nagađanja o izvanzemaljskom životu, čitatelji su ih ozbiljno shvatili. Kratko nakon objavljivanja članka, znatno se povećala prodaja The New York Sun novina [5].

Istraživanje provedeno od strane studenata Sveučilišta u Arizoni definira različite vrste dezinformacija: obmane, propaganda, „trolanje“, humor i satira, zajedno s motivima koji stoje iza njih [6]. Slika 1 prikazuje podjelu dezinformacija na navedene vrste.



Slika 1: Podjela dezinformacija
Izvor: [6]

Također su navedene podvrste koje navode motivaciju vijesti. Propaganda, „trolanje“ i obmane spadaju u podgrupu vijesti koje žele zavarati čitatelje dok satira i humor nemaju taj motiv. Druga podgrupa svrstava obmane i satiru u podgrupu vijesti koje su financijski motivirane za razliku od ostalih. U dijelu opisivanja vrsta dezinformacija, studenti definiraju satiru kao vijest koja nema namjeru zavaravati, iako ima namjerno lažan sadržaj i općenito je motivirana nematerijalnim interesom, iako financijska korist može biti sekundarni cilj. Obmane definiraju kao vijesti s namjerno lažnim sadržajem čiji je cilj zavaravanje čitatelja netočnim informacijama i financijski je motivirana. Propaganda je vijest s namjerno pristranim ili lažnim sadržajem čiji je autor namjeravao zavarati čitatelja, motivirane su promicanjem političkog cilja ili stajališta bez obzira na financijsku korist. „Trolanje“ je vrsta dezinformacija koja prikazuje vijesti pristrano ili lažno s namjerom zavaravanja čitatelja, motivirane su dobivanjem osobne humorne vrijednosti [6].

Postoje razne organizacije koje se bave detekcijom dezinformacija i provjeravanjem činjenica. PolitiFact¹ web stranica služi za provjeravanje izjava raznih političkih dužnosnika. „Truth-O-Meter“ je alat pomoću kojeg rangiraju svaki izjavu. PunditFact² je slična web stranica, koja provjerava izjave političkih stručnjaka. Obije su u vlasništvu Poynter instituta [7]. FactCheck je neprofitna organizacija koja je namijenjena glasačima u SAD-u. Žele povećati znanje građana činjenicama kako bi donijeli prave odluke prilikom glasanja [8]. Snopes³ je još jedna web stranica koja se bavi provjeravanjem činjenica. Korisnici mogu pretraživati stranicu za određene činjenice i u većini slučajeva će dobiti povratnu informaciju u obliku: istina, laž ili negdje između [9]. Iako su nabrojani alati korisni, potrebno je obrazovati ljude kako bi mogli sami primijetiti nepravilnosti u dezinformacijama. „The News Literacy Project“ je projekt novinara Alana Millera koji želi naučiti mlade ljude kritičkom razmišljanju prilikom čitanja raznih vijesti kako bi mogli prepoznati istinite od lažnih [10].

2.1.1. Primjer širenja dezinformacija: predsjednički izbori u SAD-u

2017. godine lažne vijesti su proglašene za pojam godine od strane Collins rječnika [11]. Korištenje pojma „lažna vijest“ povećalo se za 365% te godine. Jedan do glavnih razloga takovom naglom povećanju su predsjednički izbori u Sjedinjenim Američkim Državama 2016.

¹ <https://www.politifact.com/>

² <https://www.politifact.com/punditfact/>

³ <https://www.snopes.com/>

godine. Glavni kandidati te godine bili su Hillary Clinton i Donald Trump, za kojeg neki tvrde, pretežito on sam, kao stvaratelja pojma „lažna vijest“ [12]. Iako ta tvrdnja nije istinita, njegova uloga u populariziranju pojma ne može biti zanemarena. Student sa jednog Američkog sveučilišta, Cameron Harris, shvatio je kako može zaraditi novce, bez prevelike muke, širenjem dezinformacija. Odlučio je napraviti web stranicu na kojoj će objavljivati dezinformacije kako bi zaintrigirao glasače koji naginju na stranu republikanaca. Jedna od zanimljivijih objava navodi kako je nađena kutija u kojoj su unaprijed ispunjeni glasački listići u korist Hillary Clinton. Cilj je bio zaintrigirati glasače koji su već prije sumnjali u njenu korumpiranost i time potvrditi njihove sumnje. Vijest je podijelio na razne Facebook stranice koje je on sam kreirao. Objava je dosegla oko 6 milijuna ljudi. Cameron navodi kako njegova motivacija nije bila pobjeda Donalda Trumpa na izborima, nego samo zarada. Postoji mogućnost da je ta dezinformacija promijenila mišljenje nekolicine glasača prije izbora i time utjecala na same rezultate izbora [13]. Slika 2 prikazuje članak na dotičnoj web stranici.



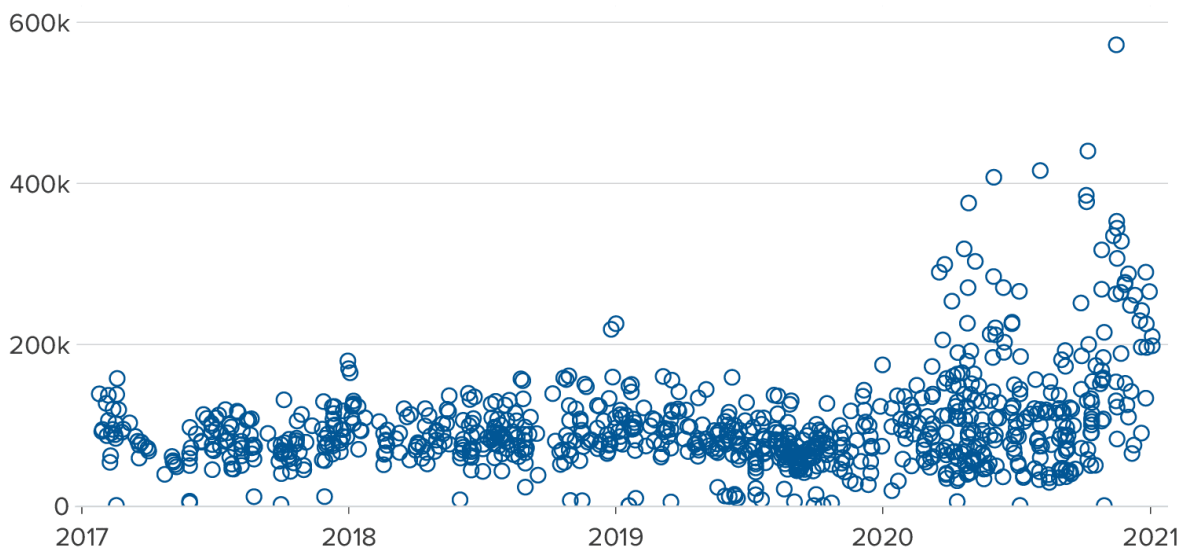
Slika 2: Članak Cammeron Harrisa
Izvor: [13]

Bitno je istaknuti kako je naslov ključan dio svake dezinformacije. Mora biti šokantan i zaintrigirati čitatelja kako bi ga naveo na otvaranje linka koji ga vodi na web stranicu. Takve web stranice su pune oglasa od strane Google-a. Ukoliko neki od čitatelja klikne na oglas, vlasnik web stranice zaradi određenu svotu novaca. U nekoliko dana Cammeron Harris je od navedenog članka zaradio 5000 dolara od ukupno 22000 dolara koje je zaradio tijekom cijele

kampanje [13]. Nakon što su izbori završili, pobjednik Donald Trump je pomoću društvene mreže Twitter dijelio svoja mišljenja s javnošću. Slika 3 prikazuje Trumpovo korištenja pojma „lažna vijest“ na Twitteru.

Trump's "fake news" tweets

Every tweet including the phrase “fake news” from @realDonaldTrump since his inauguration, with the number of favorites each post received on the vertical axis



Slika 3: Objave Donalda Trumpa na Twitteru
Izvor: [49]

Na vertikalnoj osi je prikazan broj objava koje su označene kao favoriti od strane korisnika Twittera. Jasno je vidljiv porast u 2020. godini, nakon što je Donald Trump izgubio izbore od Joe Bidena. Očito nezadovoljan porazom, Trump putem Twittera iskazuje svoje sumnje u legitimnost izvođenja glasanja. Cilj je bio uzrujati svoj dio glasača. Upravo to se i dogodilo. Nekolicina Trumpovih glasača uputilo se prema Capitolu kako bi zaustavili potvrdu pobjede Joe Bidena. Nasilno su ušli u zgradu i nanijeli veliku štetu i strah prisutnima. Ranjena je nekolicina policajaca, nanesena je šteta od 1.5 milijuna dolara i nekoliko osoba je tragično preminulo [14]. Navedeni su samo neki od mnogih slučajeva korištenja dezinformacija i posljedica koje mogu nastati njihovim širenjem. Ovaj je jedan od značajnih slučajeva koji pokazuje štetnost širenja dezinformacija jer se radi o predsjedničkim izborima jedne od najmoćnijih država na svijetu.

2.2. Pregled dosadašnjih istraživanja

Dezinformacije predstavljaju značajan problem, posebno zbog velikog razvoja društvenih mreža. Postoji 4.48 milijardi korisnika raznih društvenih mreža u cijelom svijetu. 99% njih pretražuju društvene mreže pomoću aplikacija na mobilnim uređajima. Facebook je vodeća društvena mreža sa 2.9 milijardi korisnika mjesečno [15]. Tako veliki broj korisnika omogućuje širenje vijesti, istinitih ili lažnih, rapidnom brzinom. Još do prije nekoliko desetljeća, ljudi su morali kupiti novine ili gledati televizijski program u određeno vrijeme kako bi saznali za nova događanja u svijetu. Danas je situacija znatno drugačija. Razvoj mobilne tehnologije omogućuje ljudima da u bilo kojem trenutku saznaju sve što ih zanima. Također svaka osoba sa malo informatičkog znanja može kreirati web stranicu i objavljevati dezinformacije na njoj. Te stranice mogu izgledati kao drugi pouzdani novinski portali, te zavarati ljude da su i one pouzdane. Kako bi stranica dobila na popularnosti, potrebno je dijeliti članke po društvenim mrežama. U istraživanju koje su proveli Vosougi i suradnici [16], prikazano je kako pouzdanim vijestima treba 6 puta duže nego lažnim da dosegnu 1500 korisnika Twittera. Razlog takvog rezultata može biti intrigirajući naslov objave koja sadrži dezinformacije. Dezinformacije se u većini slučajeva dijele u obliku poveznice na web stranicu. Poveznica se sastoji od naslova objave, odnosno članka. To je prvi dio informacije koji je čitateljima vidljiv. Kako bi se čitatelja navelo da otvori poveznicu, naslov mora biti napisan što „šokantnije“, bez otkrivanja puno informacija. Time se stvara znatiželja kod čitatelja. Takav naslov se zove „Clickbait“. Slika 4 prikazuje primjer „Clickbait“ naslova.



Slika 4: Clickbait
Izvor: [53]

Naslov se sastoji od 2 dijela. Prvi dio postavlja pitanje „Kako prepoznati lažne vijesti?“. Ukoliko bi tu bio kraj naslova, nekoliko čitatelja bi bilo dovoljno zaintrigirano, ako ih uistinu zanima tematika, te bi potencijalno otvorili poveznicu. Postoje i ostali čitatelji koje se ne bi odlučili na otvaranje linka. Kako bi nagovorili tu skupinu na otvaranje linka, dodaje se drugi dio naslova, koji glasi: „Odgovor će vas ZAPREPASTITI!!!“. Također na slici postoji dodatan tekst koji glasi: „ŠOKANTNO!“, „MORATE POGLEDATI!“, „NEĆETE VJEROVATI!“. Takav dodatak naslovu stvara zaintrigiranost kod čitatelja koji se može pitati što može biti toliko šokantno da će ga zaprepastiti. Istraživanje „Does Clickbait Actually Attract More Clicks?“ navodi da prilikom stvaranja dovoljno znatiželje među čitateljima, oni postanu prisiljeni kliknuti na poveznicu kako bi popunili prazninu u znanju [17].

Umjetna inteligencija, strojno učenje i NLP olakšavaju borbu protiv širenja dezinformacija. Istraživanje autora Kushal Agarwalla i suradnika koristi kombinaciju NLP-a i strojnog učenja za detekciju dezinformacija. Koriste logističku regresiju, SVM i Navni Bayes algoritme u svom radu. Usporedili su rezultate navedenih algoritama i došli do zaključka da najveću točnost ima Naivni Bayes (83%). Najniži rezultat je postigla logistička regresija (66%) [18]. Mjera točnosti, kao i ostale mjere, će se objasniti u poglavlju 2.4. Z. Khanam i njegovi suradnici provode istraživanje na približno sličnu temu. Koriste dva ista algoritma kao i u istraživanju autora Kushal Agarwalla (ne koriste logističku regresiju). Uz njih, koriste i KNN, Slučajne šume, Stabla odluke i XGBoost algoritme. Najveću točnost ima algoritam XGboost (više od 75%), dok je Naivni Bayes postigao najgore rezultate (manje od 70%) [19]. Zanimljivo je uočiti kako u dva primjera, na približno jednaku temu, isti algoritam postiže najbolji i najgori rezultat. Postoji više razloga za takve razlike u rezultatima, uzimajući u obzir da nisu trenirani na istom skupu podataka.

Korištenje tehnologije je jedan od načina detekcije dezinformacija. Drugi način je ljudsko prepoznavanje dezinformacija, odnosno sposobnost ljudi da uoče karakteristike takvih vijesti koje ne postoje kod pouzdanih vijesti. Tiziana Assenza i Alberto Cardaci provode istraživanje u kojem žele ispitati ljudsku sposobnost detektiranja dezinformacija. Ispitanici su pročitali nekolicinu kratkih članaka raznih novinskih portala. Omjer članaka koji su bazirani na pouzdanim vijestima i dezinformacijama je bio 50:50. Ispitanici su uspješno detektirali, u prosjeku, 12.45 od 20 (62.25%) članaka. 24.2% ispitanika je uspješno detektiralo manje od 10 (50%) članaka, a 14.29% njih je uspješno detektiralo 16 i više (>80%) članaka [20].

2.2.1. Pregled postojećih skupova podataka s anotacijama vezanim uz dezinformacije

Prikupljanje i priprema zadovoljavajućeg skupa podataka predstavlja jedan od najbitnijih ali i najtežih zadataka u cijelom procesu razvoja modela za detekciju dezinformacija. S obzirom da se radi o problemu klasifikacije, potrebno je odabrati postojeći anotirani skup podataka ili izraditi vlastiti anotirani skup podataka. Muhammad Yousaf i suradnici provode istraživanje koje se bavi tematikom detektiranja dezinformacija koristeći metode strojnog učenja. Za treniranje modela koriste nekoliko različitih skupova podataka [21]. „ISOT fake news dataset“ je jedan od korištenih skupova u njihovom radu. Sastoji se od popisa istinitih i lažnih vijesti. Istinite vijesti su prikupljene sa web stranice novinske agencije Reuters, koja se smatra pouzdanim izvorom. Lažne vijesti su prikupljene sa web stranica koje su označene kao nepouzdanе od strane Wikipedije i Politifact organizacije. Članci su većinom fokusirani na političke vijesti. Postoji 21.417 članaka koji predstavljaju istinite vijesti i 23.481 članaka koji predstavljaju lažne. Skup se sastoji od 4 stupca: naslov, tekst, tema i datum [22]. Korištena su još 2 skupa podataka koji su javno dostupni na web stanici Kaggle. Relativno su slične strukture. Prvi, pod nazivom „Fake News“ je skup podataka koji se koristi za natjecanje u programiranju koje je počelo 2018. godine. Cilj je ostvariti što veću točnost predikcije. Pobjednik natjecanja je ostvario točnost od 98%. Skup podataka je podijeljen na skup za treniranje (20.800 članaka) i testiranje (5200 članaka). Sastoji se od stupaca: id, naslov, autor, tekst i oznaka [23]. „Fake News detection“ je naziv drugog skupa sa stranice Kaggle, odnosno trećeg skupa koji je korišten u navedenom istraživanju. Postoji 4009 članaka od kojih je 46.7% lažnih a 53.3% istinitih [24]. Nažalost ne postoji puno informacija o načinu prikupljanja podataka za oba Kaggle skupa. Sva tri skupa su pohranjena u obliku csv (vrijednosti odvojene zarezom, engl. *comma separated values*) datoteke.

LIAR je još jedan skup podataka koji se koristi za klasifikaciju dezinformacija. Javno je dostupan za korištenje. Sastoji se od 12.800 kratkih izjava koje su ručno označene. Oznake nisu podijeljene na istinite i lažne, kao u prije navedenim skupovima, nego su uvedene dodatne oznake: „pants on fire“, lažne, jedva istinite, polu istinite, većinom istinite i istinite. Sve oznake sadrže otprilike 2300 izjava osim oznake „pants on fire“ koja sadrži 1050 izjava. Ostali stupci u skupu se odnose na informacije od govorniku (autoru) izjava. U jednom od stupaca su zapisane vrijednosti koje predstavljaju broj prijašnjih netočnih izjava govornika, u formatu od 5 brojeva. Npr. {0, 2, 1, 4, 7} označava 0 „pants on fire“, 2 lažne, 1 jedva istinitu, 4 većinom istinite i 7 istinitih izjava. Zadnji stupac sadrži informacije o kontekstu (lokacija i događanje) u

kojem je dana izjava. Intervju na televiziji ili radiju, predizborni govor, novinski članci, debate i objave na društvenim mrežama su izvori većine izjava [25]. Slika 5 prikazuje jednu od objava iz LIAR skupa.

Statement: *“The last quarter, it was just announced, our gross domestic product was below zero. Who ever heard of this? Its never below zero.”*
Speaker: Donald Trump
Context: presidential announcement speech
Label: Pants on Fire
Justification: According to Bureau of Economic Analysis and National Bureau of Economic Research, the growth in the gross domestic product has been below zero 42 times over 68 years. Thats a lot more than “never.” We rate his claim Pants on Fire!

Slika 5: Primjer izjave (LIAR skup podataka)
Izvor: [25]

Autor izjave je, bivši predsjednik SAD-a , Donald Trump. U izjavi navodi kako rast BDP-a njegove države nikada nije bio ispod nule. Izjava je dobila „pants on fire“ oznaku, jer prema činjenicama pouzdanih izvora, rast BDP-a u SAD-u je bio je bio ispod nule 42 puta u zadnjih 68 godina.

Andreas Vlachos i Sebastian Riedel provode istraživanje na temu provjere činjenica. Prema Wangu, oni su prvi konstruirali skup podataka za detekciju dezinformacija i činjenica [25]. Podatke su prikupili sa web stranica Channel 4 i PolitiFact. Obije stranice imaju preko 1000 provjerenih izjava koje su bazirane na relevantnim problemima u Ujedinjenom Kraljevstvu i SAD-u. Za svaku izjavu su zabilježili datum, autora, URL i oznaku vjerodostojnosti. S obzirom da izjave nisu bile označene na isti način, definirali su oznake: istinita, većinom istinita, polu istinita, većinom lažna i lažna izjava. Od 221 izjave koje su prikupili, odlučili su da su samo 106 prikladne za njihov model [26]. S obzirom da je to relativno mali skup podataka, jedan od autora, Andreas Vlachos zajedno sa Williamom Ferreiom konstruira novi skup podataka pod nazivom Emergent. Podatci su sakupljeni od strane novinara koji su radili na Emergent projektu [27]. Skup podataka se sastoji od 300 izjava i 2595 povezanih članaka. Za svaku izjavu postoji prosječno 8.64 članaka, najmanji broj članaka po izjavi je 1 a najveći 50 članaka. Svaka izjava sadrži jednu od oznaka: za (u člancima je navedeno

da je istinita), protiv (u člancima je navedeno da je lažna), postojeća (postoji u člancima, ali nema potvrde o njenoj istinitosti) [28]. Slika 6 prikazuje primjer izjave i donesenu oznaku.

Claim: Robert Plant ripped up an \$800 million contract offer to reunite Led Zeppelin
Source: mirror.co.uk (shares: 39,140)
Headline: Led Zeppelin's Robert Plant turns down £500MILLION to reform supergroup
Stance: <i>for</i>
Source: usnews.com (shares: 850)
Headline: No, Robert Plant Didn't Rip Up an \$800 Million Contract
Stance: <i>against</i>
Source: forbes.com (shares: 3,360)
Headline: Robert Plant Reportedly Tears Up \$800 Million Led Zeppelin Reunion Contract
Stance: <i>observing</i>
Veracity: <i>False</i>

Slika 6: Primjer izjave iz Emergent skupa podataka
Izvor: [28]

Izjava navodi da je Robert Plant pokidao ugovor u kojem mu je ponuđeno 800 milijuna dolara za ponovno ujedinjenje Led Zeppelina. Ispod su prikazani naslovi 3 članaka, svaki sa svojeg stajališta. Prvi članak tvrdi da je izjava istinita, suprotno njemu, drugi tvrdi da je lažna a zadnji samo prenosi vijest da je navodno istina. Nakon detaljnog istraživanja od strane novinara, došli su do zaključka da je izjava lažna. Zanimljivo je za primijetiti da je prvi članak, koji dokazano prenosi lažnu vijest, podijeljen otprilike 40.000 puta, dok je drugi, istinit članak podijeljen samo 850 puta.

Autori Emilio Ferrara i suradnici provode istraživanje u kojem izgrađuju ReCOVeRY skup podataka kako bi ispitali kredibilitet vijesti u doba COVID-19 pandemije. ReCOVeRY skup podataka je osmišljen i konstruiran u svrhu „borbe“ protiv dezinformacija i lažnih vijesti koje se odnose na temu COVID-19 [29]. Ispitali su 2000 novinskih portala od kojih je 60 identificirano sa izrazito visokim ili izrazito niskim stupnjem kredibiliteta. Koristili su *NewsGuard* i *Media Bias/Fact Check* (MBFC u daljnjem tekstu) alate. Oba alata imaju različite načine označavanja pouzdanosti portala. *NewsGuard* alat bodovanjem po raznim kriterijima ocjenjuje pojedine članke. Svaki kriterij donosi određen broj bodova (od 5 do 22), najviše je moguće skupiti 100 bodova. Najbitniji kriterij, odnosno onaj koji donosi najviše bodova, se odnosi na kontinuirano objavljivanje dezinformacija. Ukoliko nije zabilježeno da je novinski portal u više navrata objavljivao dezinformacije, dobiti će 22 boda. Portali koji skupe preko 60 bodova, smatraju se kredibilnima. Svi portali, sa ukupnim brojem ispod 60, se smatraju

nepouzdanima [29]. MBFC alat označuje portale u rangu od 6 razina točnosti činjenica: vrlo visoka, visoka, većinom činjenični, mješovita, niska, vrlo niska [29]. Ferrara i suradnici koriste kombinaciju oba alata prilikom konstruiranja ReCOVerry skupa podataka. Ukoliko novinski portal ima ukupnu ocjenu veću od 90 (*NewsGuard*) i spada u rang vrlo visoke ili visoke točnosti (MBFC) smatra se pouzdanim. U suprotnom, ukoliko portal ima ukupnu ocjenu manju od 30 (*NewsGuard*) i spada u rang ispod mješovit (MBFC) smatra se nepouzdanim. Nakon što su odabrali zadovoljavajuće novinske portale, morali su odabrati članke koji se odnose na COVID-19 pandemiju. To su izveli traženjem ključnih riječi („SARS-CoV-2“, „COVID-19“ i „Coronavirus“) unutar članaka. Prikupljeni članci se sastoje od 12 različitih komponenti: ID, URL, izdavač, datum izdavanja, autor, naslov i cijeli tekst, naslovna slika, država, politička pristranost, *NewsGuard* bodovi i MBFC rang. Zatim su, pomoću Twitter-ovog alata za pronalaženje objava, zabilježili sve objave na Twitter-u koje su vezane za pojedini članak, ukoliko objava sadrži URL članka i objavljena je nakon izdavanja članka [29]. Konačni skup podataka sadrži 2029 članaka, sa balansiranošću 2:1 u korist pouzdanih članaka. Ukupan broj objava na Twitter-u, koje su vezane za članke, iznosi 140.820, od kojih je 114.402 objave za pouzdane a 26.418 za nepouzdate.

2.3. Opis metodologija

Radi uspješne detekcije dezinformacija, u ovom radu se koristi NLP u kombinaciji sa strojnim učenjem. NLP kombinira računalnu lingvistiku, modeliranje ljudskog jezika sa modelima strojnog učenja i dubokog učenja. Zajedno, ove tehnologije omogućuju računalima da obrađuju ljudski jezik u obliku teksta ili glasovnih podataka i da „razumiju“ njegovo puno značenje, zajedno s namjerom govornika ili pisca [30].

2.3.1. Obrada prirodnog jezika

Obrada prirodnog jezika je područje umjetne inteligencije koje se bavi proučavanjem interakcija između računala i ljudskog jezika. Glavni cilj je omogućiti da računala razumiju pisani tekst i ljudski govor. 50-ih godina prošlog stoljeća stvara se interes za obradu prirodnih jezika. Alan Turing objavljuje članak pod nazivom „Računalni strojevi i inteligencija“ (engl. *Computing Machinery and Intelligence*). U prvom odlomku postavlja pitanje: „Mogu li strojevi misliti?“ [31]. U članku se prvi put spominje Turingov test, u kojem se pokušava saznati ako računalo može uvjeriti čovjeka da je ono također čovjek, odnosno da li računalo može voditi

razgovor sa osobom tako da ta osoba misli da razgovara sa drugom osobom a ne sa računalom. Sastoji se od 3 člana, jednog „ispitivača“ koji postavlja pisana pitanja za ostala 2 člana, koji se nalaze u drugoj prostoriji. Jedan je čovjek a drugi je računalo. Ukoliko računalo uspije zavarati ispitivača da odgovore daje čovjek, uspješno je prošlo Turingov test.

NLP cjevovod (engl. *Pipeline*) je set od sedam koraka potrebnih za čitanje i razumijevanje ljudskog jezika [32]. U prvom koraku se segmentiraju rečenice, odnosno izdvajaju se rečenice iz odlomka radi boljeg razumijevanja. U drugom koraku se izdvajaju riječi iz rečenice. Ovaj korak je jednostavan za jezike u kojima se riječi odvajaju razmacima, ali problem stvaraju jezici, kao kineski, u kojima to nije slučaj. Treći korak je korjenovanje (engl. *stemming*) u kojem se riječi svode na osnovni oblik, uklanjaju se prefiksi i sufiksi [33]. Četvrti korak je lematizacija (engl. *lemmatisation*), odnosno svođenje riječi na kanonski oblik. Kanonski oblik predstavlja riječi koji su zapisane kao imenice u nominativu, glagoli u infinitivu. U petom koraku se uklanjaju zaustavne riječi. To su riječi koje se najčešće pojavljuju a ne nose veliko značenje. U sljedećem koraku se otkrivaju ovisnosti jedne riječi o drugoj odnosno istražuje se kako je jedna riječ povezana sa drugom. U zadnjem koraku se označavaju vrste riječi. Za svaku riječ se definira da li je ona imenica, glagol, pridjev, itd. Ovime se detaljnije opisuju riječi kako bi se izbjegao problem dvoznačnosti [32].

NLP ima ulogu u puno alata i programa koji se svakodnevno koriste. Neki od takvih alata i programa su: filtriranje mailova, automatsko ispravljanje teksta, chatbot-ovi, pametni asistenti (Siri, Alexa, Bixby, itd.), strojno prevođenje teksta i mnogi drugi.

2.3.2. Strojno učenje

Postoje tri podgrupe algoritama strojnog učenja temeljenih na vrsti problema koji se nastoje riješiti. To su nadzirano učenje, nenadzirano učenje i podržano odnosno pojačano učenje.

Algoritmi koji spadaju pod nadzirano učenje su trenirani na anotiranom skupu. Anotirani skup sadrži dodatne informacije na temelju kojih model strojnog učenja može učiti kako klasificirati podatke. Na primjeru skupa podataka koji se koristi u ovom radu, postoji stupac u kojem je navedeno da li je vijest istinita ili lažna. Na temelju tih informacija model uči, a potom se može koristiti za detekciju (predikciju) dezinformacija na skupu novih podataka koji nisu anotirani. Nadzirano učenje se može dodatno podijeliti na regresiju i klasifikaciju. Regresijski algoritmi se koriste ako postoji relacija između izlazne varijable (zavisne) i ulazne varijable

(nezavisne), uzimajući u obzir da je izlazna varijabla numerička vrijednost. Korisni su za predviđanje kontinuiranih vrijednosti na temelju različitih podataka. Neki od algoritama su:

- Linearna regresija
- Lasso regresija
- Slučajne šume
- Regresija potpornih vektora (engl. *Support Vector Regression, SVR*)
- Stablo odluke [34]

Klasifikacija je proces predviđanja klasa na temelju dobivenih podataka. Glavni zadatak je identificirati klase u koje će se novi podaci smjestiti. Neki od klasifikacijskih algoritama su:

- Logistička regresija
- SVM
- KNN
- Slučajne šume
- Stablo odluke [35]

U slučaju stabla odluke i slučajnih šuma, može ih se koristiti u oba slučaja, za regresiju i za klasifikaciju.

Za razliku od nadziranog učenja, algoritmi nenadziranog učenja su trenirani na skupu podataka koji nije anotiran. Glavni im je cilj detektirati uzorke u podacima. Model traži skrivene značajke i razvrstava podatke na smislen način. Nenadzirano učenje se koristi za grupiranje, vizualizaciju, redukciju dimenzija i asocijativno učenje [36]. Grupiranje je proces organiziranja podataka, koji su na neki način slični, u grupe. Takve grupe nazivaju se klasteri. Neki od algoritama grupiranja su K-prosječna vrijednost, C- prosječna vrijednost, Hijerarhijska klaster analiza, itd. Vizualizacijski algoritmi generiraju 2D ili 3D grafove koji prikazuju veliku količinu podataka pokušavajući očuvati njihovu strukturu kako bi potencijalno mogli detektirati neočekivane uzorke [36]. Zadatak redukcije dimenzija je pojednostavljivanje podataka bez gubitka previše informacija. Da bi model bio pouzdan, dimenzije skupa podataka ne smiju biti velike [37]. Analiza glavnih komponenti (engl. *Principal Components Analysis, PCA* u daljnjem tekstu), Krenel PCA, T-distribuirano stohastičko umetanje u susjedstvo (engl. *T-distributed Stochastic Neighbor Embedding*) neki su od algoritama za vizualizaciju i redukciju dimenzija. Asocijativno učenje je vrsta nenadziranog učenja kojemu je zadatak pronaći

poveznice u velikom broju podataka. Apriori, Eclat i algoritam rasta učestalih skupova (engl. *F-P Growth Algorithm*) su najčešći algoritmi koji se koriste u asocijativnom učenju.

Pojačano učenje je zasnovano na pronalasku ravnoteže između istraživanja novog i eksploatacije poznatog znanja. Smatra se kao znanost o donošenju odluka. Radi se o učenju optimalnog ponašanja u okruženju kako bi se ostvarila maksimalna nagrada [38]. Postoje 4 komponente: agent, okruženje, nagrada i zadatak. Agent je entitet koji se nalazi u nepoznatom okruženju. Njegov zadatak je da istraži okruženje koje je promjenjivo. Ukoliko agent uspješno obavi akciju, biti će nagrađen, u suprotnom će biti kažnjen. Zadatak se smatra ispunjenim ukoliko agent nauči koje akcije će mu donijeti maksimalnu nagradu, uzimajući u obzir da agent ne zna ako će ga te akcije dovesti do konačnog cilja. Q-učenje i SARSA (engl. *State-Action-Reward-State-Action*) su jedni od algoritama pojačanog učenja. Tablica 1 prikazuje usporedbu navedenih podgrupa strojnog učenja.

Kriteriji	Nadzirano učenje	Nenadzirano učenje	Pojačano učenje
Definicija	Uči pomoću anotiranih podataka	Uči koristeći podatke koji nisu anotirani i bez navođenja	Radi na interakciji s okolinom
Tip skupa podataka	Anotirani skup podataka	Skup podataka koji nije anotiran	Nema unaprijed definiranih podataka
Tip problema	Regresija i klasifikacija	Asocijativno učenje i grupiranje	Eksploatacija ili istraživanje
Nadzor	Pod nadzorom	Nema nadzora	Nema nadzora
Algoritmi	Linearna regresija, logistička regresija, SVM, itd.	K- prosječna vrijednost, C- prosječna vrijednost, Apriori	Q-učenje, SARSA
Cilj	Izračun ishoda	Otkrivanje značajnih uzoraka	Učenje niza radnji
Upotreba	Procjena rizika, predviđanje prodaje	Sustav preporuka, detekcija anomalija	Autonomna vozila, zdravstvo

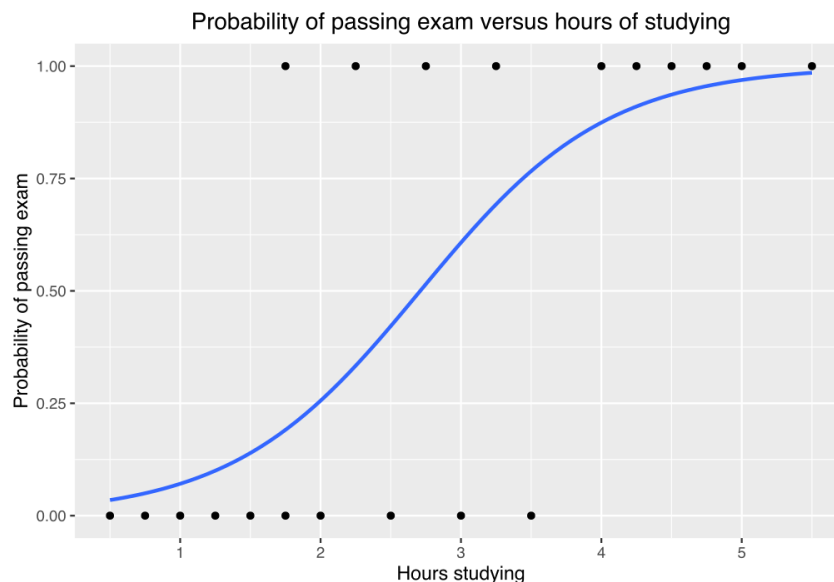
Tablica 1: Usporedba podgrupa strojnog učenja
Izvor: [50]

U nastavku će se opisati algoritmi koji su korišteni u praktičnom dijelu rada.

2.3.2.1. Logistička regresija

Logistička regresija je algoritam strojnog učenja koji se koristi za probleme binarne klasifikacije. Model logističke regresije procjenjuje matematičku vjerojatnost pripadanja nekoj kategoriji. Ukoliko je vjerojatnost veća od 0.5 (50%) ishod će biti pozitivan (pripada kategoriji), u suprotnom će biti negativan (ne pripada kategoriji). Logističku regresiju možemo podijeliti u tri vrste: binarna logistička regresija, multinomna logistička regresija, ordinalna logistička regresija. Binarna logistička regresija predviđa odnos između nezavisne varijable i zavisne (binarne) varijable. Primjer je predviđanje košarkaške utakmice, jedini ishodi su pobjeda ili poraz za određeni tim. Multinomna logistička regresija se razlikuje od binarne po broji mogućih ishoda, ima ih više od 2. Primjer je odabir javnog prijevoznog sredstva na putu do posla (autobus, vlak, tramvaj, itd.). Ordinalna logistička regresija se razlikuje od ostalih po tome što ima zavisnu varijablu ordinalnog tipa. Primjer takve varijable je veličina majice (S, M, L, XL) [39].

Algoritam je moguće pokazati pomoću sigmoidne funkcije koja je prikazana na slici 7.



Slika 7: Graf logističke regresije
Izvor: [51]

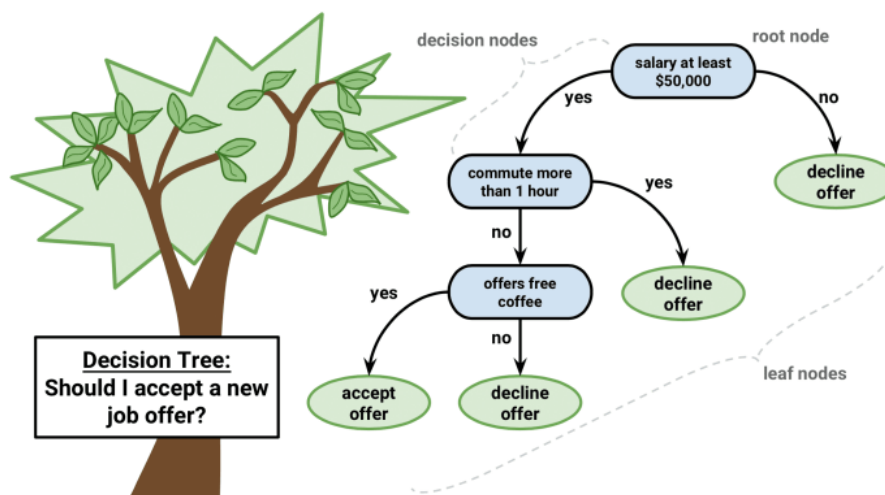
Na apscisi je definiran broj sati učenja a na ordinati vjerojatnost prolaska ispita. Apscisa predstavlja nezavisnu varijablu a ordinata zavisnu. Vrijednosti na ordinati se nalaze između 0 i 1. Točke na grafu predstavljaju uspjeh na ispitu. Ukoliko se točka nalazi na dnu grafa, student nije prošao ispit, a ukoliko se nalazi na vrhu, student je uspješno prošao ispit. Iz grafa

se može zaključiti da su studenti, koji su proveli više vremena učeći za ispit, imali veću prolaznost, uz nekoliko izuzetaka.

2.3.2.2. Stablo odluke

Stabla odluke se koriste za probleme klasifikacije i regresije. U sklopu ovog rada algoritam će se koristiti za klasifikaciju, uzimajući u obzir prirodu problema. Struktura algoritma nalikuje stablu okrenutom naopako, unutar kojeg postoje 3 tipa čvora: korijenski čvor (engl. *Root node*), čvor odluke (engl. *Decision node*) i krajnji čvor (engl. *Leaf node*). Postoje i grane koje prenose vrijednosti sa jednog čvora na drugi.

Slika 8 prikazuje primjer stabla odluke.



Slika 8: Primjer stabla odluke
Izvor: [52]

U primjeru prikazanom na slici 8 zadatak prikazanog stabla je odlučiti da li se treba prihvatiti ili odbiti ponuda za posao. Korijenski čvor predstavlja početak svakog stabla. U njemu je zapisan glavni uvjet, odnosno najupotrebljiviji atribut za navedeni zadatak. U ovom slučaju ispituje se da li je plaća minimalno 50.000 dolara. Ukoliko je, korijenski čvor se razdvaja na čvorove odluke. U suprotnom se donosi odluka da se obje ponuda za posao. Ta odluka je pohranjena u krajnji čvor. U prvom čvoru odluke, na slici 4, se ispituje ukoliko je potrebno putovati više od jednog sata do radnog mjesta. Ako je odluka da, ponuda za posao se odbija. Ako je odgovor ne, ponovno se prethodni čvor razdvaja na nove čvorove odluke, u primjeru sa slike je samo jedan novi čvor odluke, a može ih biti više. U njemu se ispituje ako posao nudi

besplatnu kavu. Ako je odgovor ne, ponuda za posao se odbija. Ako je odgovor da, svi uvjeti su zadovoljeni i donosi se odluka da će se prihvatiti ponuda za posao.

Entropija i informacijska dobit bitni su faktori prilikom stvaranja stabla odluke. Informacijska dobit je usko povezana s entropijom. Prilikom smanjivanja entropije, povećava se informacijska dobit. Entropija je mjera kojom se mjeri čistoća ili nasumičnost nekog skupa [40]. Formula za entropiju binarne klasifikacije prikazana je izrazom:

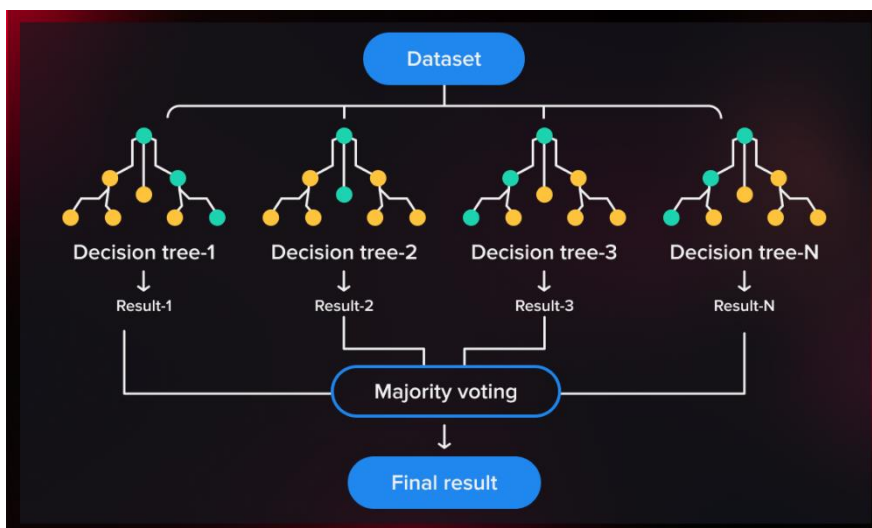
$$E(S) = -p_p \log_2 p_p - p_n \log_2 p_n$$

Gdje E predstavlja entropiju, S zadani skup, p_p postotak pozitivnih članova u S , p_n postotak negativnih članova u S . Na primjer, postoji skup od ukupno 50 objava, 20 je istinitih a 30 lažnih. Ukoliko se navedeni brojevi uvrste u formulu, entropija će iznositi 0.97 što je vrlo blizu maksimalnoj vrijednosti 1. Entropija će biti 1 kada skup sadrži jednak broj pozitivnih i negativnih članova. Informacijska dobit se može računati kao razliku entropije prije i poslije neke podjele. Koristi se za odabir atributa koji donose najviše informacija, kako bi model stabla preciznije donijelo odluku.

2.3.2.3. Slučajne šume

Isto kao i kod stabla odluke, slučajne šume mogu se koristiti za probleme klasifikacije i regresije. Zasnovane su na stablima odluke, odnosno funkcioniranju na način kombiniranja više stabla odluke u jedan zajednički sustav. Slučajne šume stvaraju stabla odluke nasumično odabirući njihove attribute. Svako stablo bira jednu od klasa za svoj rezultat. Klasa koja se pojavljuje u najviše stabla je izabrana za rezultat cijelog modela [41]. Primjer korištenja slučajne šume za donošenje odluka može biti odabir boje novog auta. Peteročlana obitelj se odluči za kupovinu novog auta i pokušavaju odabrati boju auta. Svaki član obitelji predstavlja jedno stablo odluke u šumi stabala. Ponuđeno je nekoliko boja u kojima dolazi auto. Zadatak svakog člana je glasati za jednu od boja. Boja koja dobije najviše glasova biti će odabrana. U nekim slučajevima 5 stabala može biti premalo za donošenje odluke, stoga je poželjno imati što veći broj stabala u šumi.

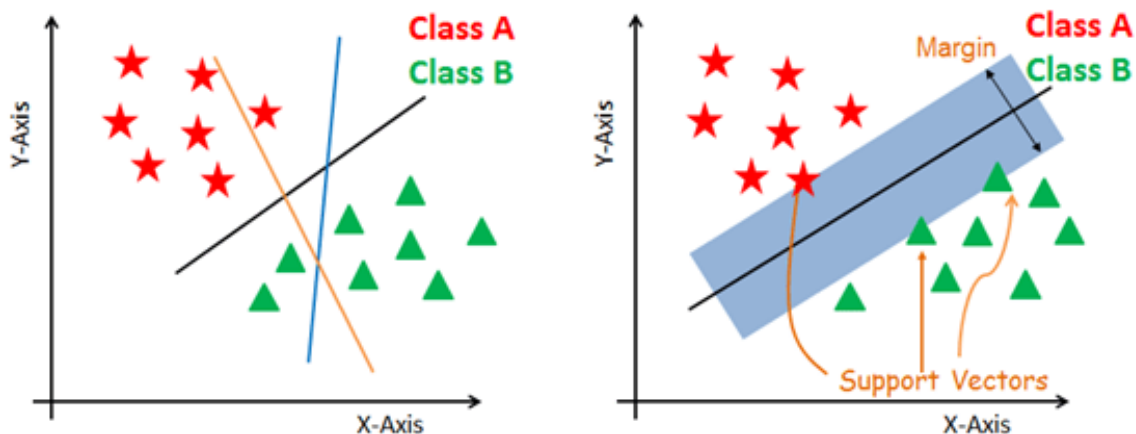
Slika 9 prikazuje primjer slučajne šume. Na vrhu je prikazan skup podataka kao početni čvor. Iz njega se stvaraju nekolicina stabala odluka kojima je nasumično podijeljen dio podataka. Svako stablo generira rezultat. Glasovanjem se odabire rezultat koji se smatra finalnim rezultatom.



Slika 9: Primjer slučajne šume
Izvor: [41]

2.3.2.4. Stroj potpornih vektora (SVM)

SVM algoritmi se koriste za probleme klasifikacije i regresije, kao i prošla 2 algoritma. Zadatak SVM algoritma je detekcija hiper-ravnine (engl. *Hyperplane*) u n-dimenzijском prostoru koja najbolje klasificira podatke u dvoje skupine (klase). Postoji više odgovarajućih hiper-ravnina. Bira se ona koja je najudaljenija od podatkovnih točaka. Dimenzije hiper-ravnine ovise o broju ulaznih varijabli. Ukoliko postoje 2 ulazne varijable, hiper-ravnina će biti linija, ako ih je 3, hiper-ravnina će biti dvodimenzionalna ravnina u trodimenzionalnom prostoru [42]. „Support“ vektori su podatkovne točke koje se nalaze na najmanjoj udaljenosti od hiper-ravnine, i smatraju se bitnijim podatkovnim točkama od ostalih. Slika 10 prikazuje odabir hiper-ravnine i „support“ vektore.



Slika 10: SVM primjer
Izvor: [43]

Lijevi dio slike prikazuje podatkovne točke u obliku narančastih zvjezdica (klasa A) i zelenih trokuta (klasa B). Također su prikazane 3 hiper-ravnine koje dijele podatkovne točke. S obzirom da žuta i plava hiper-ravnina imaju manje udaljenosti od podatkovnih točaka, bira se crna hiper-ravnina. Ona jedina potpuno točno odvaja podatkovne točke klase A i B. Na desnom djelu slike prikazani su „support“ vektori na odabranoj hiper-ravnini. Ukoliko podatkovne točke nije moguće podijeliti linearnom hiper-ravninom, koristi se „kernel trick“ pomoću kojeg se niže dimenzijski prostor pretvara u više dimenzijski kako bi se moglo uspješno razdvojiti podatkovne točke [43].

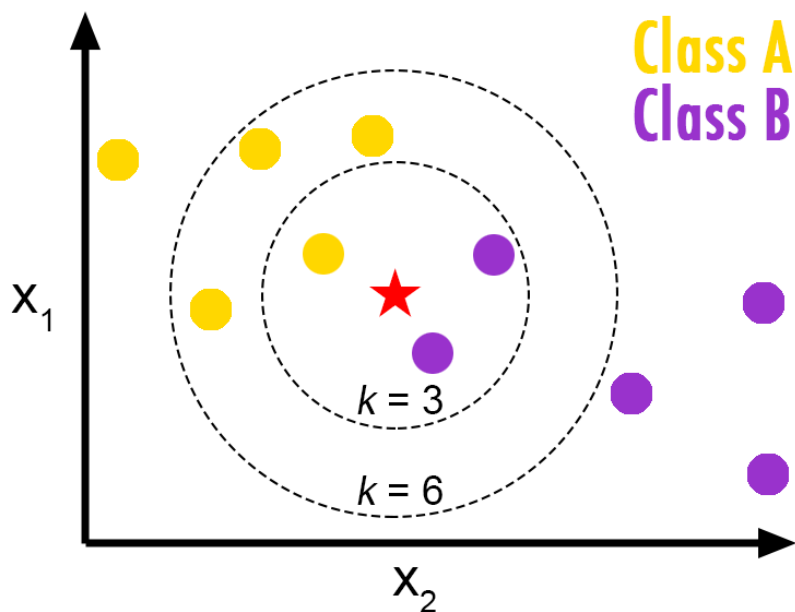
2.3.2.5. K-najbližih susjeda (KNN)

KNN algoritam se koristi za probleme klasifikacije i regresije. Funkcionira na principu najmanje udaljenosti između podatkovnih točaka. Smatra se „lijenim“ algoritmom zbog toga što ne trenira podatke sve dok ne dobije nove, neviđene podatke. Prije toga samo pohrani podatke u fazi treniranja [44]. Nove podatkovne točke se klasificiraju u klasu koja ima najveći broj susjeda. Potrebno je definirati broj najbližih susjeda prije procesa klasifikacije. Najbliži susjed se odredi mjerenjem euklidske udaljenosti (engl. *Euclidean distance*). Euklidska udaljenost predstavlja udaljenost 2 točke u Euklidskom prostoru. Formula za računanje euklidske udaljenosti prikazana je izrazom:

$$d(p, q)^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$$

za koje vrijedi $p(x_1, y_1)$, $q(x_2, y_2)$ [45].

Slika 11 prikazuje primjer KNN algoritma za 2 klase podataka. Postoji klasa A koja je prikazana žutim kružićima i klasa B prikazana ljubičastim kružićima. U sredini grafa se nalazi crvena zvjezdica koja predstavlja novu podatkovnu točku. Oznaka „k“ predstavlja zadanu vrijednost najbližih susjeda. Ukoliko je k postavljen na 3, zvjezdica će pripadati klasi B zato što postoje 2 ljubičasta kružića a samo jedan žuti kružić unutar 3 najbliža susjeda. Kada se k poveća na 6, zvjezdica, pripada klasi A. Sad postoji 4 žuta kružića, a samo 2 ljubičasta unutar 6 najbližih susjeda.



Slika 11: Primjena KNN algoritma
Izvor: [45]

2.3.2.6. Naivni Bayes

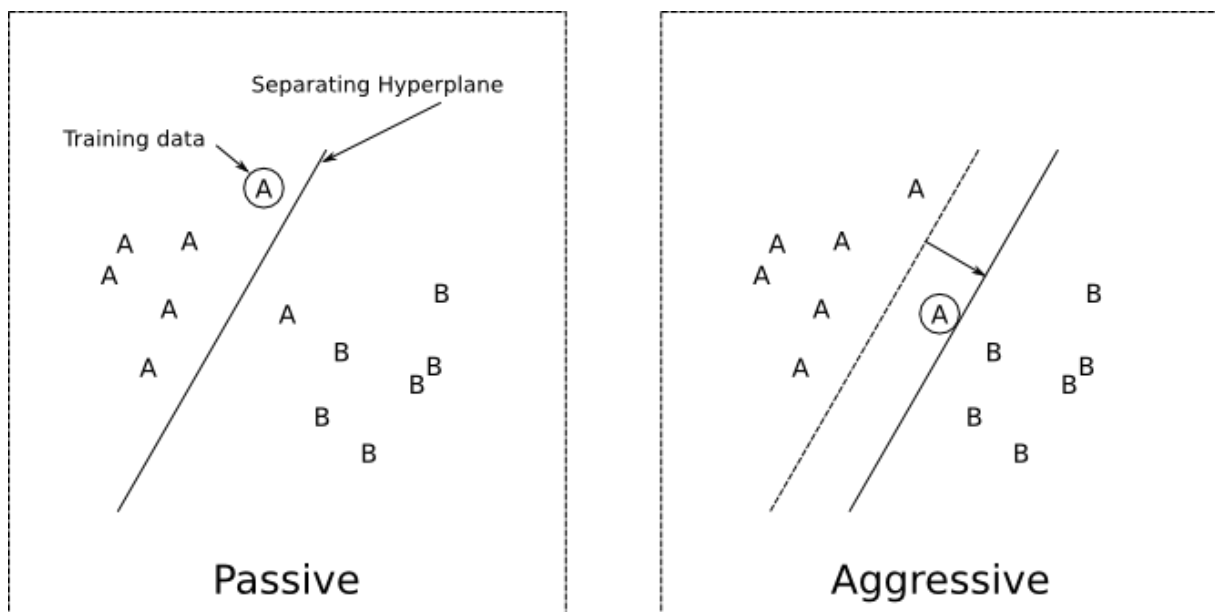
Naivni Bayes algoritmi se koriste za problem klasifikacije. Bazirani su na Bayes-ovom teoremu, po kome je algoritam dobio ime. Postoji više algoritama, ali u svrhu ovog rada opisati će se multinomni Naivni Bayes-ov algoritam. Kako bi se razumio algoritam potrebno je razumjeti Bayes-ov teorem. Teorem opisuje vjerojatnost nekog događaja, na temelju prethodnog znanja o uvjetima koji se odnose na događaj [46]. Formula na kojoj je baziran teorem prikazana je izrazom:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

gdje A i B predstavljaju događaje i $P(B) \neq 0$, $P(A | B)$ je vjerojatnost da se A dogodio ako je istina da se B dogodio, $P(B | A)$ je vjerojatnost da se B dogodio ako je istina da se A dogodio, $P(A)$ i $P(B)$ su vjerojatnosti događanja bez dodatnih uvjeta [47]. Multinomni Naivni Bayes-ov algoritam omogućuje klasifikaciju podataka koji ne mogu biti prikazani numerički. Cilj klasifikacije je dodijeliti dijelove teksta klasama tako što određuju vjerojatnost da taj dio teksta pripada klasi dugih tekstova koji imaju istu temu. Pridjev „naivni“ je pridodan nazivu zbog pretpostavke da je pojavljivanje svake riječi neovisno o drugim riječima, što u pravom svijetu nije istina.

2.3.2.7. Pasivno agresivni algoritam

Pasivno agresivni algoritam spada u skupinu online algoritama strojnog učenja. Kod online strojnog učenja, podatci su dostupni u sekvencijalnom redoslijedu, odnosno model se trenira postepeno, ovisno o dostupnosti podataka. Dobar primjer je detekcija dezinformacija na društvenim mrežama. Model je treniran na trenutno dostupnom skupu i kada novi podatci postanu dostupni, model se trenira na njima, s obzirom da se na društvenim mrežama konstanto generiraju novi podaci. Korisni su za jako velike skupove podataka, gdje je potrebna velika količina slobodnog mjesta za pohranu, jer nije potrebno trenirati model na cijelom skupu, nego u manjim količinama u više navrata [48]. Pasivno agresivni algoritam je dobio ime zbog načina izvođenja, ostaje „pasivan“ kod točnih predikcija, odnosno ne rade se promjene u modelu, a postane „agresivan“ kod netočnih, izvode se izmjene kako bi predikcija postala točna. Slika 12 prikazuje 2 faze algoritma. U pasivnoj fazi se ništa ne izvodi. S obzirom da je jedan podatak krivo klasificiran, odnosno nalazi se na krivoj strani razdvajajuće hiper-ravnine, prelazi u agresivnu fazu. Hiper-ravnina se pomiče minimalno, sve dok se ne popravi kriva klasifikacija.

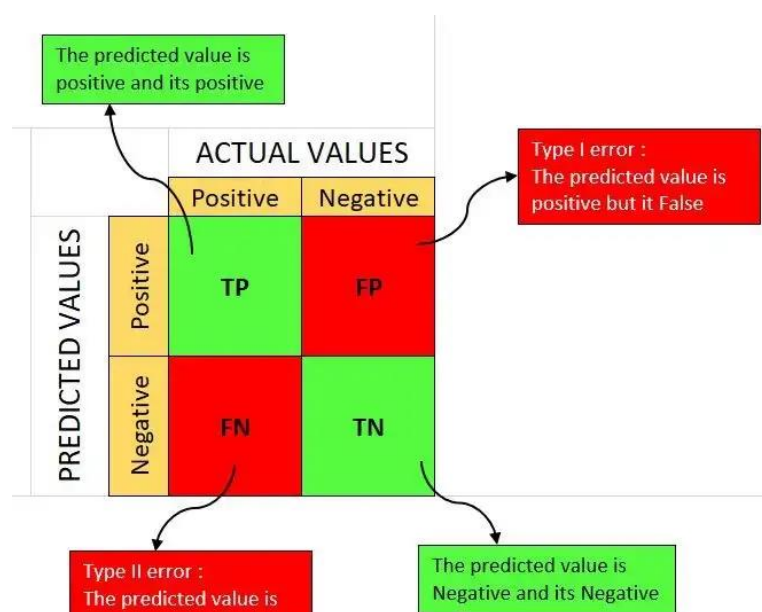


Slika 12: Izvođenje pasivno agresivnog algoritma
Izvor: [54]

2.4. Opis evaluacije modela

U ovom djelu rada će se objasniti pojmovi kojima se evaluiraju modeli strojnog učenja. Prije treniranja modela, potrebno je anotirani skup podijeliti na skupove za treniranje i testiranje. U većini slučajeva, skup za treniranje (engl. *train*) će sadržavati otprilike 70% podataka, dok će ostali postotak sadržavati skup za testiranje (engl. *test*). U ovom radu, svi skupovi podataka su podijeljeni u omjeru 75:25. Nasumično su podijeljeni podaci uzimajući u obzir balansiranost skupa. Novi skupovi, *train* i *test*, će imati istu balansiranost kao i originalni skup. Ukoliko skup nije dobro balansiran, skup podataka može biti „prenaučen“ za jednu oznaku. Na primjeru detekcije lažnih vijesti, ukoliko skup sadrži puno više lažnih vijesti, model će imati problema sa detekcijom istinitih vijesti. *Train* skup služi za razvoj modela, odnosno model se trenira koristeći neki algoritam strojnog učenja kako bi mogao stvarati predikcije. S obzirom da je skup anotiran, model detektira uzorke za obje oznake (npr. istina i laž) i na temelju detekcija, stvara zaključke koje služe za predikcije nad novim podacima. Pošto ne bi imalo smisla evaluirati model sa istim podacima nad kojima je treniran, koristi se *test* skup za tu svrhu. *Test* skup sadrži nove, modelu ne viđene podatke, nad kojima on radi predikcije. Uspoređivanjem predikcija sa stvarnim vrijednostima, računaju se mjere evaluacije koje će biti opisane u nastavku.

Matrica konfuzije prikazuje predviđene i stvarne vrijednosti. Dimenzije matrice su 2x2 za probleme binarne klasifikacije. Slika 13 prikazuje primjer matrice konfuzije sa opisima svakog polja.



Slika 13: Primjer matrice konfuzije
Izvor: [55]

U prvo polje, označeno sa TP (engl. *true positive*), spadaju sve vrijednosti koje su predviđene kao pozitivne i jesu pozitivne. U polje označeno sa FP (engl. *false positive*) spadaju sve vrijednosti koje su predviđene kao pozitivne ali su zapravo negativne. U polje označeno FN (engl. *false negative*) spadaju sve vrijednosti koje su predviđene kao negativne ali su ustvari pozitivne. U zadnje polje, označeno sa TN (engl. *true negative*), spadaju sve vrijednosti koje su označene kao negativne i jesu negativne. Na temelju vrijednosti u poljima matrice konfuzije računaju se mjere za točnost, preciznost, odziv i F1-rezultat. Točnost se računa tako da podijelimo broj točnih predikcija sa ukupnim brojem predikcija. Točnost je bitna mjera ako je skup podataka dobro balansiran. Preciznost se računa tako da se ukupan broj točno predviđenih pozitivnih vrijednosti podijeli sa ukupnim brojem pozitivno predviđenih vrijednosti. Odziv se računa tako da se ukupan broj točno predviđenih podjeli sa ukupnim brojem pozitivnih (relevantnih) vrijednosti (TP + FN). F1-rezultat se računa sljedećom formulom:

$$2 * \frac{\text{preciznost} * \text{odaziv}}{\text{preciznost} + \text{odaziv}}$$

3. Eksperiment

Ovaj dio rada će obuhvatiti sve procese i alate koji su korišteni u praktičnom djelu rada. Opisati će se skupovi podataka i usporediti sa drugim skupovima koji se mogu koristiti za zadatak detekcije lažnih vijesti. Prikazat i objasniti će se NLP metode koje su korištene kao i algoritmi strojnog učenja.

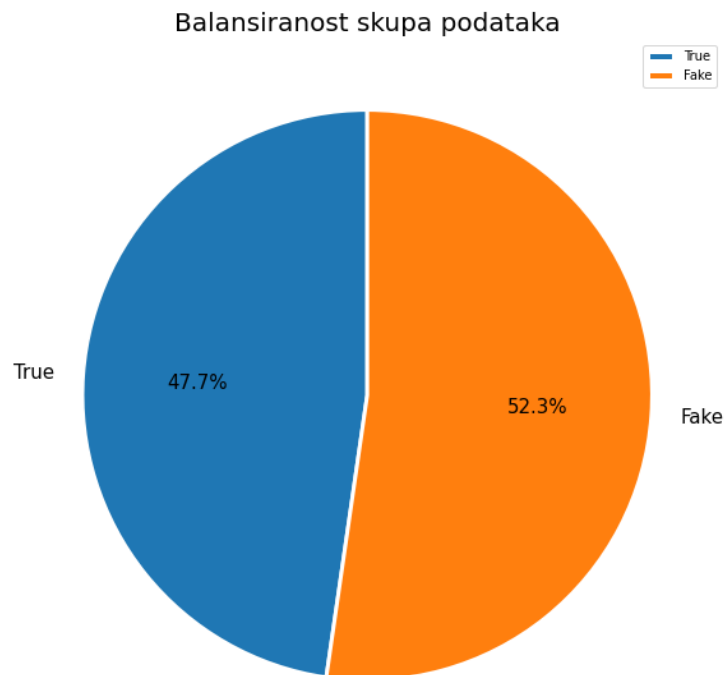
3.1. Python biblioteke

Python programski jezik se koristi za izradu projekta. Sadrži puno moćnih biblioteka koje olakšavaju rad programerima. Sve biblioteke je prvo potrebno instalirati unutar radnog okruženja. Za instaliranje biblioteka u ovom projektu se koristi „pip“. „Pip“ je Python-ov upravitelj biblioteka koji instalira biblioteke koje se nisu unaprijed instalirane u Python okruženju. U sklopu projekta koriste se sljedeće biblioteke: Pandas, Numpy, PIL, Wordcloud Re, Contractions, Matplotlib.pyplot, Nltk, Sklearn i Seaborn. Pandas biblioteka se koristi za učitavanje i manipulaciju skupom podataka. Učitavaju se csv datoteke i pretvore se u DataFrame objekt. Numpy biblioteka se koristi za stvaranje polja u kojem će biti pohranjena slika koja se koristi za prikaz oblaka riječi. PIL se koristi za učitavanje slike koja predstavlja „oblak“. Wordcloud biblioteka omogućuje stvaranje oblaka riječi. Re omogućuje korištenje regularnih izraza. Regularni izlazi omogućuju procese pripreme teksta tako što predstavljaju uzorke u tekstu koje je potrebno detektirati. Contractions biblioteka se koristi za razdavanje riječi koji su stvorene od kombinacije dviju riječi. Matplotlib.pyplot se koristi za prikaz grafova, kao i Seaborn biblioteka. Nltk biblioteka sadrži alate koji se koriste za obradu prirodnog jezika. Prvo se preuzima skup zaustavnih riječi (engl. *stopwords*). Učitavaju se alati za pretvaranje rečenica u tokene (engl. *tokenize*), svođenje riječi u njihov korijen i svođenje riječi u njihov kanonski oblik. Zadnja biblioteka koja se koristi je Sklearn. Prvo se koristi za podjelu skupa na skup za treniranje i testiranje. Također omogućuje pretvorbu riječi u vektore (engl. *vectorization*) i korištenje algoritama strojnog učenja.

3.2. Skupovi podataka

Postoje 3 skupa podataka koji će se koristiti za treniranje modela. Svi su detaljno opisani u poglavlju koje se odnosi na pregled postojećih skupova podataka. ISOT je prvi skup koji se koristi. Sastoji se od dvije datoteke: Fake.csv (23481 članaka koji su označeni kao lažne vijesti)

i True.csv (21417 članaka koji su označeni kao istinite vijesti). Slika 14 prikazuje balansiranost skupa podataka.



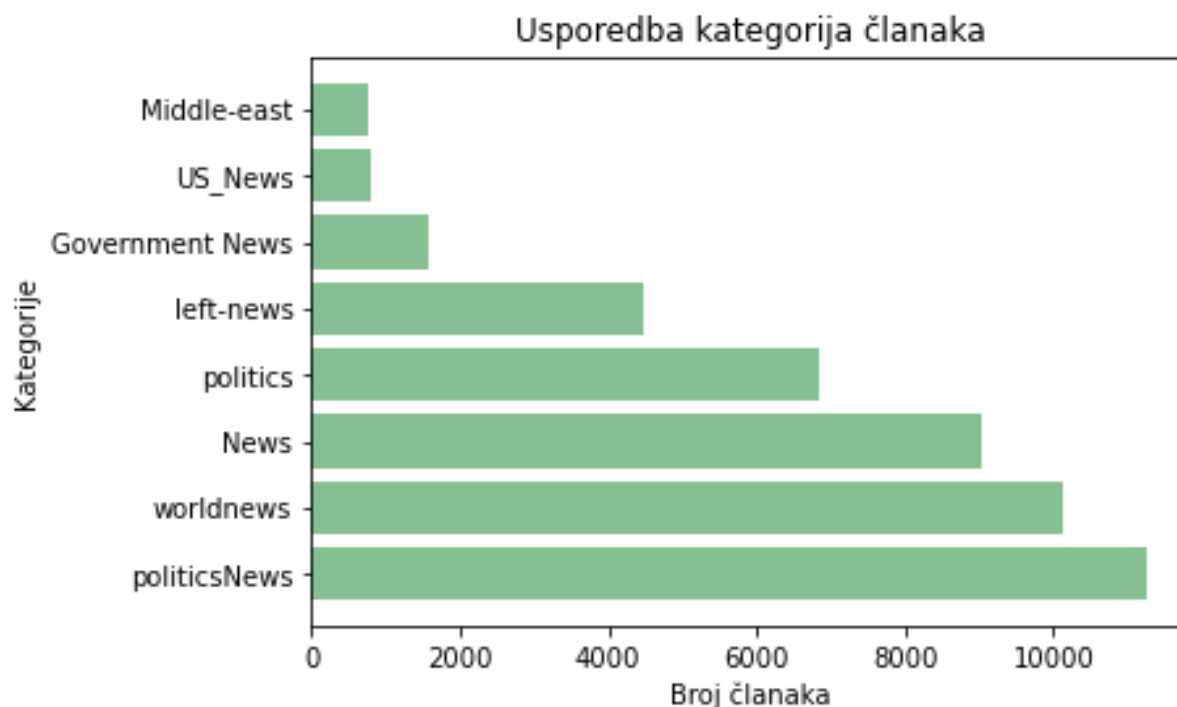
Slika 14: Balansiranost ISOT skupa podataka
Izvor: Vlastiti izvor

S obzirom da se skup sastoji od 2 csv datoteke, bilo ih je potrebno povezati u jedan skup za korištenje. Datoteke su učitane korištenjem Pandas biblioteke. Potom su povezane u jedan zajednički DataFrame koji predstavlja cjelokupni skup podataka. Slika 15 prikazuje izgled skupa podataka.

	title	text	subject	date	label
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	0
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	0
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	0
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	0
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	0

Slika 15: ISOT skup podataka
Izvor: Vlastiti izvor

Stupci u tablici predstavljaju naslov članka, cijeli tekst unutar članka, kategoriju u koju spada članak, datum objavljivanja i oznaku (0 označuje istinite a 1 lažne članke, odnosno vijesti). S obzirom da nedostaje nekoliko podataka, oni su zamijenjeni praznim poljem (stringom). Postoji 8 različitih kategorija u koje su svrstani članci. Slika 16 prikazuje omjer kategorija unutar skupa podataka.



Slika 16: Usporedba kategorija članaka
Izvor: vlastiti izvor

Najveći broj članaka spada u političke vijesti a najmanji broj u vijesti iz bliskog istoka. Bitno je spomenuti da istiniti članci spadaju samo u kategorije vijesti iz svijeta i političke vijesti a lažni članci u sve ostale kategorije. Svi članci su napisani između 2016. i 2017. godine. Drugi skup podataka, pod nazivom „Fake News“ je preuzet sa Kaggle web stranice. Sastoji se od 3 csv datoteke (train.csv, test.csv, submit.csv). Train skup se sastoji od 20.800 članaka, a test od 5200 članaka. Test skup nije anotiran ali se oznake za svaki članak unutar test skupa nalaze u submit skupu (5200 oznaka). S obzirom na računalnu zahtjevnost pripreme i treniranja podataka, koristi se samo train skup. On je dodatno podijeljen na skupove za treniranje i testiranje u omjeru 75% treniranje i 25% testiranje. Slika 17 prikazuje stupce i količinu podataka u svakom stupcu. Stupcima naslov, autor i tekst fali nekoliko podataka vezano za članke. Podaci koji nedostaju su zamijenjeni praznim stringom, kao i u prethodnom skupu. Samo id i label (oznaka) imaju sve podatke, što je i za očekivati u anotiranom skupu podataka. Skup se sastoji od 49.9% istinitih članaka i 50.1% lažnih članaka, što predstavlja skoro savršenu balansiranost. Za razliku od prethodnog skupa, ne postoje podatci za kategorije u koje spada članak, kao ni datum, ali postoje podatci za autore članaka.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20800 entries, 0 to 20799
Data columns (total 5 columns):
#   Column   Non-Null Count  Dtype
---  -
0   id        20800 non-null   int64
1   title     20242 non-null   object
2   author    18843 non-null   object
3   text      20761 non-null   object
4   label     20800 non-null   int64
dtypes: int64(2), object(3)
memory usage: 812.6+ KB
```

Slika 17: Kaggle Fake News skup podataka
Izvor: Vlastiti izvor

ReCOVeRY je naziv zadnjeg skupa koji će se koristiti. Sastoji se od 2 csv datoteke. Prva sadrži novinske članke a druge podatke sa Twitter-a. Koristiti će se samo prvi skup sa novinskim člancima. Sadrži stupce u kojima su pohranjeni linkovi, izdavačke kuće, datum izdavanja, autor, naslov, link naslove slike, politička pristranost, država, i pouzdanost (1 predstavlja pouzdane, a 0 nepouzdan članke). Oznake su zamijenjene kako bi bile iste kao u dva prethodna skupa (0 za pouzdane, 1 za nepouzdan). Značajno je manji od prethodna 2 skupa, sa ukupno 2029 članaka. To je otprilike deset puta manje od prethodnog skupa. Također nije toliko dobro balansirano kao prethodna 2 skupa. 67.2% članaka su pouzdani a 32.8% članaka predstavljaju nepouzdan članke. S pozitivne strane, većina članaka nema nedostajućih podataka, a mali broj onih koji fale su većinom podatci koji nisu toliko potrebni za treniranje skupa (npr. nedostaje podatak o državi).

3.3. Priprema podataka

Svaki skup podataka koji se koristi za klasifikaciju je potrebno prilagoditi kako bi se dobili što bolji rezultati predikcija. U prvom skupu podataka (ISOT) koristiti će se stupci naslov i tekst članka. Stvara se novi stupac u kojem će se spremiti naslov i tekst. Zatim se stvaraju funkcije koje će prilagoditi tekst u novom stupcu. Slika 18 prikazuje funkcije koje se koriste za pripremu teksta. Prva funkcija uklanja sve znakove osim slova iz teksta. Samo je ostavljen apostrof zbog riječi kao npr. „don't“. Druga funkcija razdvaja riječi poput „don't“ na „do“ i „not“. Zatim je potrebno razdvojiti rečenice na zasebne riječi ili tokene(engl. *tokens*), pomoću

naredbe „word_tokenize“. Treća funkcija uklanja kratke riječi, odnosno one koje imaju manje od 3 slova zato što ne doprinose preveliku vrijednost.

```
def cleaning(text):
    text = re.sub('[^a-zA-Z]', ' ', text)
    text = text.lower()
    return text
✓ 0.9s
```

Uklanjanje contractions-a (riječi koje su kombinacija dviju riječi npr. "do" + "not" = "don't")

```
def remove_contractions(text):
    return ' '.join([contractions.fix(word) for word in text.split()])
✓ 0.6s
```

Uklanjanje kratkih riječi (manje od 3 slova)

```
def remove_small_words(text):
    return [x for x in text if len(x) > 3 ]
✓ 0.7s
```

Slika 18: Funkcije za pripremu teksta
Izvor: Vlastiti izvor

Slika 19 prikazuje funkciju za uklanjanje zaustavnih riječi. Zaustavne riječi su riječi koje nemaju veliku važnost a pojavljuju se često su skupu podataka. Na slici su prikazani neki primjeri zaustavnih riječi u engleskom jeziku.

```
Funkcija za uklanjanje stopwords-a (zaustavnih riječi)
```

```
baza['r_stopwords']=baza['r_tokens'].apply(lambda x:([word for word in x if word not in (stopwords.words('english'))])
✓ 148m 54.6s
```

```
stopwords.words('english')
✓ 0.1s
```

Output exceeds the [size limit](#). Open the full output data [in a text editor](#)

```
['i',
 'me',
 'my',
 'myself',
 'we',
 'our',
 'ours',
 'ourselves',
 'you',
 "you're",
 "you've",
 "you'll",
 "you'd",
 'your',
 'yours',
```

Slika 19: Funkcija za uklanjanje zaustavnih riječi
Izvor: Vlastiti izvor

Sljedeći skup funkcija spada u tehnike normalizacije teksta. Slika 20 prikazuje 3 funkcije za normalizaciju teksta. Prve dvije funkcije svode riječi na njihov korijen. Uklanjaju prefikse i sufikse riječi, ali nažalost nisu savršeni. Npr. engleske riječi „university“ i „universe“ mogu imati isti korijen „univers“, ali imaju potpuno različita značenja. Prikazana su 2 različita algoritma: Porter i Snowball stemmer. Snowball stemmer se smatra poboljšanom verzijom Porter stemmera, ali se ne koriste u radu jer donose lošije rezultate nego bez njihova korištenja.

```
def porter_stemmer(text):
    stemmed_tokens = [ps.stem(token) for token in text]
    return ' '.join(stemmed_tokens)
✓ 0.6s
```

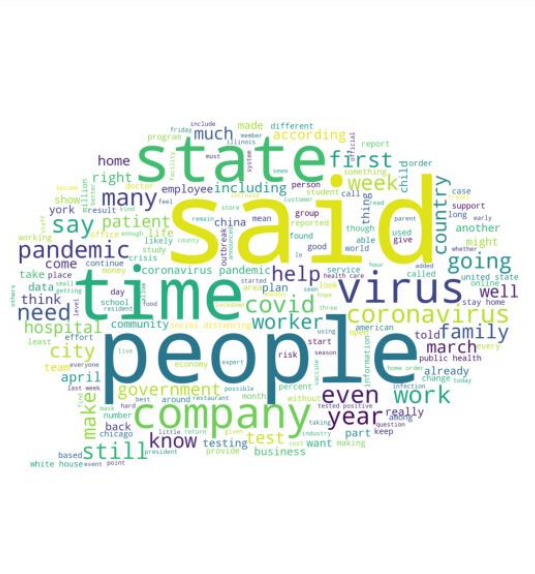
```
def snowball_stemmer(text):
    stemmed_tokens = [ss.stem(token) for token in text]
    return ' '.join(stemmed_tokens)
✓ 0.6s
```

```
def lemmatize_text(text):
    lemmantized = ' '.join([wnl.lemmatize(words) for words in text])
    return lemmantized
✓ 0.1s
```

Slika 20: Funkcije za normalizaciju teksta
Izvor: Vlastiti izvor

Umjesto njih, koristi se treća funkcija sa slike koja pretvara riječi u njihov kanonski oblik. Za razliku od prve dvije, riječi su svedene na točan gramatički oblik. Prilikom pokretanja funkcije, tokeni ulaze u funkciju a kao izlaz se spajaju natrag u rečenice. Nakon izvođena svih funkcija, skup podataka je spreman za treniranje. Sve funkcije su primijenjene na sva tri skupa. Jedine razlike su u stupcima koji postoje unutar skupova. Drugi skup „Fake News“ koristi stupce autor, naslov i tekst članka. Svi podatci unutar tih skupova su spojeni u novi stupac. Isto je napravljeno i za zadnji skup ReCOVeRY. Na slici 21 je prikazana usporedba oblaka riječi za istinite i lažne vijesti unutar ReCOVeRY skupa. Veličina fonta predstavlja broj pojavljivanja riječi u skupu. Zanimljivo je za primijetiti kako se riječi kao „virus“ i „covid“ puno više pojavljuju u skupu lažnih vijesti.

Oblak riječi za istinite vijesti



Oblak riječi za lažne vijesti



Slika 21: Usporedba istinitih i lažnih vijesti (ReCOvery)

Izvor: Vlastiti izvor

3.4. Algoritmi strojnog učenja

Nakon što je tekst pretvoren u optimalno stanje, može se početi sa treniranjem skupova. Prvo se definira varijabla X , u koju se pohranjuju vrijednosti iz stupca u kojem se nalazi obrađeni tekst, i varijabla y , u koju se pohranjuju vrijednosti koje predstavljaju istinitost vijesti (0 ili 1). S obzirom da računalo ne razumije ljudski jezik, potrebno je prevesti tekst u računalo razumljiv format. To se izvodi pretvorbom teksta u vektore. Taj proces se naziva vektorizacija (engl. *vectorization*). Koristi se TF-IDF tehnika vektorizacije. Slika 22 prikazuje korištenje TF-IDF vektorizatora. U funkciju *fit()* se pohranjuju vrijednosti iz varijable X pa se potom tekstualne vrijednosti pretvaraju u vektore funkcijom *transform()* i spremaju u varijablu X .

```
vectorizer = TfidfVectorizer()
vectorizer.fit(X)
X=vectorizer.transform(X)
```

Slika 22: TF-IDF vektorizator

Izvor: Vlastiti izvor

Sljedeći korak je definiranje skupa za treniranje i testiranje. To se izvodi koristeći funkciju *train_test_split()*. Sva tri skupa su podijeljeni u omjeru 75 train : 25 test. Prvi algoritam koji se koristi je logistička regresija. Slika 23 prikazuje algoritam logističke regresije. U model se

učitavaju X_train i y_train varijable. Nakon što model završi treniranje, izvodi proces predikcije nad X_test varijablom i uspoređuje predikcije sa y_test varijablom kako bi izračunao točnost modela.

```
model_LR = LogisticRegression()  
model_LR.fit(X_train, y_train)  
X_test_pred_LR = model_LR.predict(X_test)
```

Slika 23: Model logističke regresije
Izvor: Vlastiti izvor

Stablo odluka je drugi algoritam koji se koristi. Postupak je sličan prošlom modelu, samo je dodatno definirano da se koristi mjera entropije. Slika 24 prikazuje model stabla odluke.

```
model_DT = DecisionTreeClassifier(criterion='entropy')  
model_DT.fit(X_train, y_train)  
X_test_pred_DT = model_DT.predict(X_test)
```

Slika 24: Model stabla odluke
Izvor: Vlastiti izvor

Slika 25 pokazuje model slučajnih šuma. Postupak je sličan prošlim modelima. Također je potrebno definirati korištenje entropije.

```
model_RF = RandomForestClassifier(criterion='entropy')  
model_RF.fit(X_train, y_train)  
X_test_pred_RF = model_RF.predict(X_test)
```

Slika 25: Model slučajne šume
Izvor: Vlastiti izvor

SVM je sljedeći algoritam koji se koristio. Potrebno je postaviti linearni kernel i definirati da se koristi za klasifikaciju. Slika 26 prikazuje SVM model.

```
model_SVM = svm.SVC(kernel='linear')  
model_SVM.fit(X_train, y_train)  
X_test_pred_SVM = model_SVM.predict(X_test)
```

Slika 26: SVM model
Izvor: Vlastiti izvor

KNN je sljedeći algoritam koji se koristio. Potrebno je definirati broj susjeda, postavljeno je na 2. Također se definira korištenje cosine udaljenosti i mjerenje udaljenosti između točaka. Slika 27 prikazuje KNN model.

```
model_KNN = KNeighborsClassifier(n_neighbors=2, metric='cosine',  
                                weights='distance')  
model_KNN.fit(X_train, y_train)  
X_test_pred_KNN = model_KNN.predict(X_test)
```

Slika 27: KNN model

Izvor: Vlastiti izvor

Naivni Bayes je predzadnji korišteni algoritam. Koristi se multinomni Naive Bayes algoritam. Nije potrebno podešavati dodatne parametre. Slika 28 prikazuje Naivni Bayes model.

```
model_NB = MultinomialNB()  
model_NB.fit(X_train, y_train)  
X_test_pred_NB = model_NB.predict(X_test)
```

Slika 28: Naivni Bayes model

Izvor: Vlastiti izvor

Zadnji algoritam korišten u ovom radu je pasivno agresivni algoritam. Postavljen je maksimalni broj iteracija na 100. Slika 29 prikazuje pasivno agresivni model.

```
model_PA = PassiveAggressiveClassifier()  
model_PA.fit(X_train, y_train)  
X_test_pred_PA = model_PA.predict(X_test)
```

Slika 29: Pasivno agresivni model

Izvor: Vlastiti izvor

4. Rezultati

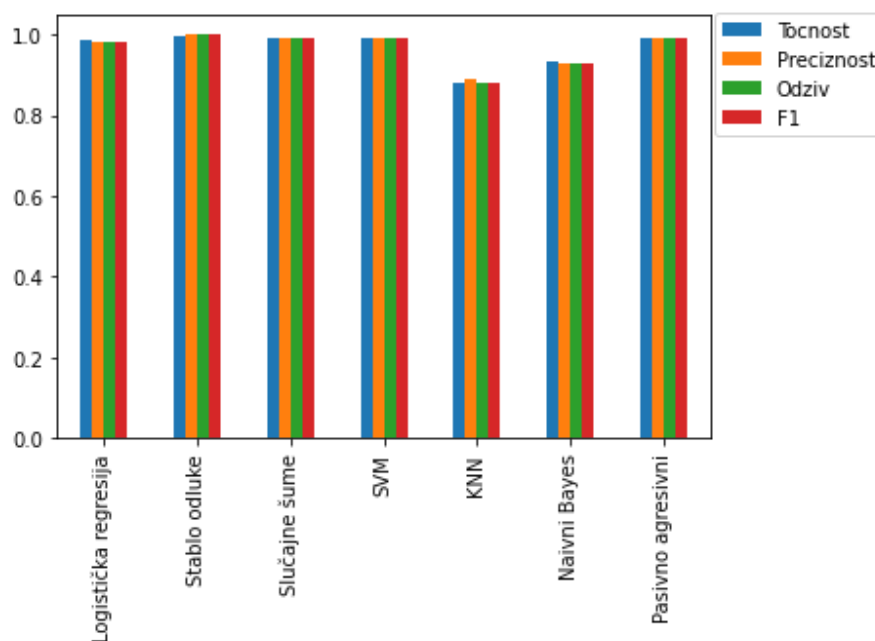
U ovom djelu rada će se prikazati dobiveni rezultati za svaki skup podataka posebno. Također će se usporediti rezultati sa rezultatima koji su postignuti u drugim radovima.

Tablica 2 prikazuje rezultate za sve trenirane modele strojnog učenja. Nisu sortirani po točnosti nego su prikazani redoslijedom kojim su bili trenirani. Rezultati su prikazani za skup podataka ISOT. Najveću točnost su postigli stablo odluke, SVM i pasivno agresivni modeli. Također su postigli najbolje rezultate za preciznost, odziv i F1. Najlošije rezultate je postigao KNN model. Generalno su svi modeli postigli jako dobre rezultate za skup podataka ISOT.

Algoritam	Točnost	Preciznost	Odziv	F1
Logistička regresija	0.9847	0.98	0.98	0.98
Stablo odluke	0.9963	1.00	1.00	1.00
Slučajne šume	0.9898	0.99	0.99	0.99
SVM	0.9923	0.99	0.99	0.99
KNN	0.8804	0.89	0.88	0.88
Naivni Bayes	0.9338	0.93	0.93	0.93
Pasivno agresivni	0.9928	0.99	0.99	0.99

Tablica 2: Točnost modela za skup podataka ISOT
Izvor: Vlastiti izvor

Slika 30 prikazuje usporedbu točnosti, preciznosti, odziva i F1 za modele trenirane na skupu ISOT.



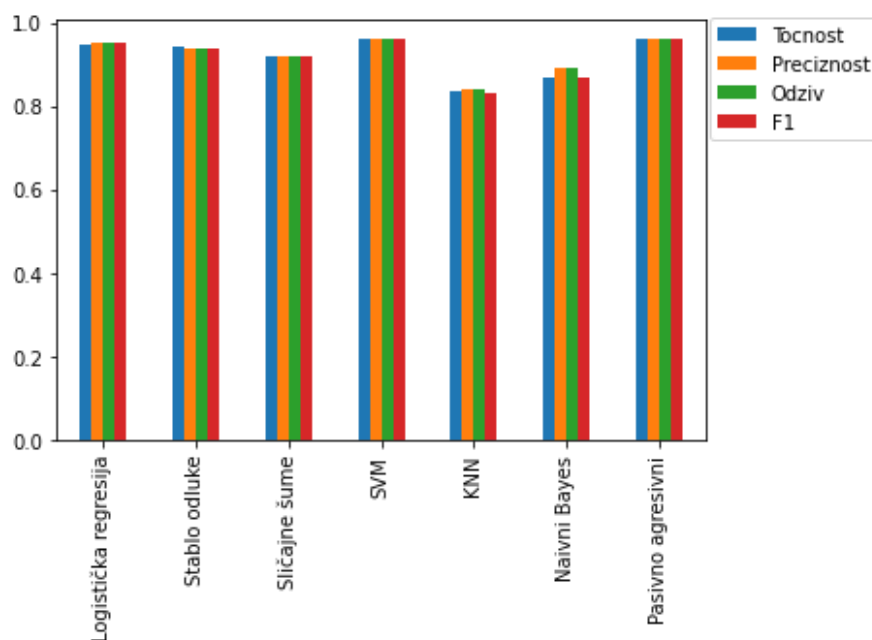
Slika 30: Usporedba mjera (ISOT)
Izvor: Vlastiti izvor

Tablica 3 prikazuje rezultate za sve trenirane modele strojnog učenja. Nisu sortirani po točnosti nego su prikazani redosljedom kojim su bili trenirani. Rezultati su prikazani za skup podataka Fake News. Nabolje rezultate su postigli SVM i pasivno agresivni modeli. Najgori rezultat je postigao KNN model. Ponovno su modeli postigli jako dobre rezultate iako su malo lošiji od modela treniranih na skupu ISOT.

Algoritam	Točnost	Preciznost	Odziv	F1
Logistička regresija	0.9463	0.95	0.95	0.95
Stablo odluke	0.9433	0.94	0.94	0.94
Slučajne šume	0.9198	0.92	0.92	0.92
SVM	0.9606	0.96	0.96	0.96
KNN	0.8350	0.84	0.84	0.83
Naivni Bayes	0.8688	0.89	0.87	0.87
Pasivno agresivni	0.9596	0.96	0.96	0.96

Tablica 3: Točnost algoritama za skup podataka Fake News
Izvor: Vlastiti izvor

Slika 31 prikazuje usporedbu točnosti, preciznosti, odziva i F1 za modele trenirane na skupu Fake News.



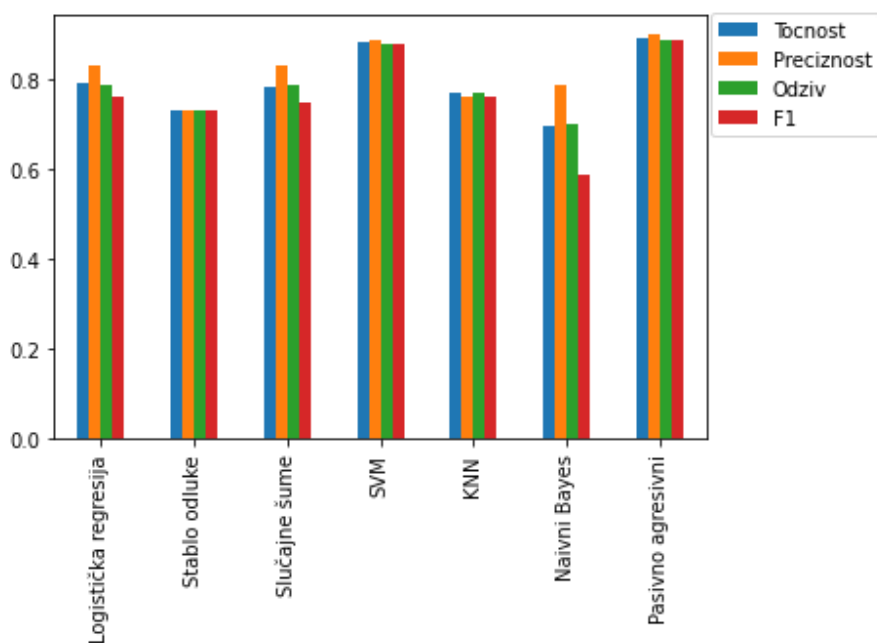
Slika 31: Usporedba mjera (Fake News)
Izvor: Vlastiti izvor

Tablica 4 prikazuje rezultate za sve trenirane modele strojnog učenja. Algoritmi nisu sortirani po točnosti nego su prikazani redosljedom kojim su bili trenirani. Rezultati su prikazani za skup podataka ReCOVeRY. Najbolje rezultate su postigli SVM i pasivno agresivni modeli, kao i za skup Fake News, ali su rezultati lošiji nego za prethodna 2 skupa podataka.

Algoritam	Točnost	Preciznost	Odziv	F1
Logistička regresija	0.7913	0.83	0.79	0.76
Stablo odluke	0.7303	0.73	0.73	0.73
Slučajne šume	0.7854	0.83	0.79	0.75
SVM	0.8819	0.89	0.88	0.88
KNN	0.7717	0.76	0.77	0.76
Naivni Bayes	0.6969	0.79	0.70	0.59
Pasivno agresivni	0.8937	0.90	0.89	0.89

Tablica 4: Točnost algoritama za skup podataka ReCOVery
Izvor: Vlastiti izvor

Slika 32 prikazuje usporedbu točnosti, preciznosti, odziva i F1 za modele trenirane na ReCOVery skupu.



Slika 32: Usporedba mjera (ReCOVery)
Izvor: Vlastiti izvor

4.1. Usporedba rezultata s ostalim radovima

Najbolje rezultate imaju svi modeli trenirani na ISOT skupu, što je i očekivano s obzirom da je najveći skup i relativno dobro je balansiran. U radu [21] autor M. Yousaf i suradnici postižu jako dobre rezultate za modele trenirane na ISOT skupu. Najvišu točnost postižu slučajne šume (99%) i SVM (98%), što je približno jednako modelima treniranim u ovom radu, iako je SVM model (99%) minimalno bolji od slučajnih šuma (98%). Ti modeli su postigli najbolje rezultate i za ostale mjere. Model slučajne šume, u radu navedenih autora,

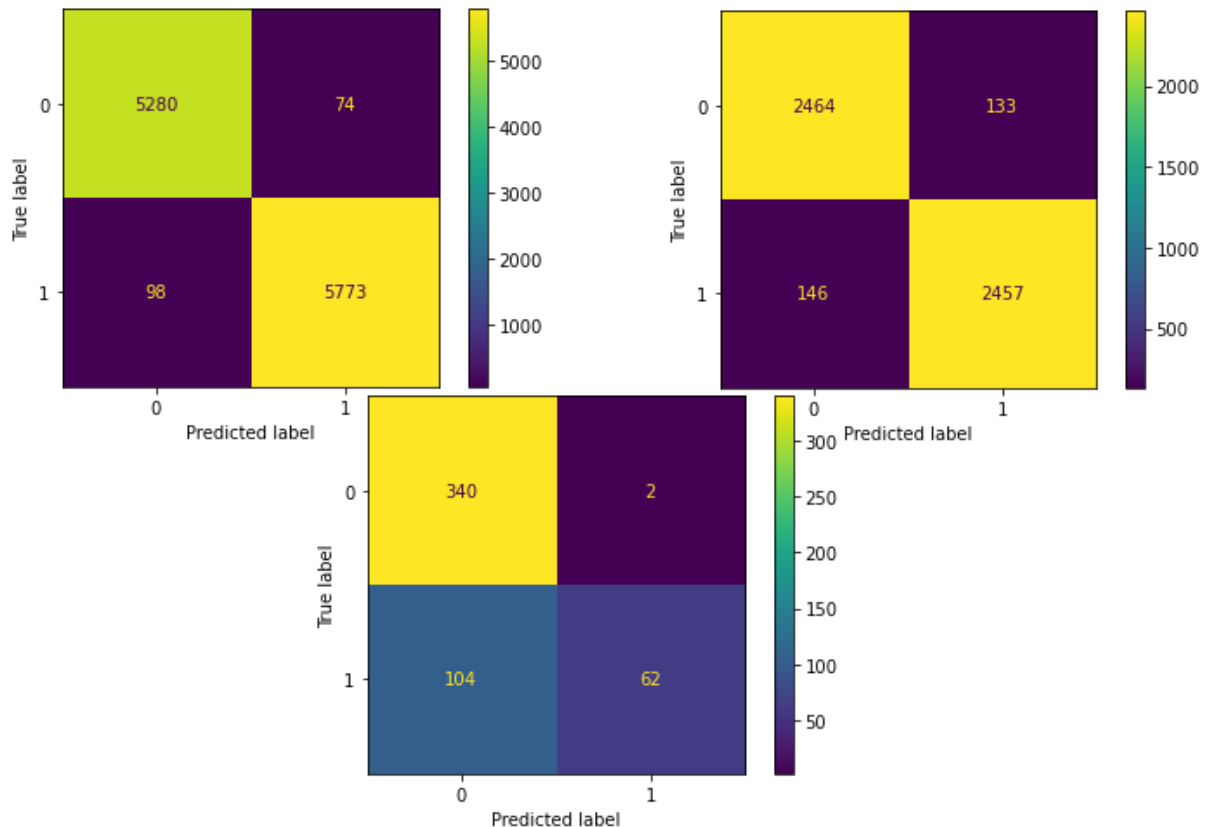
postiže rezultate za preciznost: 99%, odziv: 100% i F1: 99% što je gotovo jednako modelu slučajnih šuma u ovom radu, osim odziva koji iznosi 99%. Suprotno tome, SVM model u ovom radu postiže minimalno bolje rezultate za navedene mjere od SVM modela iz navedenog rada.

Modeli trenirani na skupu Fake New imaju lošije rezultate ali s obzirom da je skup podataka otprilike duplo manji od skupa ISOT, rezultati su zadovoljavajući. Autori rada [21], također treniraju modele i na ovom skupu podataka. Najveću točnost postižu koristeći stablo odluke (94%) što je jednako točnosti modela stabla odluke u ovom radu. Zanimljivo je za usporediti SVM model korišten u njihovom radu, koji postiže točnost od 37%, što je puno manje od točnosti SVM modela koji je postigao najbolje rezultate u ovom radu (96%). Najgoru točnost, u njihovom radu, postiže KNN model sa 28%. U ovom radu, KNN model također postiže najgore rezultate za skup Fake News, sa točnosti 84%, što je ipak puno više u usporedbi sa njihovim radom. S obzirom da je skup korišten u svrhu natjecanja, dostupni su najbolji postignuti rezultati [23]. Pobjednik natjecanja je postigao točnost od 99%. SVM model iz ovog rada bi bio šesti na rang ljestvici.

Skup ReCOVery je puno manji od druga dva skupa pa je i očekivano da će imati najgore rezultate. Također nije dobro balansiran sa otprilike duplo više članaka koji su pouzdani, pa se može pretpostaviti da je puno uspješniji u detekciji članaka koji su pouzdani nego nepouzdan, što je vidljivo na prikazu matrica konfuzije u nastavku rada. Autori u radu [29] također postižu bolje rezultate za detekciju pouzdanih vijesti, što potvrđuje navedenu pretpostavku. Modele treniraju koristeći neke od algoritama koji su korišteni u ovom radu, kao stablo odluke, SVM, logistička regresija itd. Prikazali su rezultate samo za najbolje modele. Model stabla odluke je postigao rezultate za preciznost: 78%, odziv: 77% i F1: 78% za pouzdane vijesti. Za nepouzdan vijesti je postigao rezultate za preciznost: 54%, odziv: 55% i F1: 55%. U ovom radu, model stabla odluke postiže približno jednake rezultate za pouzdane vijesti. Za nepouzdan vijesti, rezultati su minimalno bolji. Preciznost iznosi: 59%, odziv: 56% i F1: 58%. Nisu prikazali rezultate točnosti modela, jer bi mogli biti zavaravajući, zbog loše balansiranosti skupa.

4.2. Matrice konfuzije za logističku regresiju

Slika 33 prikazuje matrice konfuzije za sva 3 skupa podataka. Prva predstavlja skup ISOT, druga skup Fake News a treća skup ReCOVerry.

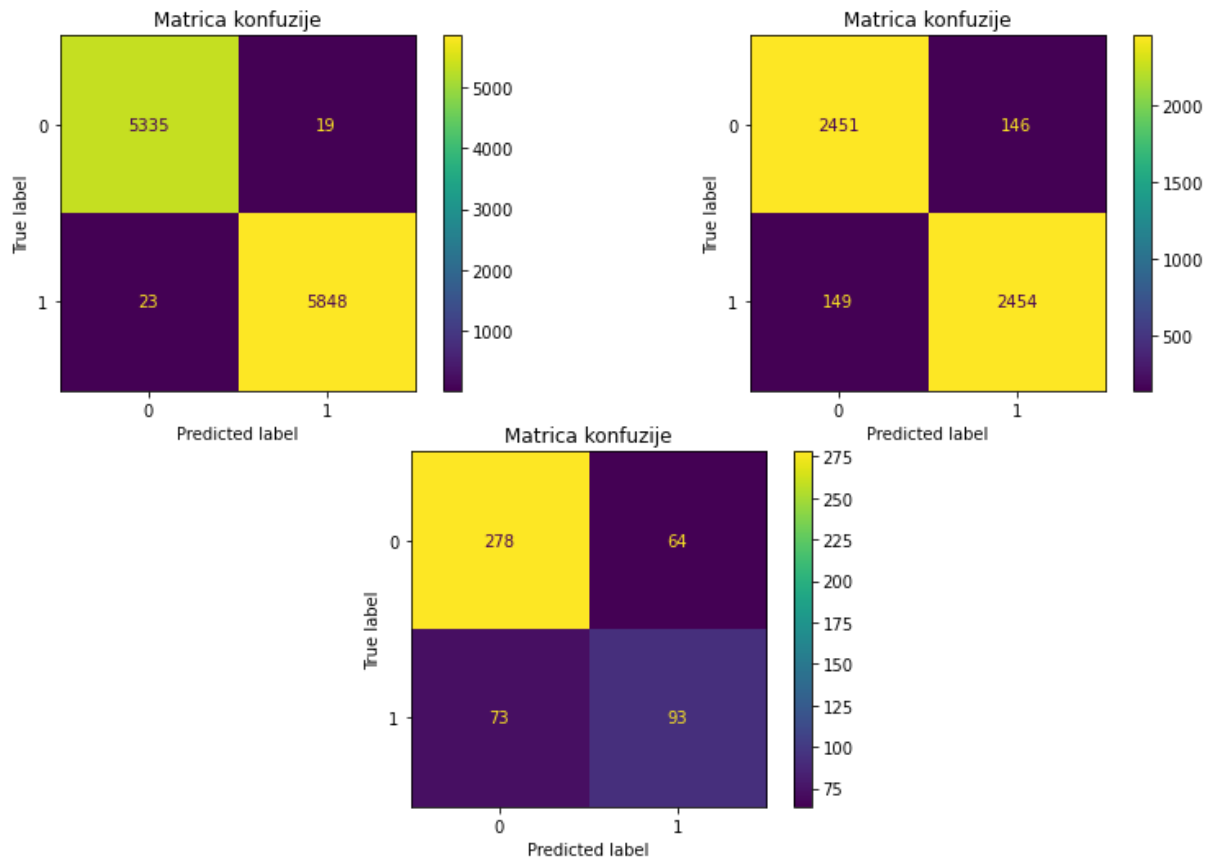


Slika 33: Matrice konfuzije za logističku regresiju
Izvor: Vlastiti izvor

Za skup ISOT, broj točnih predikcija za pozitivne vrijednosti iznosi 5280, odnosno točno je predvidio 5280 članaka kao istinite vijesti. 74 članaka je predvidio kao istinite vijesti ali su lažne, 98 članaka je predvidio kao lažne vijesti ali su istinite i 5773 članaka je točno predvidio kao lažne vijesti. Za skup Fake News, broj točnih predikcija za pozitivne vrijednosti iznosi 2464, odnosno točno je predvidio 2464 članaka kao istinite vijesti. 133 članaka je predvidio kao istinite vijesti ali su lažne, 146 članaka je predvidio kao lažne vijesti ali su istinite i 2457 članaka je točno predvidio kao lažne vijesti. Za skup ReCOVerry, broj točnih predikcija za pozitivne vrijednosti iznosi 340, odnosno točno je predvidio 340 članaka kao istinite vijesti. 2 članaka je predvidio kao istinite vijesti ali su lažne, 104 članaka je predvidio kao lažne vijesti ali su istinite i 62 članaka je točno predvidio kao lažne vijesti. Omjer TN i FN vrijednosti, pokazuje da je model ima problema sa detekcijom lažnih vijesti, iako ima relativno visoku točnost (80%).

4.3. Matrice konfuzije za stablo odluka

Slika 34 prikazuje matrice konfuzije za sva 3 skupa podataka. Prva predstavlja skup ISOT, druga skup Fake News a treća skup ReCOVerry

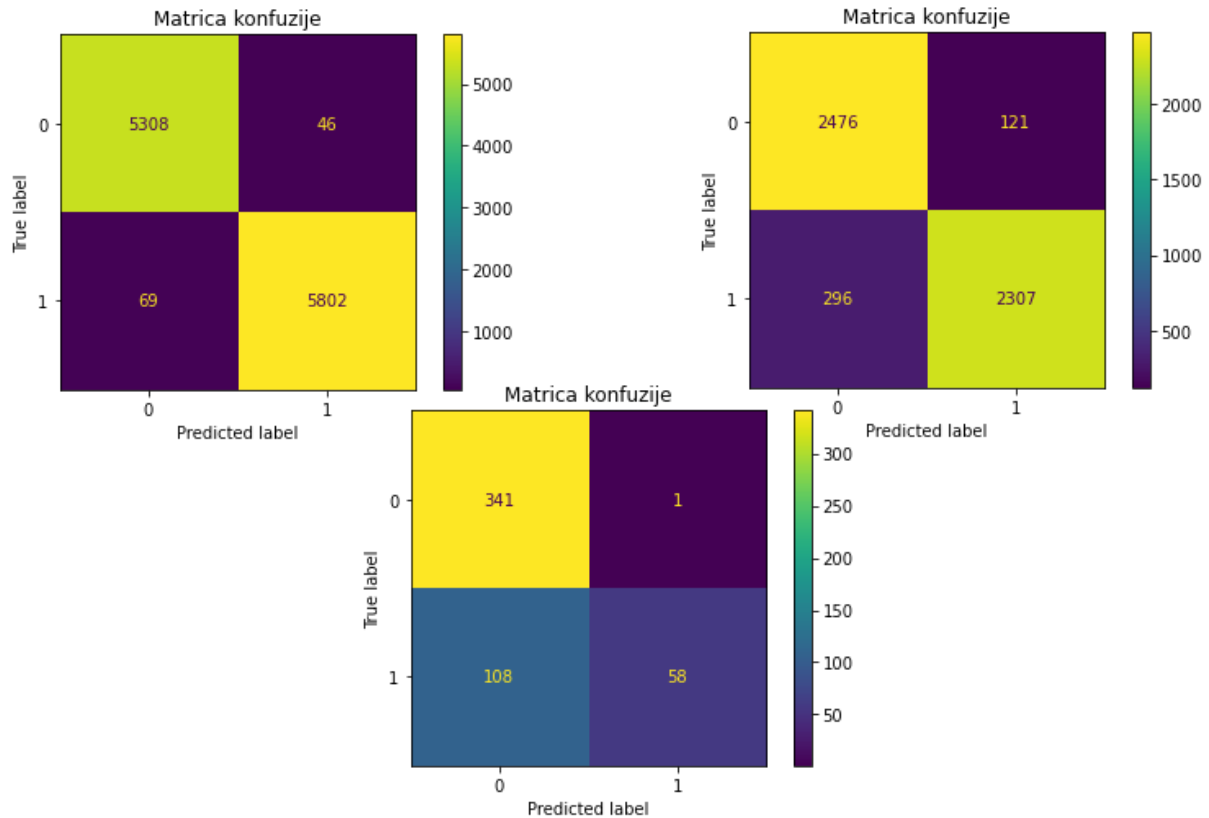


Slika 34: Matrica konfuzije za stablo odluke
Izvor: Vlastiti izvor

Za skup ISOT, broj točnih predikcija za pozitivne vrijednosti iznosi 5335, odnosno točno je predvidio 5335 članaka kao istinite vijesti. 19 članaka je predvidio kao istinite vijesti ali su lažne, 23 članaka je predvidio kao lažne vijesti ali su istinite i 5848 članaka je točno predvidio kao lažne vijesti. Za skup Fake News, broj točnih predikcija za pozitivne vrijednosti iznosi 2451, odnosno točno je predvidio 2451 članaka kao istinite vijesti. 146 članaka je predvidio kao istinite vijesti ali su lažne, 149 članaka je predvidio kao lažne vijesti ali su istinite i 2545 članaka je točno predvidio kao lažne vijesti. Za skup ReCOVerry, broj točnih predikcija za pozitivne vrijednosti iznosi 278, odnosno točno je predvidio 278 članaka kao istinite vijesti. 64 članaka je predvidio kao istinite vijesti ali su lažne, 73 članaka je predvidio kao lažne vijesti ali su istinite i 93 članaka je točno predvidio kao lažne vijesti. Omjer TN i FN je bolji od vrijednosti postignutih logističkom regresijom, ali i dalje je visok broj FN vrijednosti.

4.4. Matrice konfuzije za slučajne šume

Slika 35 prikazuje matrice konfuzije za sva 3 skupa podataka. Prva predstavlja skup ISOT, druga skup Fake News a treća skup ReCOVeRY.

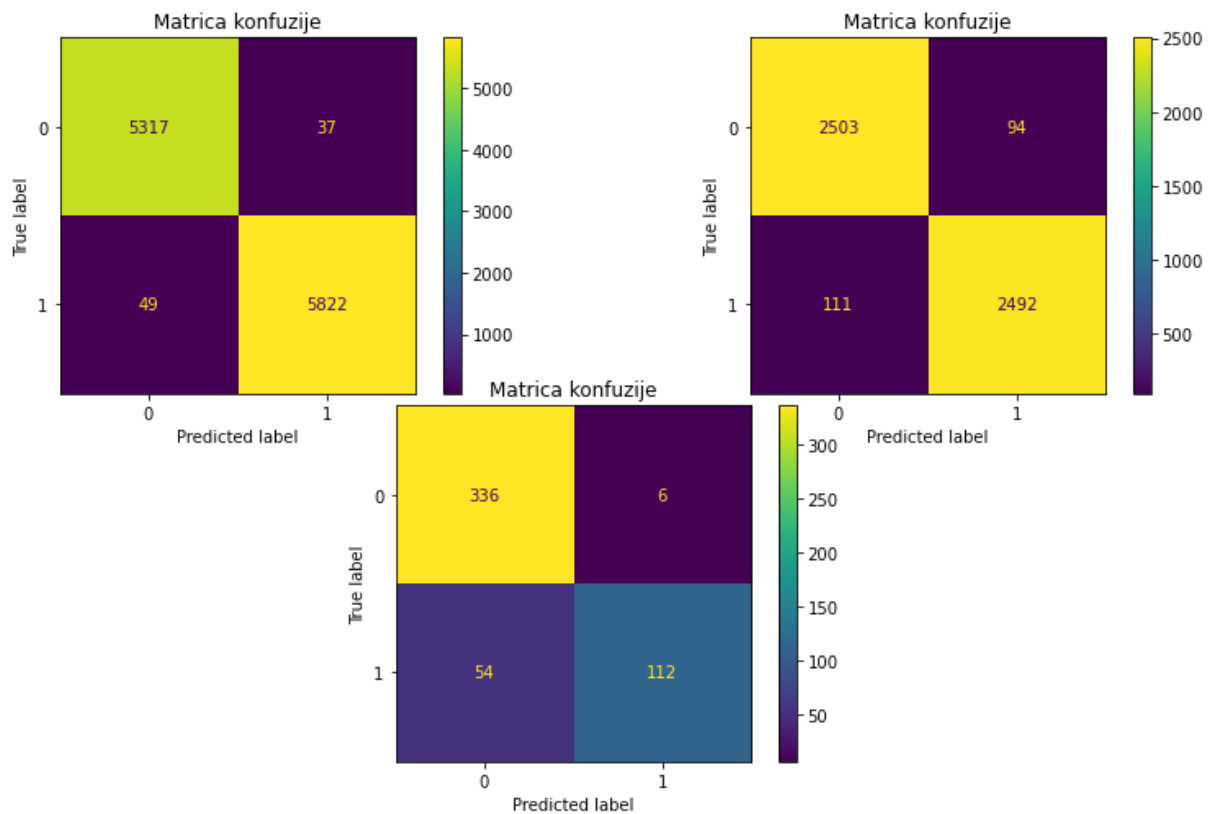


Slika 35: Matrica konfuzije za slučajne šume
Izvor: Vlastiti izvor

Za skup ISOT, broj točnih predikcija za pozitivne vrijednosti iznosi 5308, odnosno točno je predvidio 5308 članaka kao istinite vijesti. 46 članaka je predvidio kao istinite vijesti ali su lažne, 69 članaka je predvidio kao lažne vijesti ali su istinite i 5802 članaka je točno predvidio kao lažne vijesti. Za skup Fake News, broj točnih predikcija za pozitivne vrijednosti iznosi 2476, odnosno točno je predvidio 2476 članaka kao istinite vijesti. 121 članaka je predvidio kao istinite vijesti ali su lažne, 296 članaka je predvidio kao lažne vijesti ali su istinite i 2307 članaka je točno predvidio kao lažne vijesti. Za skup ReCOVeRY, broj točnih predikcija za pozitivne vrijednosti iznosi 341, odnosno točno je predvidio 341 članaka kao istinite vijesti. 1 članaka je predvidio kao istinite vijesti ali su lažne, 108 članaka je predvidio kao lažne vijesti ali su istinite i 58 članaka je točno predvidio kao lažne vijesti. Omjer TN i FN vrijednosti lošiji nego kod stabla odluke iako postiže bolje rezultate za sve mjere. Razlog tome je uspješnija detekcija istinitih vijesti.

4.5. Matrice konfuzije za SVM

Slika 36 prikazuje matrice konfuzije za sva 3 skupa podataka. Prva predstavlja skup ISOT, druga skup Fake News a treća skup ReCOVeRY.



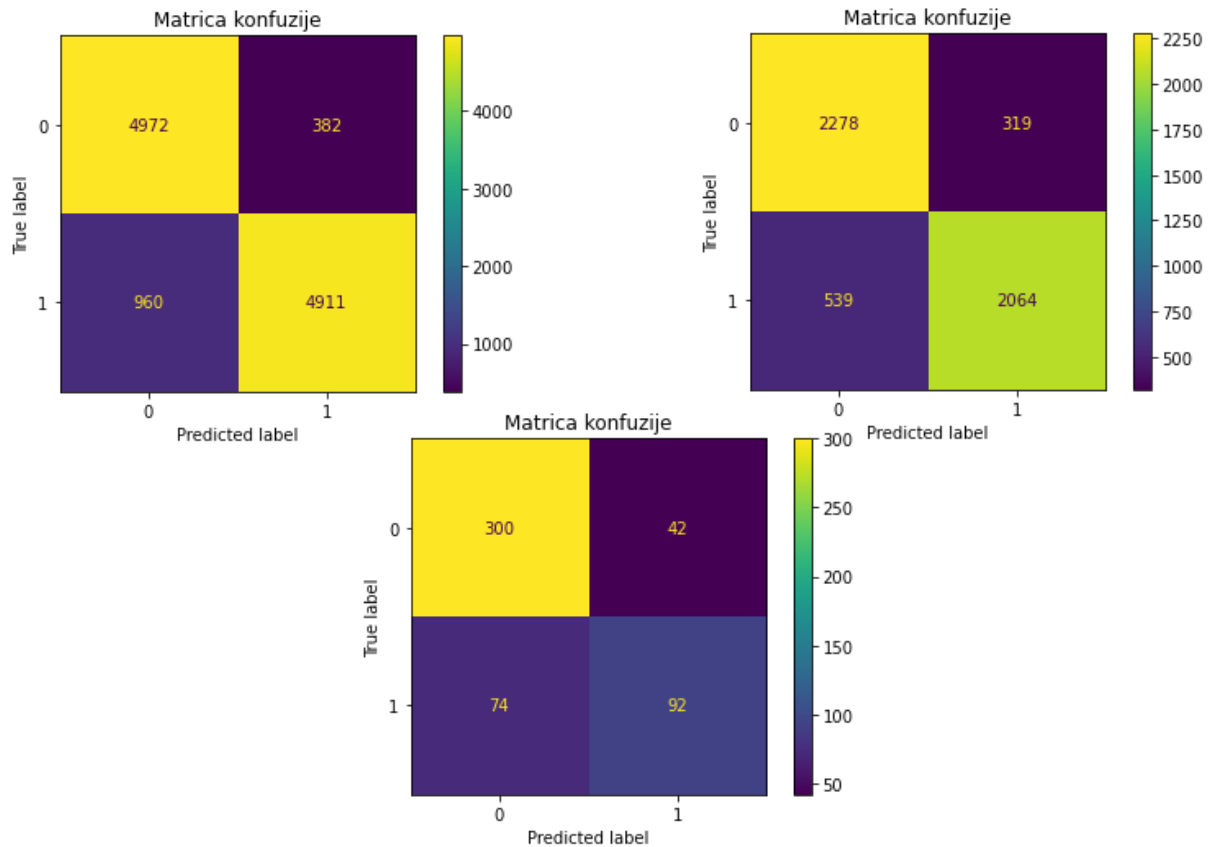
Slika 36: Matrica konfuzije za SVM

Izvor: Vlastiti izvor

Za skup ISOT, broj točnih predikcija za pozitivne vrijednosti iznosi 5317, odnosno točno je predvidio 5317 članaka kao istinite vijesti. 37 članaka je predvidio kao istinite vijesti ali su lažne, 49 članaka je predvidio kao lažne vijesti ali su istinite i 5822 članaka je točno predvidio kao lažne vijesti. Za skup Fake News, broj točnih predikcija za pozitivne vrijednosti iznosi 2503, odnosno točno je predvidio 2503 članaka kao istinite vijesti. 94 članaka je predvidio kao istinite vijesti ali su lažne, 111 članaka je predvidio kao lažne vijesti ali su istinite i 2492 članaka je točno predvidio kao lažne vijesti. Za skup ReCOVeRY, broj točnih predikcija za pozitivne vrijednosti iznosi 336, odnosno točno je predvidio 336 članaka kao istinite vijesti. 6 članaka je predvidio kao istinite vijesti ali su lažne, 54 članaka je predvidio kao lažne vijesti ali su istinite i 112 članaka je točno predvidio kao lažne vijesti. Postiže puno veći broj točno detektiranih lažnih vijesti, u usporedbi sa prijašnjim algoritima. Generalno je postigao dobre rezultate za svaki skup podataka.

4.6. Matrice konfuzije za KNN

Slika 37 prikazuje matrice konfuzije za sva 3 skupa podataka. Prva predstavlja skup ISOT, druga skup Fake News a treća skup ReCOVerry.



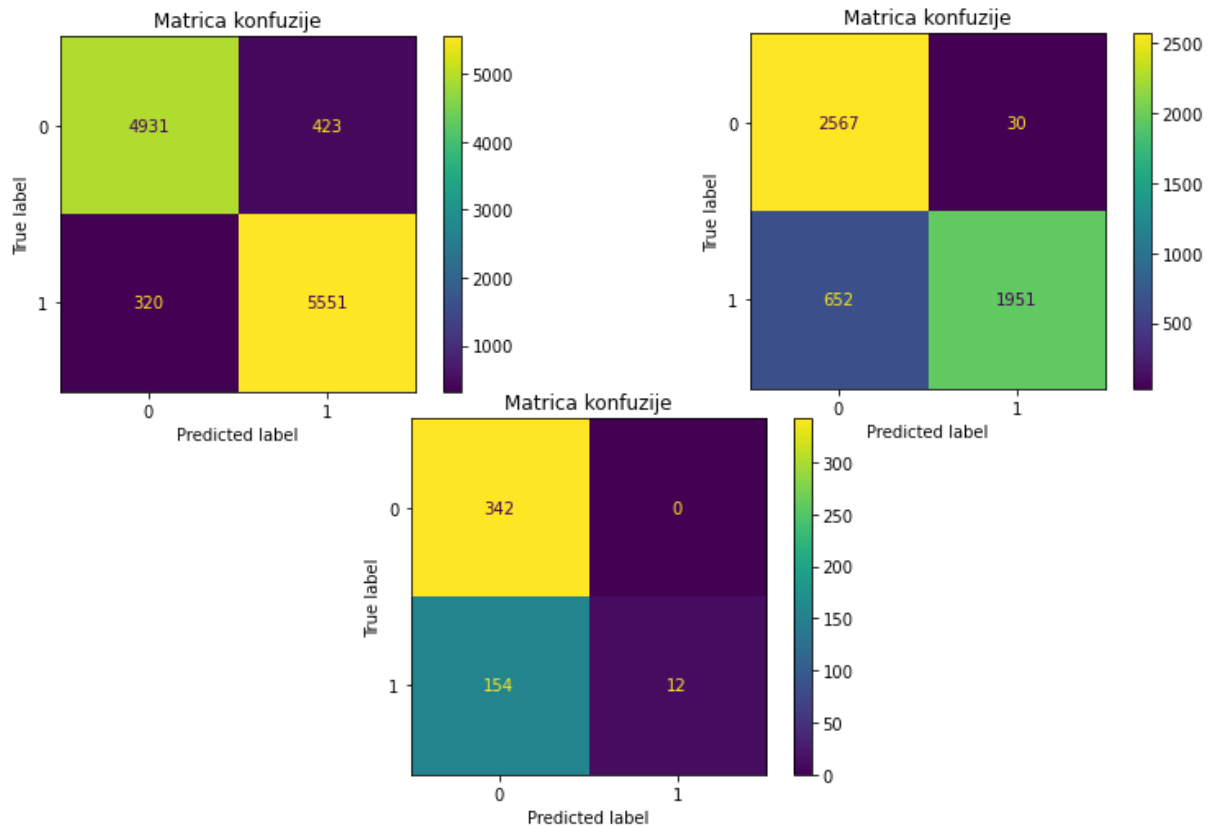
Slika 37: Matrice konfuzije za KNN

Izvor: Vlastiti izvor

Za skup ISOT, broj točnih predikcija za pozitivne vrijednosti iznosi 4972, odnosno točno je predvidio 4972 članaka kao istinite vijesti. 382 članaka je predvidio kao istinite vijesti ali su lažne, 960 članaka je predvidio kao lažne vijesti ali su istinite i 4911 članaka je točno predvidio kao lažne vijesti. Za skup Fake News, broj točnih predikcija za pozitivne vrijednosti iznosi 2278, odnosno točno je predvidio 2278 članaka kao istinite vijesti. 319 članaka je predvidio kao istinite vijesti ali su lažne, 539 članaka je predvidio kao lažne vijesti ali su istinite i 2064 članaka je točno predvidio kao lažne vijesti. Za skup ReCOVerry, broj točnih predikcija za pozitivne vrijednosti iznosi 300, odnosno točno je predvidio 300 članaka kao istinite vijesti. 42 članaka je predvidio kao istinite vijesti ali su lažne, 74 članaka je predvidio kao lažne vijesti ali su istinite i 92 članaka je točno predvidio kao lažne vijesti. Algoritam u većini slučajeva postiže najgore rezultate, u usporedbi sa ostalim algoritmima.

4.7. Matrice konfuzije za Navini Bayes

Slika 38 prikazuje matrice konfuzije za sva 3 skupa podataka. Prva predstavlja skup ISOT, druga skup Fake News a treća skup ReCOVery.

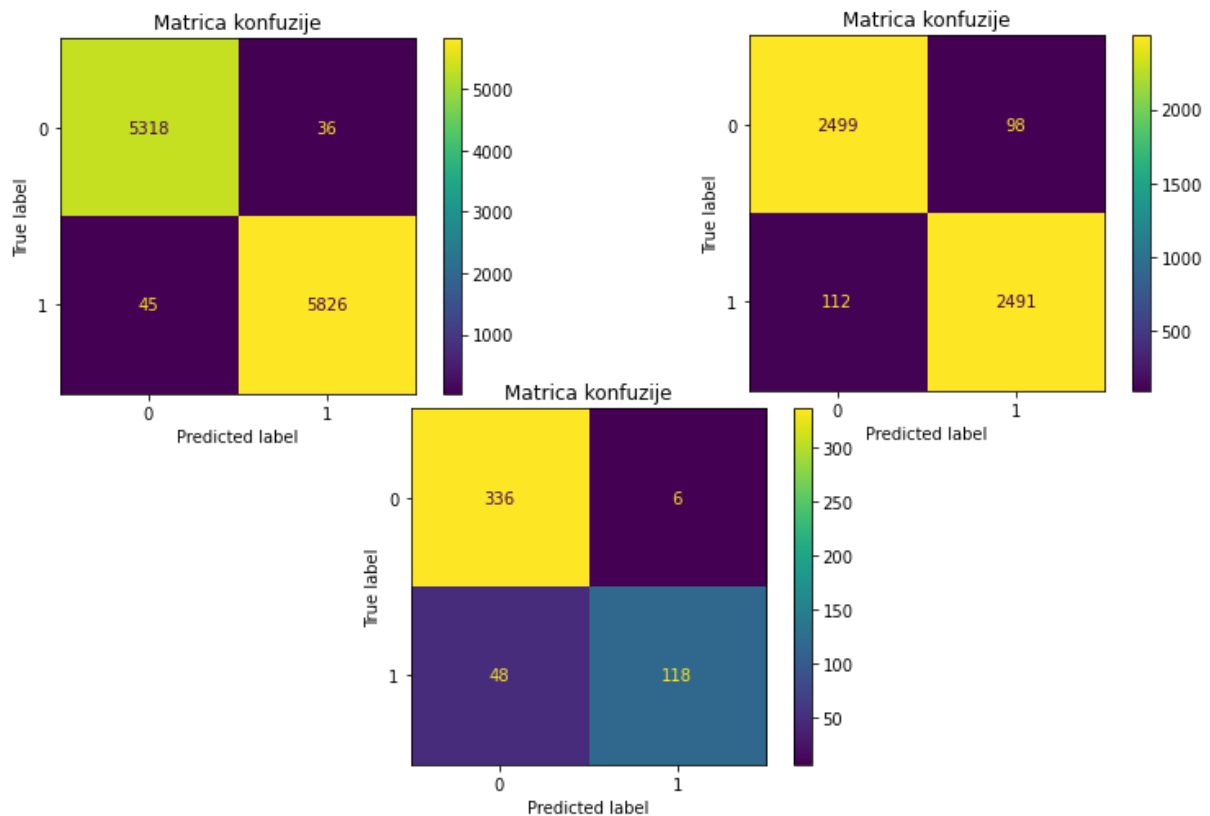


Slika 38: Matrica konfuzije za Navini Bayes
Izvor: Vlastiti izvor

Za skup ISOT, broj točnih predikcija za pozitivne vrijednosti iznosi 4931, odnosno točno je predvidio 4931 članaka kao istinite vijesti. 423 članaka je predvidio kao istinite vijesti ali su lažne, 320 članaka je predvidio kao lažne vijesti ali su istinite i 5551 članaka je točno predvidio kao lažne vijesti. Za skup Fake News, broj točnih predikcija za pozitivne vrijednosti iznosi 2567, odnosno točno je predvidio 2567 članaka kao istinite vijesti. 30 članaka je predvidio kao istinite vijesti ali su lažne, 652 članaka je predvidio kao lažne vijesti ali su istinite i 1951 članaka je točno predvidio kao lažne vijesti. Postiže najgore rezultate za detekciju lažnih vijesti u ovom skupu, sa najvećim brojem FN vrijednosti. Za skup ReCOVery, broj točnih predikcija za pozitivne vrijednosti iznosi 324, odnosno točno je predvidio 342 članaka kao istinite vijesti. 0 članaka je predvidio kao istinite vijesti ali su lažne, 154 članaka je predvidio kao lažne vijesti ali su istinite i 12 članaka je točno predvidio kao lažne vijesti. Kao i u Fake News skupu, postiže najveći broj FN vrijednosti, te je uspješno detektirao samo 12 lažnih vijesti.

4.8. Matrice konfuzije za pasivno agresivni

Slika 39 prikazuje matrice konfuzije za sva 3 skupa podataka. Prva predstavlja skup ISOT, druga skup Fake News a treća skup ReCOVery



Slika 39: Matrice konfuzije za pasivno agresivni algoritam
Izvor: Vlastiti izvor

Za skup ISOT, broj točnih predikcija za pozitivne vrijednosti iznosi 5318, odnosno točno je predvidio 5318 članaka kao istinite vijesti. 36 članaka je predvidio kao istinite vijesti ali su lažne, 45 članaka je predvidio kao lažne vijesti ali su istinite i 5826 članaka je točno predvidio kao lažne vijesti. Za skup Fake News, broj točnih predikcija za pozitivne vrijednosti iznosi 2499, odnosno točno je predvidio 2499 članaka kao istinite vijesti. 98 članaka je predvidio kao istinite vijesti ali su lažne, 112 članaka je predvidio kao lažne vijesti ali su istinite i 2491 članaka je točno predvidio kao lažne vijesti. Za skup ReCOVery, broj točnih predikcija za pozitivne vrijednosti iznosi 336, odnosno točno je predvidio 336 članaka kao istinite vijesti. 6 članaka je predvidio kao istinite vijesti ali su lažne, 48 članaka je predvidio kao lažne vijesti ali su istinite i 118 članaka je točno predvidio kao lažne vijesti. Postiže najbolje rezultate u većini slučajeva za sve skupove podataka.

5. Zaključak

Lažne vijesti su značajan problem koji je postao posebno aktualan u zadnjih nekoliko godina zahvaljujući razvoju tehnologije i popularizaciji društvenih mreža. Takav problem je potrebno riješiti i pristupiti mu na različite načine. Nekada se moglo vjerovati vijestima jer su izvori bili pouzdani i nije bilo potrebe za sumnjom. Danas je situacija drugačija. U ljudima se stvorila potreba za provjeravanjem svega što pročitaju na internetu, kako bi uistinu bili sigurni da to što čitaju, nije laž, izmišljena kako bi napakostila nekome pojedincu ili organizaciji, nego istinita vijest potvrđena činjenicama. Iako je tehnologija olakšala širenje lažnih vijesti, ona također omogućuje da se stane na kraj tom problemu.

Umjetna inteligencija daje mogućnost razvoja modela koji su sposobni detektirati lažne vijesti. U radu su prikazani neki od mogućih modela koji se koriste za rješavanje navedenog problema. Modeli su učeni na 3 različita skupa podataka. Najbolje rezultate su postigli modeli učeni na skupu sa najviše novinskih članaka. Može se zaključiti da je odabir dovoljno velikog skupa podataka, sa relativno dobrom balansiranošću jedan od ključnih koraka u razvoju modela za detekciju lažnih vijesti. Najbolje rezultate su postigli SVM i pasivno agresivni modeli, sa približno savršenom točnošću predikcije. Stoga bi se mogli uspješno implementirati u alate koji bi otkrivali lažne vijesti na razne načine, ukoliko postoji zadovoljavajući skup podataka koji je dobro strukturiran. Visoka točnost modela ne garantira uspješnu detekciju lažnih vijesti u svim situacijama. Model bi trebao biti treniran na podacima koji su napisani drugim jezicima. S obzirom da postoji jako puno različitih jezika, od kojih su neki pisani drugačijim pismom (npr. ćirilicom) i imaju drugačija gramatička pravila, sastavljanje ili pronalazak zadovoljavajućeg skupa podataka bi bio izrazito zahtjevan zadatak. Također postoji mogućnost korištenja strojnog učenja za stvaranje lažnih vijesti. Modeli strojnog učenja detektiraju uzorke u tekstu pomoću kojih razlikuju istinite i lažne vijesti. Na temelju tim detekcija, moguće je stvaranje lažnih vijesti koji bi bile sličnije istinitim i time bi zavarali modele za detekciju lažnih vijesti.

Rad bi se mogao dodatno nadograditi tako da se modeli implementiraju u web stranicu koja bi služila kao provjera pouzdanosti novinskih članaka. Korisnici bi unijeli tekst članka u definirano polje i dobili bi povratu informaciju ukoliko članak potencijalno sadrži lažne vijesti. Još jedna mogućnost je implementiranje modela u alate koji bi se koristili u preglednicima. Kada se otvori web stranica, npr. novinski portal, i odabere se jedan članak, alat bi automatski označio da li je web stranica pouzdana ili sumnjiva.

Kako bi detekcija lažnih vijesti korištenjem umjetne inteligencije imala smisla, potrebna je ljudska suradnja. Algoritmi strojnog učenja, koji se koriste za automatsko otkrivanje dezinformacija, mogu biti uspješni do određene mjere. Međutim, uvijek je potrebna dodatna provjera od strane čovjeka. Dodatno, kritičko razmišljanje i osobno istraživanje su ključni koraci prema sprječavanju širenja lažnih vijesti. Prije dijeljenja vijesti na društvenim mrežama, neophodno je uvjerit se da je vijest istinita, da postoje činjenice koje potvrđuju vijest. Svako dijeljenje vijesti koja nije istinita, dodaje njejoj pouzdanosti u očima drugih.

6. Reference

- [1] J. Penner, »How Tehnology Allows Fake News to Spread,« [Mrežno]. Available: <https://www.activeco.com/how-technology-allows-fake-news-to-spread/>. [Pokušaj pristupa 2 12 2022].
- [2] N. Lavrač, V. Podpečan i M. Robnik-Šikonja, »Representation Learning: Propositionalization and Embeddings,« Springer Nature, 2021.
- [3] M. Shaw, »Truly Project Hate: the third scandal of the official Vote Leave campaign headed by Boris Johnson,« 30 8 2018. [Mrežno]. Available: <https://www.opendemocracy.net/en/dark-money-investigations/truly-project-hate-third-scandal-of-official-vote-leave-campaign-headed-by-/>. [Pokušaj pristupa 2 12 2022].
- [4] J. Soll, »Politico Magazine,« 16 12 2016. [Mrežno]. Available: <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535/>. [Pokušaj pristupa 11 17 2022].
- [5] M. Wills, »Jstor Daily,« 7 11 2017. [Mrežno]. Available: <https://daily.jstor.org/how-the-sun-conned-the-world-with-the-moon-hoax/>. [Pokušaj pristupa 17 11 2022].
- [6] M. Verstraete, J. R. Bambauer i D. E. Bambauer, »IDENTIFYING AND COUNTERING FAKE NEWS,« 24 7 2017. [Mrežno]. Available: <https://news.arizona.edu/story/ua-report-examines-fake-news-and-how-stop-it>. [Pokušaj pristupa 17 11 2022].
- [7] T. P. Institute, »Politifact,« [Mrežno]. Available: <https://www.politifact.com/>. [Pokušaj pristupa 26 11 2022].
- [8] A. P. P. Center, »FactCheck,« [Mrežno]. Available: <https://www.factcheck.org/spindetectors/about/>. [Pokušaj pristupa 26 11 2022].
- [9] S. M. Group, »Snopes,« [Mrežno]. Available: <https://www.snopes.com/about/>. [Pokušaj pristupa 26 11 2022].
- [10] S. v. d. Lans, »18 Organizations Leading the Fight Against Fake News,« [Mrežno]. Available: <https://thetrustedweb.org/organizations-leading-the-fight-against-fake-news/>. [Pokušaj pristupa 26 11 2022].
- [11] Collins, »Collins Dictionary,« 2017. [Mrežno]. Available: <https://blog.collinsdictionary.com/language-lovers/collins-2017-word-of-the-year-shortlist/>. [Pokušaj pristupa 15 11 2022].
- [12] N. Frankland, »The new Daily,« 9 10 2017. [Mrežno]. Available: <https://thenewdaily.com.au/news/world/2017/10/09/donald-trump-fake-news/>. [Pokušaj pristupa 15 11 2022].

- [13] S. Shane, »The New York Times,« 18 1 2017. [Mrežno]. Available: <https://www.nytimes.com/2017/01/18/us/fake-news-hillary-clinton-cameron-harris.html>. [Pokušaj pristupa 15 11 2022].
- [14] R. Farley, »FactCheck.org,« 1 11 2021. [Mrežno]. Available: <https://www.factcheck.org/2021/11/how-many-died-as-a-result-of-capitol-riot/>. [Pokušaj pristupa 16 11 2022].
- [15] B. Dean, »Social Network Usage & Growth Statistics: How Many People Use Social Media in 2022?,« 10 10 2021. [Mrežno]. Available: <https://backlinko.com/social-media-users>. [Pokušaj pristupa 22 11 2022].
- [16] S. Vosoughi, D. Roy i S. Aral, »The Spread of True and False News Online,« Massachusetts Institute of Tehnology, 2018.
- [17] M. D. Molina, N. Hassan, S. S. Sundar, T. Le, M. M. U. Rony i D. Lee, »Does Clickbait Actually Attract More Clicks? Three Clickbait Studies You Must Read,« Yokohama, 2021.
- [18] K. Agarwalla, S. Nandan, V. A. Nair i D. D. Hema, »Fake News Detection using Machine Learning and Natural Language Processing,« International Journal of Recent Technology and Engineering (IJRTE), 2019.
- [19] Z. Khanam, B. N. Alwasel, H. Sirafi i M. Rashid, »Fake News Detection Using Machine Learning Approaches,« IOP Publishing, 2021.
- [20] A. C. Tiziana Assenza, »The Ability to Distill the Truth,« 2022.
- [21] I. Ahmad, M. Yousaf, S. Yousaf i M. Ovais, »Fake News Detection Using Machine Learning Ensemble Methods,« Hindawi, 2020.
- [22] H. Ahmed, I. Traore i S. Saad, »Detecting opinion spams and fake news using text,« Wiley, 2018.
- [23] U. M. L. Club, »Kaggle,« 2018. [Mrežno]. Available: <https://www.kaggle.com/competitions/fake-news/data>. [Pokušaj pristupa 23 11 2022].
- [24] Jruvika, »Kaggle,« 2018. [Mrežno]. Available: <https://www.kaggle.com/datasets/jruvika/fake-news-detection?resource=download>. [Pokušaj pristupa 23 11 2022].
- [25] W. Y. Wang, »Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection,« Vancouver, 2017.
- [26] A. Vlachos i S. Riedel, »Fact Checking: Task definition and dataset construction,« Association for Computational Linguistics, Baltimore, 2014.
- [27] C. Silverman, »Lies, Damn Lies and Viral Content.,« Columbia University, 2015.
- [28] W. Ferreira i A. Vlachos, »Emergent: a novel data-set for stance classification,« 2016.

- [29] X. Zhou, E. Ferrara, A. Mulay i R. Zafarani, »ReCOVerry: A Multimodal Repository for COVID-19 News Credibility Research,« Association for Computing Machinery, New York, 2020.
- [30] I. C. Education, »Natural Language Processing (NLP),« 2 7 2020. [Mrežno]. Available: <https://www.ibm.com/cloud/learn/natural-language-processing>. [Pokušaj pristupa 18 11 2022].
- [31] A. M. Turing, »COMPUTING MACHINERY AND INTELLIGENCE,« Mind, 1950.
- [32] Turing, »Turing for Developers,« [Mrežno]. Available: <https://www.turing.com/kb/natural-language-processing-function-in-ai>. [Pokušaj pristupa 20 11 2022].
- [33] I. Boban, »Unaprijeđene metode pronalaženja rečenica temeljene na djelomičnom preklapanju riječi i postupcima pretprocesiranja teksta,« Sveučilište u Mostaru, Mostar, 2022.
- [34] G. Sharma, »Analytics Vidhya,« 26 5 2021. [Mrežno]. Available: <https://www.analyticsvidhya.com/blog/2021/05/5-regression-algorithms-you-should-know-introductory-guide/>. [Pokušaj pristupa 19 11 2022].
- [35] M. Waseem, »Edureka,« 28 3 2022. [Mrežno]. Available: <https://www.edureka.co/blog/classification-in-machine-learning/>. [Pokušaj pristupa 19 11 2022].
- [36] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow, Sebastopol: O'Reilly Media Inc., 2017.
- [37] L. Gašparović, »Redukcija dimenzija u dubinskoj analizi,« Rijeka, 2020.
- [38] S. D. Piyush Verma, »Synopsys,« 27 4 2021. [Mrežno]. Available: <https://www.synopsys.com/ai/what-is-reinforcement-learning.html>. [Pokušaj pristupa 19 11 2022].
- [39] V. Kanade, »Spiceworks,« 18 4 2022. [Mrežno]. Available: <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/>. [Pokušaj pristupa 20 11 2022].
- [40] B. Žitković, »Primjena stabla odlučivanja na skupu podataka iz obrazovanja,« Fakultet organizacije i informatike, Varaždin, 2020.
- [41] I. Logunova, »Serokell,« 23 6 2022. [Mrežno]. Available: <https://serokell.io/blog/random-forest-classification>. [Pokušaj pristupa 21 11 2022].
- [42] A. Sasidharan, »Geeksforgeeks,« [Mrežno]. Available: <https://www.geeksforgeeks.org/support-vector-machine-algorithm/?tab=article>. [Pokušaj pristupa 21 11 2022].
- [43] A. Navlani, »Datacamp,« 12 2019. [Mrežno]. Available: <https://www.datacamp.com/tutorial/svm-classification-scikit-learn-python>. [Pokušaj pristupa 21 11 2022].

- [44] I. Mierswa, »Rapidminer,« 8 5 2017. [Mrežno]. Available: <https://rapidminer.com/blog/k-nearest-neighbors-laziest-machine-learning-technique/>. [Pokušaj pristupa 21 11 2022].
- [45] A. Ali, »Medium,« 21 7 2018. [Mrežno]. Available: <https://medium.com/machine-learning-researcher/k-nearest-neighbors-in-machine-learning-e794014abd2a>. [Pokušaj pristupa 21 11 2022].
- [46] J. Joyce, »Stanford Encyclopedia of Philosophy Archive,« 28 6 2003. [Mrežno]. Available: <https://plato.stanford.edu/archives/spr2019/entries/bayes-theorem/>. [Pokušaj pristupa 21 11 2022].
- [47] A. Stuart i K. Ord, Kendall's Advanced Theory of Statistics, Volume 1, London: Edward Arnold, 1994.
- [48] A. Kharwal, »Passive Aggressive Classifier in Machine Learning,« 10 2 2021. [Mrežno]. Available: <https://thecleverprogrammer.com/2021/02/10/passive-aggressive-classifier-in-machine-learning/>. [Pokušaj pristupa 26 11 2022].
- [49] N. Rattner, »CNBC,« 13 1 2021. [Mrežno]. Available: <https://www.cnbc.com/2021/01/13/trump-tweets-legacy-of-lies-misinformation-distrust.html>. [Pokušaj pristupa 15 11 2022].
- [50] S. Arora, »Aitude,« 29 1 2020. [Mrežno]. Available: <https://www.aitude.com/supervised-vs-unsupervised-vs-reinforcement/>. [Pokušaj pristupa 19 11 2020].
- [51] Canley, »Wikipedia,« [Mrežno]. Available: https://en.wikipedia.org/wiki/Logistic_regression#/media/File:Exam_pass_logistic_curve.svg. [Pokušaj pristupa 20 11 2022].
- [52] E. Hatipoglu, »Medium,« 13 7 2018. [Mrežno]. Available: <https://medium.com/@ekrem.hatipoglu/machine-learning-classification-decision-tree-random-forest-part-12-8c9515d811b9>. [Pokušaj pristupa 21 11 2022].
- [53] L. Sučić, »9 Korona Kid pametnih trikova: Kako prepoznati lažne (i čudno bizarne) vijesti?,« 6 17 2020. [Mrežno]. Available: <https://brickzine.hr/9-korona-kid-pametnih-trikova-kako-prepoznati-lazne-i-cudno-bizarne-vijesti/>. [Pokušaj pristupa 22 11 2022].
- [54] Kazoo04, »Hatena blog,« 20 02 2012. [Mrežno]. Available: <https://kazoo04.hatenablog.com/entry/2012/12/20/000000>. [Pokušaj pristupa 27 11 2022].
- [55] A. Suresh, »What is a confusion matrix?,« Analytics Vidhya, 2022.

7. Popis slika

Slika 1: Podjela dezinformacija.....	3
Slika 2: Članak Cammeron Harrisa.....	5
Slika 3: Objave Donalda Trumpa na Twitteru	6
Slika 4: Clickbait.....	7
Slika 5: Primjer izjave (LIAR skup podataka)	10
Slika 6: Primjer izjave iz Emergent skupa podataka.....	11
Slika 7: Graf logističke regresije.....	16
Slika 8: Primjer stabla odluke	17
Slika 9: Primjer slučajne šume	19
Slika 10: SVM primjer	19
Slika 11: Primjena KNN algoritma	21
Slika 12: Izvođenje pasivno agresivnog algoritma	22
Slika 13: Primjer matrice konfuzije	23
Slika 14: Balansiranost ISOT skupa podataka	26
Slika 15: ISOT skup podataka.....	26
Slika 16: Usporedba kategorija članaka	27
Slika 17: Kaggle Fake News skup podataka	28
Slika 18: Funkcije za pripremu teksta	29
Slika 19: Funkcija za uklanjanje zaustavnih riječi	29
Slika 20: Funkcije za normalizaciju teksta.....	30
Slika 21: Usporedba istinitih i lažnih vijesti (ReCOVerry).....	31
Slika 22: TF-IDF vektorizator.....	31
Slika 23: Model logističke regresije.....	32
Slika 24: Model stabla odluke.....	32
Slika 25: Model slučajne šume.....	32
Slika 26: SVM model.....	32
Slika 27: KNN model.....	33
Slika 28: Naivni Bayes model.....	33
Slika 29: Pasivno agresivni model	33
Slika 30: Usporedba mjera (ISOT).....	34
Slika 31: Usporedba mjera (Fake News).....	35
Slika 32: Usporedba mjera (ReCOVerry)	36
Slika 33: Matrice konfuzije za logističku regresiju.....	38
Slika 34: Matrica konfuzije za stablo odluke	39
Slika 35: Matrica konfuzije za slučajne šume	40
Slika 36: Matrica konfuzije za SVM.....	41
Slika 37: Matrice konfuzije za KNN.....	42
Slika 38: Matrica konfuzije za Navini Bayes.....	43
Slika 39: Matrice konfuzije za pasivno agresivni algoritam	44

8. Popis tablica

Tablica 1: Usporedba podgrupa strojnog učenja.....	15
Tablica 2: Točnost algoritama za prvi skup podataka.....	34
Tablica 3: Točnost algoritama za drugi skup podataka.....	35
Tablica 4: Točnost algoritama za drugi skup podataka.....	36