

Analiza sentimenta u hrvatskim tvitovima na temu potresa

Blazsetin, Danijel Sztjepan

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:131:813940>

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

Download date / Datum preuzimanja: **2024-11-16**



Sveučilište u Zagrebu
Filozofski fakultet
University of Zagreb
Faculty of Humanities
and Social Sciences

Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences](#)



SVEUČILIŠTE U ZAGREBU

FILOZOFSKI FAKULTET

ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI

Ak. god. 2021./2022.

Danijel Blazsetin

Analiza sentimenta u hrvatskim tvitovima na temu potresa

Diplomski rad

Mentor: dr. sc. Petra Bago, doc.

Zagreb, lipanj 2022.

Izjava o akademskoj čestitosti

Izjavljujem i svojim potpisom potvrđujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

(potpis)

Sažetak

1. Uvod.....	5
2. Analiza sentimenta.....	6
2.1. Računalni izazovi u analizi sentimenta.....	6
3. Strojno učenje.....	9
3.1. Odabrani pristupi za razvoj alata za analizu sentimenta.....	10
3.1.1. Metoda potpornih vektora.....	10
3.1.2. Naivni Bayesov klasifikator.....	11
3.1.3. Metoda povratnih neuronskih mreža.....	12
3.2. Mogući pristupi izradi alata temeljenih na povratnim neuronskim mrežama.....	13
4. Razvoj Twitter korpusa.....	15
4.1. Analiza sentimenta na Twitteru nakon katastrofa.....	17
5. Izrada anotiranoga Twitter korpusa.....	20
5.1. Metodologija.....	20
5.2. Rezultati.....	24
5.3. Interpretacija rezultata.....	27
5.3.1. Kontekstualizacija Cohenove kappe.....	27
5.3.2. Dvosmisleni tvitovi.....	28
5.3.3. Omjer kategorija.....	30
5.3.4. Razlike u neodlučnosti anotatora.....	31
6. Izrada alata za određivanje sentimenta.....	33
6.1. Metodologija.....	33
6.2. Rezultati.....	39
6.3. Interpretacija rezultata.....	42
7. Zaključak.....	46

Literatura	48
Popis tablica.....	53
Popis slika.....	53
Popis formula.....	53

1. Uvod

Prepoznavanje je emocija u jeziku oduvijek predstavljalo nesigurnost istraživačima jezika bili oni lingvisti, psiholozi, sociolozi ili istraživači iz područja obrade prirodnoga jezika. Emocije i sentiment uvijek su prisutni u ljudskoj jezičnoj proizvodnji, stvarajući tako iskaze koji s jedne strane nose informaciju, a s druge strane prenose govornikov stav prema izgovorenome ili napisanome. Prepoznavanje tih često prekrivenih i zamršenih emocija nužno je za razumijevanje svake složenije komunikacije, uključujući, ali ne i isključivo, humor, ironiju, sarkazam, uvrede, jezične inovacije i dr. Popularizacijom računalnih metoda lingvisti i drugi istraživači jezika sve se više koriste računalnim metodama za obradu jezika. Međutim, jezikoslovci koji gradi pristupaju tradicionalnim i konzervativnijim metodama održali su distancu prema ovakvim metodama proučavanja jezika, ističući kako računalni algoritmi ne mogu razaznati i protumačiti složene ljudske emocije i namjere koje se kriju iza iskaza, implicirajući time da ljudski um ostaje nezamjenjiv u analizi i interpretaciji sentimenta. Unatoč ovakvim primjedbama, obrada se prirodnoga jezika, predvođena često istraživačima interdisciplinarne znanstvene pozadine, od provjere pravopisa i izrade korpusa pomaknula prema složenijim i zahtjevnijim analizama jezika, uključujući analizu sintaktičkih elemenata, prevođenje i prepoznavanje emocije ili sentimenta u nekome jezičnom iskazu. Razvojem pristupačnosti računalnih metoda područje je analize sentimenta postajalo sve popularnije, pa se danas analizom sentimenta osim istraživača iz akademskih krugova bavi i šira publika.

U sklopu ovoga rada razvit će se alat za prepoznavanje sentimenta u *tvitovima* na temu potresa. Prvi, teorijski, dio govori o položaju analize sentimenta u obradi prirodnoga jezika i računalne lingvistike. Potom se opisuju najčešći pristupi analizi sentimenta i najpopularnije metode za izradu i obradu jezičnih resursa. Nakon toga slijedi kratki pregled strojnoga učenja, kao jednoga od temeljnih pristupa modernoj i suvremenoj analizi sentimenta. Drugi dio rada govori o izradi jezičnoga resursa, točnije korpusa *tvitova* na temu potresu. U ovome se dijelu opisuje postupak prikupljanja podataka, njihova anotacija i obrada te analiza korpusa. Naposljetku, u trećem dijelu rada opisuje se razvoj alata na temelju izrađenoga jezičnoga korpusa te se dobiveni rezultati evaluiraju i kontekstualiziraju s pomoću sličnih istraživanja. Na samome kraju prikazuju se zaključci rada i govori se o mogućim nadogradnjama sustava.

2. Analiza sentimenta

Obrada prirodnoga jezika (engl. *natural language processing*) bavi se proučavanjem, analizom i obradom prirodnoga jezika u računalnome okruženju. Analiza sentimenta (engl. *sentiment analysis*)¹ čini posebnu granu obrade prirodnog jezika koja je uslijed razvoja digitalne umreženosti svijeta i povećane aktivnosti na društvenim mrežama zaslužila zainteresiranost i uzbuđenje istraživača. Budući da se bavi jezikom, ova je grana računarstva naišla na niz prepreka čije rješavanje zahtijeva interdisciplinarni pristup (usp. Cambria 2014).

Analiza sentimenta bavi se proučavanjem mišljenja, stavova i emocija ljudi prema pojedincima, problemima, događajima i pojavama (Liu 2011:474). Pri analizi sentimenta istraživači nailaze na već uvriježene probleme pri obradi prirodnog jezika koje uključuju nepravilnosti u pravopisu te korištenje dijalekta i drugih jezičnih inačica. Dodatno, oni u fokus svoga istraživanja u ovome slučaju stavljaju sentiment nekoga teksta koji je najčešće teško sa sigurnošću odrediti, jer je on zapravo tek interpretacija raznih osjećaja koje ljudi prepoznaju u nekom jezičnome iskazu.

Analizom se sentimenta ne bavi samo znanstvena zajednica, već se metodama koje se u ovome području koriste služe privatne tvrtke i državne institucije koje, analizirajući sentiment novinskih članaka, recenzija, komentara i izjava na društvenim mrežama, donose zaključke o općem mnijenju i razmišljanju u društvu. Analiza sentimenta o nekom proizvodu, društvenoj pojavi ili političkoj opciji mogu zamijeniti ili nadopuniti ankete čija izrada zahtijeva neusporedivo više ljudskih i financijskih resursa.

2.1. Računalni izazovi u analizi sentimenta

Analiza je sentimenta prosudba o osjećaju koji dominira u nekome jezičnom iskazu. Automatsku analizu sentimenta vrše računalne aplikacije koje ne zahtijevaju korisničku prosudbu kako bi odredile sentiment nekoga iskaza. Označavanje s pomoću ovakvih aplikacija ima nekoliko prednosti u odnosu na ručno određivanje sentimenta. Aplikacije za analizu sentimenta u vrlo kratkom vremenu mogu obraditi veliku količinu jezičnog materijala, dok bi obrada iste količine podataka ljudskim određivanjem zahtijevalo puno više vremena. Alati su za analizu sentimenta stoga puno isplativiji i dostupniji (Liu 2011: 459–460).

¹ Liu (2012:7) kao sinonim koristi i termin rudarenje mišljenja (engl. *opinion mining*). Radi konzistentnosti, u ovome ćemo se radu služiti terminom analiza sentimenta.

Dakako, aplikacije se za analizu sentimenta razvijaju na temelju korpusa koje anotiraju ljudi. Točnost aplikacije u velikoj mjeri ovisi o kvaliteti jezičnih materijala na kojima je sustav učen. Prva je faza izrade aplikacije zbog toga često anotacija. Pri označavanju sentimenta jezičnoga materijala istraživač se oslanja na intuiciju i subjektivnost anotatora. Anotatori su međutim stavljeni pred težak zadatak jer sentiment u jezičnim iskazima često nije jednoznačan². Teško je stoga govoriti o korpusima u kojima je sentiment *loše* označen, jer je riječ o subjektivnim doživljajima anotatora, te bit ovakvih korpusa upravo i leži u toj subjektivnosti. Nakon početne faze izrade, anotatori se *isključuju* iz procesa označavanja, a njihova se subjektivnost aplikacijama predstavlja kao objektivni podatak na kojemu se aplikacija *uči* i trenira.

Mišljenja i emocije nikada nisu jednoznačni. Oni se uvijek javljaju u jezičnome, kulturološkome, ideološkome ili drugim kontekstima. Sentiment nekoga jezičnog iskaza u velikoj mjeri ovisi o pravilima platforme na kojoj se javlja ili forme u kojoj se javlja. Isti jezični iskaz u različitim društvima može izazvati druge reakcije. Ono što će određenoj grupi biti ironično, sarkastično i humoristično, drugoj može biti uvredljivo i agresivno. Događa se da se u jednome jezičnom iskazu istovremeno izraze pozitivne i negativne emocije. Kako će se odrediti sentiment ako je *osjećajnost* iskaza kontradiktoran? Hoće li se istraživač koncentrirati na riječi, sintagme, rečenice, odlomke, na čitav tekst ili na cijeli kulturološki kontekst kada analizira sentiment? To su pitanja bez točnoga odgovora pa se u praksi često koristi metoda pokušaja i pogreške (engl. *hit and miss*) (Liu 2011: 476). Iako se kategorije sentimenta u radovima često predstavljaju kao jasno definirane kategorije, naravno, riječ je o spektru emocija od krajnje pozitivnoga do krajnje negativnoga. Najčešće se koristi trojna podjela na pozitivno, neutralno i negativno, ali se također može koristiti binarna podjela na pozitivno i negativno ili višeklasna podjela s elementima potpuno pozitivno, više pozitivno nego neutralno, neutralno, više negativno nego neutralno, potpuno negativno³. Odluka je arbitrarna (iako se binarna većinom ipak odbacuje) i u velikoj mjeri ovisi o svrsi analize, količine jezičnog resursa, broju anotatora pri izradi jezičnih materijala za alate i slično. Uz svoje neupitne prednosti, automatska analiza sentimenta pred istraživače stavlja niz novih izazova. Alati za analizu sentimenta, općenito govoreći, nisu univerzalni. Najtočniji su na tipu jezičnoga odsječka

² Postoji rasprava o tome koji bi kriteriji *trebali* utjecati na prosudbu sentimenta neke jezične jedinice. Liu (2012: 16-24) predlaže četveročlani skup (engl. *quadruple*) koji može opisati sentiment jednoga jezičnog izraza, a on se sastoji od (i) predmeta sentimenta, (ii) sentimenta o predmetu, (iii) mišljenika, (iv) vremena izricanja mišljenja. Za više o *opisima sentimenta* vidi Kim i Hovy 2004.

³ Ovakva se podjela koristi u Elbagir i Yang 2019.

na kojima se alat i trenirao (usp. Labille, Gauch i Alfarhood 2017. i Liu 2020:81–84). Za izradu kvalitetnoga alata za određivanje sentimenta istraživači se moraju osloniti na dobro sastavljene korpuse koji su istovremeno opsežni, ali i kvalitetno označeni.

2.1.1. Problemi s jezicima bez razvijenih jezičnih resursa

Za izradu i razvoj jezičnih alata, uključujući alata za analizu sentimenta, nužno je imati kvalitetne jezične resurse. Među jezične resurse ubrajamo korpuse, leksikone, gramatike, ali u širem smislu i alate koji mogu biti od velike pomoći pri izradi drugih jezičnih alata (Tadić, Brozović-Rončević i Kapetanović 2012: 25–37). Istraživači koji žele razviti jezične alate za svoje jezike nažalost nailaze na prepreke jer izrada većine naprednih alata zahtijeva jezične resurse koji za manje jezike često nisu dostupni. Istraživači se često okreću proučavanju engleskoga jezika kako bi mogli predstaviti nove pristupe i alate, a istovremeno svoje nacionalne jezike zapostavljaju i time se rascjep između engleskoga jezika i manjih jezika povećava. Prema istraživanju iz 2012. hrvatski je jezik vrlo daleko od optimalne razvijenosti jezičnih tehnologija (alata i resursa) (Tadić, Brozović-Rončević i Kapetanović 2012: 38–41). Iako je riječ o sada već deset godina starome istraživanju, istraživanje Rehm et al. (2020) potvrđuje da se stanje nije promijenilo te se hrvatski jezik u računalnome okruženju, uz sve druge jezike govorene na području Europske unije, nije razvijao željenom brzinom te je podrška za hrvatski jezik sporadična.

Što se tiče jezičnih resursa koji su relevantni pri izradi sustava za analizu sentimenta za hrvatski jezik, možemo govoriti o dvama resursima: Korpus sentimentom naznačenih hrvatskih vijesti (engl. *Sentiment Annotated Dataset of Croatian News*) (Pelicon et al. 2020) i Twitter Sentiment za 15 europskih jezika (engl. *Twitter sentiment for 15 European languages*) (Mozetić, Grčar i Smailović 2016), oba objavljena na platformi CLARIN.SI.

3. Strojno učenje

Proteklih godina računalna tehnologija, pa tako i obrada prirodnoga jezika okrenula se prema novim metodama izrade raznih alata i aplikacija, strojnome učenju (engl. *machine learning*). Strojno učenje postavlja dva pitanja: Kako se može napraviti sustav koji se automatski poboljšava na temelju povratnih informacija i koja su ta temeljna načela koja upravljaju misaonim procesima učenja unutar svih sustava, uključujući računala i ljude. Tražeći odgovore na ta pitanja pojavilo se niz načina za izradu sustava koji će moći učiti na podacima (Jordan i Mitchell 2015: 255–256). Jednostavnije, „strojno učenje je znanost (i umjetnost) programiranja računala da bi oni mogli učiti *na temelju podataka*“⁴ (Géron 2019: 20). Većinom se pri izradi ovakvih sustava koristi nadzirano strojno učenje (engl. *supervised machine learning*). Posebna podvrsta nadziranog strojnog učenja je duboko učenje (engl. *deep learning*). Neuronske su mreže (engl. *neural network*) tip sustava izgrađen s pomoću dubokoga učenja. One se sastoje od mnoštva čvorova (engl. *neuron*) i vektora koji ih povezuju. Vektore definira njihova težina koja se tijekom razvoja i učenja sustava mijenja te oni upravljaju radom aplikacije. Sustav svoje izračune, odluke *vraća* u sustav i uzima ih u obzir pri budućim radnjama (engl. *backpropagation*) (Lecun, Bengio i Hinton 2015: 436–438). Drugu veliku skupinu sustava izgrađenih dubokim učenjem čine one razvijene metodom nenadziranog strojnog učenja (engl. *unsupervised machine learning*). Danas je izgradnja ovakvih sustava i dalje u eksperimentalnoj fazi, ali dobar dio znanstvenika i stručnjaka budućnost dubokog učenja upravo vidi u nenadziranom strojnom učenju. Ovaj tip strojnoga učenja ne zahtijeva ljudski nadzor. Sustavi se hrane neoznačenim podacima koje onda sustav, pronalazeći zajedničke osobine, sortira. Na ovaj se način otkrivaju skupine ili klasteri (engl. *cluster*) koje su ljudima možda neintuitivne, ali, dakako, postoje i mogu pružiti jedinstveni uvid u podatke. Nad ovakvim sustavima istraživač nema kontrolu. Ne postoji mogućnost da se takvi sustavi podešavaju (engl. *tweak*) i zbog toga znaju biti vrlo skupi (vremenski, energetska, hardverski) za izradu. U slučaju da sustav ne daje zadovoljavajuće rezultate, on se iznova gradi. Ovakvi sustavi zahtijevaju veliku količinu podataka koja često nije dostupna istraživačima (Lecun, Bengio i Hinton 2015: 441–442; Jordan i Mitchell 2015: 257–258).

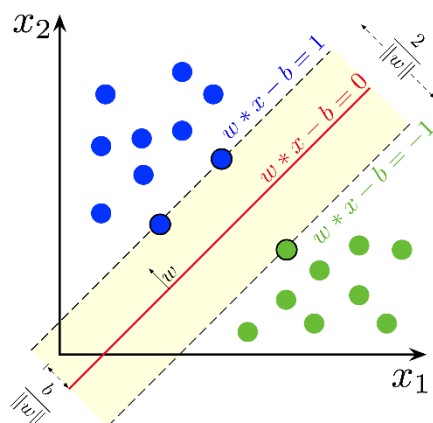
⁴ Definicije i termine koji su u izvornome obliku pisane na engleskome jeziku donosimo u slobodnome prijevodu autora.

Većina alata za analizu sentimenta izgrađena je metodom nadziranog strojnog učenja. To su alati koji klasificiraju. U slučaju nadziranog strojnog učenja sustav se uči na anotiranim podacima te sustav klasificira podatke u već definirane kategorije. Sustav pokušava naučiti (i pronaći) najbolji algoritam za kategorizaciju. Tri najpopularnije metode za izradu alata za analizu sentimenta su metoda potpornih vektora (engl. *support vector machines*, *SVM*), naivni Bayesov klasifikator i povratna neuronska mreža (engl. *recurrent neural network*, *RNN*).

3.1. Odabrani pristupi za razvoj alata za analizu sentimenta

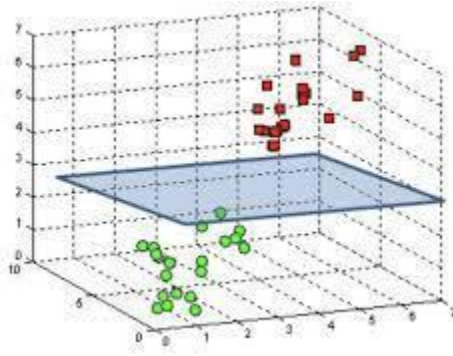
3.1.1. Metoda potpornih vektora

Metoda potpornih vektora ili metoda maksimalno granične hiperravnine nastoji pronaći hiperravninu koja razdvaja hiperprostor primjera u dva poluprostora koji odgovaraju dvjema klasama podataka tako da je udaljenost te hiperravnine od najbližih točaka podataka maksimalna. Problem klasifikacije može se definirati u konačnom dimenzionalnom prostoru, ali moguće je da klase nisu razdvojive u tome prostoru. Tada se dodaje varijabla koja problem preslikava u dodatnu dimenziju (Cetinić 2013: 12–16). Jednostavnije, ova metoda nastoji pronaći *najtočniju* granicu između dviju kategorija čija će udaljenost biti jednaka prema objema kategorijama. Na temelju te će se granice ulazne vrijednosti razvrstati u kategorije.



Slika 1: Dvodimenzionalan potporni vektor (izvor: https://en.wikipedia.org/wiki/File:SVM_margin.png)

Može se dogoditi da su klase nerazdvojive u dvodimenzionalnome prostoru (koristeći samo osi x i y) pa se tada uvode dodatne osi kako bi se mogla definirati granica. Rezultati s metodom potpornih vektora izrazito su dobri, a implementacija nije prezahtjevna, stoga je ova metoda jedna od najpopularnijih pri izradi alata za analizu sentimenta (Šnajder 2020: 1–4).



Slika 2: Trodimenzionalni potporni vektor (izvor: <https://www.xilinx.com/developer/articles/exploring-support-vector-machine-acceleration-with-vitis.html>)

3.1.2. Naivni Bayesov klasifikator

Naivni Bayesov klasifikator koristi se u slučajevima kada podaci čine konačni skup vrijednosti, a njima se također dodjeljuje vrijednost iz konačnog skupa vrijednosti. Klasifikator na temelju jednostavnih statističkih podataka i podataka o vjerojatnosti iz skupa za učenje novim primjerima računa vjerojatnost pripadanja nekoj od kategorija. Odnosno, „on omogućuje izračun vjerojatnosti hipoteze (*koja je vjerojatnost da je zeleno voće jabuka*) iz njene *a priori* vjerojatnosti (*koja je vjerojatnost da je voće jabuka*), vjerojatnosti pojave podataka ako je hipoteza točna (*koja je vjerojatnost da je jabuka zelena*) i predočenih podataka (*koja je vjerojatnost da je voće zeleno*)“⁵ (Ahel 2003: 12). Naivni Bayesov klasifikator smatra se naivnim, jer attribute skupova tretira izolirano i ne proučava korelaciju, što ga istovremeno čini jednostavnim.

$$P(c|x) = \frac{P(c)P(x|c)}{P(x)}$$

Formula 1: Naivni Bayesov klasifikator

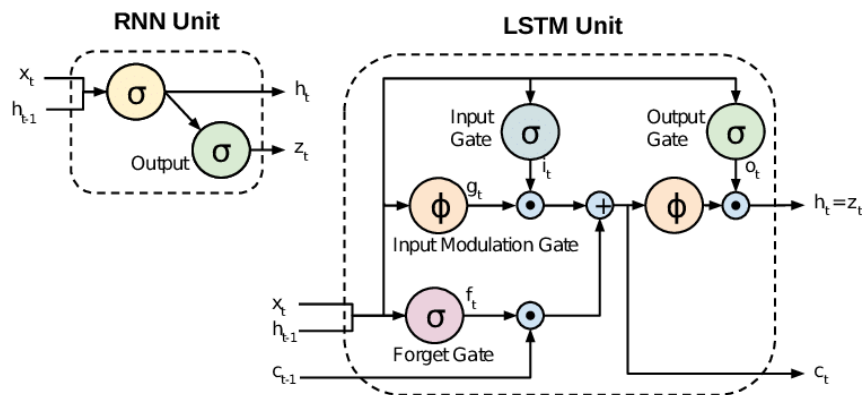
Iznenadujuće, naivni Bayesov klasifikator daje dobre rezultate pri analizi sentimenta i spada među bolje klasifikatore s obzirom na njegovu pristupačnost (Rish 2001: 41). Također, Naivni Bayesov klasifikator predvidljiv je i stabilan (rezultati, suprotno od nenadziranih neuronskih mreža, ne

⁵ Kurzivirani tekst dodao D.B.

mogu u velikoj mjeri odudarati od očekivanog). Većina se (domaćih) radova koji analiziraju sentiment oslanjaju na naivni Bayesov klasifikator.⁶

3.1.3. Metoda povratnih neuronskih mreža

Povratne neuronske mreže (engl. *recurrent neural network*, *RNN*) spadaju u kategoriju sustava izrađenih nadziranom strojnim učenjem i kao takvi, slično prijašnjim modelima, treniraju se označenim korpusom. Povratne neuronske mreže razlikuju se od sustava potpornih vektora, naivnog Bayesovog klasifikatora i unaprijednih neuronskih mreža (engl. *feed-forward neural network*) u tome što imaju mogućnost *pamćenja* pri klasifikaciji. Povratne neuronske mreže na ovaj način dijelom su i kružne, a upravo zato se i zovu povratne. One istovremeno primaju *sadašnju* i *prošlu* informaciju i na temelju nje klasificiraju podatke. Povratne bi se neuronske mreže mogle promatrati i kao niz unaprijednih neuronskih mreža. Povratne neuronske mreže obično se nadopunjuju mrežama duge kratkoročne memorije (engl. *long short-term memory*, *LSTM*) koja omogućuje povratnim neuronskim mrežama da zapamte mnoštvo klasifikacijskih odluka (Škrlec 2019: 18–21). Povratne neuronske mreže zapravo ovako primaju mnoštvo ulaznih podataka, među kojima uspostavljaju odnos na temelju kojih donose odluku.



Slika 3: Povratna neuronska mreža i mreže duge kratkoročne memorije (izvor: <https://www.researchgate.net/profile/Aliaa-Rassem/publication/317954962/figure/fig2/AS:667792667860996@1536225587611/RN-N-simple-cell-versus-LSTM-cell-4.png>)

Kao što je već spomenuto, naivni Bayesov klasifikator svaki atribut promatra izolirano, bez da uzme u obzir kontekst onoga predmeta koji treba klasificirati. Povratne neuronske mreže su daleko bolje rezultate kod podataka koji su po svojoj prirodi sekvencijalni (ili slijedni) (LeCun, Bengio i

⁶ Primjeri ovakvih analiza se mogu naći u Krajcar (2014), Šikač (2020) i Ivaninić (2021).

Hinton 2015: 441). Povratne su neuronske mreže i dalje stabilni i nadzirani sustavi, ali u usporedbi su se s drugim klasifikatorima dokazale naprednijima⁷.

3.2. Mogući pristupi izradi alata temeljenih na povratnim neuronskim mrežama

U prijašnjem smo poglavlju predstavili neke od pristupa izradi alata za analizu sentimenta. Spomenuti se pristupi mogu primijeniti na različite načine. Kao i u slučaju drugih računalnih programa, naravno i alati se za analizu sentimenta mogu napisati pomoću različitih programskih jezika, među kojima su najpopularniji Python⁸, C++⁹, C¹⁰, CUDA¹¹. Puno važnije od samoga jezika u kojemu je alat izrađen je dostupnost različitih knjižnica (engl. *library*) koje istraživačima omogućuju lakšu i bržu izradu alata za analizu sentimenta (ili alata drugih tipova koji se temelje na dubokom učenju). Nekolicina je knjižnica koje se danas koriste za izradu povratnih neuronskih mreža. Među najpopularnijima se nalaze Python i C++ knjižnice TensorFlow¹², PyTorch¹³ i Theano¹⁴.

Kada govorimo o alatima za analizu sentimenta, uvijek moramo imati na umu jezični resurs pomoću kojega će se on učiti. Temeljni korpus pri izradi alata može biti opći ili specijalizirani, tj. specifičan za onaj tip teksta na kojemu će se kasnije izrađeni alat primijeniti. Najkvalitetniji se alati dobiju kada se sustavi treniraju na specijaliziranim korpusima, međutim, to često zbog manjka označenih specijaliziranih korpusa nije moguće (Han et. al 2018: 1–3). Za izradu se nekoga alata uz specijalizirani korpus može koristiti i opći korpus za učenje. Također se pri izradi alata za analizu sentimenta mogu koristiti rječnici¹⁵ koji sadržavaju podatke o sentimentu riječi.

Bilo kakav se resurs koristi za izradu sustava, on se prije samoga učenja mora normalizirati¹⁶. Sama izrada sustava obično započinje određivanjem arhitekture samoga sustava: broja slojeva, tipova slojeva, broja neurona, raznih funkcija itd.¹⁷). Kada je arhitektura sustava definirana, te

⁷ O usporedbi Naivnog Bayesovog klasifikatora i klasifikatora izrađenih dubokim učenjem vidi Neppalli, Caragea C. i Caragea D. (2018).

⁸ <https://www.python.org/>

⁹ <https://wwwcplusplus.com/>

¹⁰ <https://devdocs.io/c/>

¹¹ <https://docs.nvidia.com/cuda/>

¹² <https://www.tensorflow.org/>

¹³ <https://pytorch.org/>

¹⁴ <https://github.com/Theano/Theano>

¹⁵ SentiWordNet (<https://github.com/aesuli/SentiWordNet>)

¹⁶ Najčešće je riječ o standardizaciji teksta resursa, tokenizaciji, POS i MSD označavanju, vektorizaciji, standardizacije duljine vektora itd.

¹⁷ Više o tome vidi u poglavlju 6.

podaci pripremljeni za učenje sustava, obično se korpus koji stoji na raspolaganju podijeli na skup za učenje i skup za provjeru. Kao što i njihovo ime govori, skup za učenje dio je anotiranoga korpusa koji se koristi za razvijanje ili treniranje nekoga sustava, dok je skup za provjeru dio anotiranog korpusa koji se isključivo koristi za evaluaciju sustava. Skup za provjeru dio je podataka koji se nije koristio za razvoj sustava te u analizi toga skupa vidimo sustav na *djelu*. Dok je kod skupa za učenje vrlo vjerojatno da će se svakom novom nadogradnjom dobiti sustav koji bolje poznaje skup za učenje, kod skupa za provjeru tome nije tako. Na skupu za provjeru svaki se put ispituju stvarne mogućnosti sustava. Moguće je dakle da je sustav na kraju razvojne faze slabiji nego što je bio u nekoj ranijoj verziji razvojnoga procesa. Razliku između skupa za učenje i skupa za provjeru možemo usporediti s učenjem napamet. Time što sustav zapamti cijeli skup za učenje, njegovo razumijevanje podataka (koji nisu uključeni u skup za učenje) pada, stoga kada se pred njega stavi nepoznat problem veća je mogućnost da on neće donijeti dobru odluku. Stoga, za procjenu nekoga alata relevantne su mjere vrednovanja koje se odnose na skup za provjeru. Evaluacija sustava jedan je od najvažnijih postupaka u izradi alata s dubokim učenjem i na temelju analize se odlučuje o tome kako se određeni sustav treba i može nadograditi, popraviti, ažurirati. Najjednostavnija mjera točnosti je dakako točnost u postocima, odnosno mjera podudaranja točnih i predviđenih vrijednosti iz skupa za provjeru. Međutim, ova mjera ne uzima u obzir pogreške sustava, raspodjelu odgovora, te se pri evaluaciji koriste i mjere poput F1-mjere, mjere gubitka i interpretira se i matrica zabune.

Funkcija gubitka (engl. *loss function*) nam govori o pogreškama sustava. Ovom se mjerom promatra koliko su grube pogreške sustava. Što je funkcija gubitka manja, to je sustav točniji. Uz navedene mjere, često se promatra i F1-mjera (engl. *F1-score*). Ova mjera istovremeno uzima u obzir odziv (engl. *recall*) (udio točno klasificiranih primjera u skupu svih pozitivnih primjera) i preciznost (engl. *precision*) (udio točno klasificiranih primjera u skupu pozitivno klasificiranih primjera). Naposljetku, često se interpretira matrica zabune koja prikazuje raspodjelu i odnos točnih i predviđenih vrijednosti.

4. Razvoj Twitter korpusa

Kada se govori o analizi sentimenta u internetskom okruženju istraživači, ovisno o potrebama klijenata i interesa, mogu se baviti recenzijama, novinskim člancima, komentarima, objavama s raznih društvenih mreža i slično. Analiza sentimenta je područje koje se brzo širi, a tome su zaslužne privatne tvrtke u čijem je interesu što prije doći do povratne informacije o nekome proizvodu. Dokazalo se da na dobar dio kupaca izravno utječu recenzije koje čitaju na društvenim mrežama, odnosno da mišljenja izrečena na društvenim mrežama mogu biti značajna za uspjeh ili neuspjeh nekog proizvoda ili političke opcije (Sarlan, Nadam i Basri 2014: 212–213).

Twitter¹⁸ je usluga za prijatelje, obitelj i kolege koji žele ostati u kontaktu i komunicirati razmjenom jednostavnih i čistih poruka. Twitter se u praksi također koristi kao mikroblogging platforma na kojoj pojedinci na specifičan način prenose situacije iz svoga života, reflektiraju na događanja iz svakodnevice, novosti iz dnevne politike, ekonomije i vijesti općenito (Java et. al 2007: 55). Twitter je stoga društvena mreža na kojoj korisnici pišu o sebi i o svijetu u kojemu žive. Objava se na Twitteru naziva *tvit* (engl. *tweet*) te uz sam tekst *tvita*, objava sadržava informacije o autoru, vremenu objave, broju lajkova, odgovora (engl. *reply*) i *ritvitova* (engl. *retweet*). *Tvitovi* imaju ograničenu duljinu od 280 znakova. Twitter koriste i osobe iz javnog života, pjevači, glumci, političari pa čak i organizacije i vlade. Twitter je najpopularniji u Sjedinjenim Američkim Državama, a najviše je objava objavljeno na engleskome jeziku. Prema statistikama statista.com¹⁹ i statcounter Global stats²⁰ u 2021. godini s područja Hrvatske je bilo otprilike 48 300 aktivnih korisnika Twittera. Za usporedbu Facebook je 2021. godini u Hrvatskoj imao otprilike 2 425 000 korisnika, dok Instagram 1 323 000 korisnika.

Twitter kao mikroblogging platforma puna je reakcija na proizvode i događanja u svijetu i time je idealna platforma za analizu sentimenta. Velika je prednost Twittera njegova ažurnost koja ga čini beskonačnim rudnikom novih komentara. *Tvitovi* su izrazito raznoliki i tako se u korpusu *tvitova* pronalaze iskazi koji pripadaju različitim jezičnim registrima (Java et al. 2007: 56–58). *Tvitovi* stoga čine heterogen korpus kratkih jezičnih iskaza koji je pogodan i interesantan za ljudsku, ali i automatsku analizu.

¹⁸ <https://twitter.com/>

¹⁹ <https://www.statista.com/statistics/568872/predicted-number-of-social-network-users-in-croatia/>

²⁰ <https://gs.statcounter.com/social-media-stats/all/croatia>

Tvitovima koji su objavljeni na Twitteru može se pristupiti pomoću Twitterovog API-ja (*Application Programming Interface*)²¹. Dobar dio istraživanja se upravo i oslanja na Twitterov API, međutim, standardna, besplatna verzija API-ja omogućuje samo preuzimanje (engl. *scraping*) *tvitova* nastalih u proteklih sedam dana, a i broj poslanih zahtjeva prema poslužitelju je ograničen. Dobar se dio zainteresiranih zato okreće skriptama i knjižnicama programskih jezika koji koriste Twitterov API, ali zaobilaze njegova ograničenja. Među ovim knjižnicama valja istaknuti Python knjižnice *twint*²², *TweetCrawler*²³ i *Tweepy*²⁴. Podaci (*tvitovi*) preuzeti pomoću ovih alata preuzimaju se u JSON (JavaScript Object Notation) formatu²⁵. *Tweet Object*²⁶ sadrži podatke o *tvitovom* identifikacijskom broju (*id*), tekstu *tvita* (*text*), autoru-korisniku (*user*), mjestu objave (*place*), jeziku (*lang*), ključnim riječima (*hashtag*) i drugo. Preuzimanje *tvitova* preko API-ja moguće je zadajući kriterij bilo kojemu od atributa, ali i pretraživanjem teksta *tvita*, pa se tako mogu pronaći *tvitovi* s određenim riječima i nizovima znakova, ključnim riječima, korisnicima. Pri pretraživanju se može zadati više kriterija pa je tako moguće pretražiti na primjer sve *tvitove* pisane hrvatskim jezikom, objavljene na području Francuske u razdoblju od 1.1.2021. do 1.1.2022.

Sentiment se *tvitova* može označiti na različite načine. Jedan od najočitijih je onaj s ručnom anotacijom. *Tvitovi* se dodjeljuju anotatorima koji potom odrede sentiment *tvitova*. Pri ovakvoj metodi, istraživači višestrukim označavanjem mogu nadzirati uspješnost pojedinih anotatora. Istraživači su pri ovakvoj izradi *u kontroli* i mogu pratiti čitav proces izrade korpusa. Očita je mana ovakve izrade što je ona ručna i zato izrada traje dugo. Uz to, za ovakvu anotaciju potrebni su anotatori, po mogućnosti više njih, stoga je jasno da ovakav tip anotacije zahtijeva pozamašne ljudske resurse (što često podrazumijeva i financijske).

Sentiment *tvitova* moguće je odrediti i pomoću ključnih riječi (engl. *hashtag*). Alati za preuzimanje sadržaja s interneta će ovako na primjer preuzeti sve *tvitove* koji su označeni ključnom riječju *#sreća*. Ovom metodom se jednostavno mogu dobiti korpusi koji sadrže samo pozitivne ili negativne *tvitove*, ali je već na prvi pogled jasno da ovakav pristup nije idealan za izradu

²¹ U literaturi se koriste i hrvatski termini *aplikacijsko programsko sučelje* i *sučelje za programiranje aplikacije*, međutim većinom se koristi engleski akronim.

²² <https://github.com/twintproject/twint>

²³ <https://github.com/jonbakerfish/TweetScraper>

²⁴ <https://www.tweepy.org/>

²⁵ <https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Objects/JSON>

²⁶ <https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet>

visokokvalitetnih korpusa. Izradom ovakvoga korpusa istraživač pretpostavlja da korisnici Twittera *hashtagovima* opisuju generalnu poruku samoga *tvita*, a u praksi to ne mora biti tako. Korisnici (i) na Twitteru često su sarkastični pa se ovom metodom mogu pod kamuflažom *hashtaga* provući *tvitovi* krivoga sentimenta. Ovakvi modeli mogu uključiti automatski generirane *tvitove* i neželjene poruke²⁷. Drugi automatizirani pristup prikupljanju korpusa *tvitova* s označenim sentimentom radi se filtriranjem sentimenta na temelju emotikona koji se nalaze u tekstovima *tvitova*. Ovakav pristup predviđa da će korisnici ovisno o sentimentu koji dominira u njihovome iskazu koristiti odgovarajući emotikon, pa će tako na primjer emotikoni “:-)”, “:)”, “=)”, “:D” biti u *tvitovima* s pozitivnom emocijom, a emotikoni “:-(”, “:(”, “=(”, “;(” u *tvitovima* s negativnom emocijom (Pak i Paroubek 2010).²⁸ Kao i u svakodnevnoj komunikaciji neverbalna se komunikacija (emotikoni) ne mora u svim slučajevima podudarati sa značenjem verbalne komunikacije. Nije tomu drugačije ni u komunikaciji na internetu i Twitteru. Emotikoni, iako često *jačaju* i *potkrepljuju* značenje pisanog teksta, mogu biti i oznaka suprotnih osjećaja. Emotikon „:)“ na primjer prvenstveno označava *sreću* i *radost*. U nekim kontekstima ovaj emotikon ipak može zaprimiti i neka druga značenja poput *sarkazma*, *iznenađenja* i *tjeskobe* (Walther i D’Addario 2001: 335). Kao posljedica, na ovaj način anotirani korpusi mogu negativno utjecati na izradu sustava za analizu sentimenta.

4.1. Analiza sentimenta na Twitteru nakon katastrofa

U radu je već više puta istaknuto kako se analizom sentimenta Twitter poruka istraživači prvenstveno služe kako bi brzo dobili povratnu informaciju od krajnjeg korisnika ili potrošača. Istraživači se međutim bave i drugim aspektima ljudskoga života. Jedna od proučavanih tema veže se uz ljudske traume i reakcije na ljudske traume. Posebnu skupinu ljudskih traume čine one kolektivne među koje ubrajamo ratove, smrt znamenitih i štovanih ljudi, a naposljetku i prirodne katastrofe čime se i ovaj rad bavi. Twitter se koristi za izražavanje vlastitih emocija, on funkcionira kao svojevrsni osobni dnevnik te ga korisnici u traumatičnim situacijama, dogodile se one na osobnoj razini ili na razini zajednice, koriste za izražavanje svojih osjećaja. U ovakvim slučajevima možemo govoriti o kolektivnom osjećaju, kolektivnoj reakciji na doživljenu katastrofu ili nesreću (Gaspere et al. 2016: 180). Rezultati dobiveni analizom sentimenta u

²⁷ Hasan et. al (2018) prikazuje sustav izgrađen na korpusu koji se sastavljen ovakvom metodom.

²⁸ Detaljnije o pristupima izradi korpusa vidi poglavlje 5.

katastrofama ipak ne mogu u potpunosti zrcaliti realno *stanje stvari* i društva. Istraživači moraju biti svjesni da je Twitter i dalje društvena mreža i da korisnike ne možemo izjednačavati u potpunosti sa stvarnim ljudima. Zaki (2020: 588) ističe:

Istraživači iz polja društvenih znanosti zabilježili su dva tipa ljudskoga ponašanja tijekom kriza. Prema prvoj pojedinci (i) paničare, (ii) ignoriraju društveni red, i (iii) ponašaju se sebično. Ovaj skup mišljenja karakterizira mišljenje popularnih medija za vrijeme nesreća i predviđanju ljudskoga ponašanja. (...) Drugo mišljenje je potkrijepljeno povijesnim činjenicama. (...) U uzrujanome stanju, oni koji su preživjeli katastrofu formiraju zajednice za međusobnu pomoć, sudjeluju u mnogobrojnim altruističnim činovima i pokazuju istaknutu solidarnost jedni prema drugima.

Prihvatimo li ove misli, tvitovi nisu zrcalo stvarnoga svijeta. *Tvitovi* često hipertrofiraju negativne pojave u društvu i zanemaruju pozitivne reakcije zajednice. *Tvitovi* prikazuju *jednu* sliku svijeta. Analizom *tvitova* uslijed katastrofa možemo promatrati generalnu reakciju zajednice i kolektivne osjećaje prema društvenim pojavama *na društvenim mrežama*. Ti nam podaci mogu govoriti i o *stvarnoj* slici, ali moramo biti svjesni konteksta u kojemu se ovi jezični iskazi javljaju. Na temelju ovakvih istraživanja može se zaključiti kako pojedinci reagiraju na nesreće na društvenim mrežama, kojim se mehanizmima koriste za procesuiranje katastrofe, tko je najčešće predmet ruganja i kako se općenito nose s novonastalom situacijom.

22. ožujka 2020. godine Zagreb je pogodio potres magnitude 5,3 po Richteru. Štete su u gradskome centru i naseljima podno Sljemena bile velike, a ljudi, infrastruktura i zgrade nepripremljene za potres koji je rezultirao kaosom. Ljudski je strah i stres bio jasan svima koji žive na ovim prostorima te je njega povećavao i strah od tada još novoga i nepoznatoga koronavirusa. 19. ožujka 2020. godine Stožer civilne zaštite RH je donio odluku o potpunom zatvaranju Republike Hrvatske u svrhu sprječavanja širenja zaraze novim koronavirusom. Medijskim prostorom i društvenim mrežama dominirale su vijesti vezane za potres. Tijekom nadolazećih mjeseci Zagrepčane su i dalje budili manji potresi te su oni i dalje bili svakodnevna tema u medijskome prostoru.

Za manje od godinu dana, 29. prosinca 2020. dogodio se drugi, još razorniji potres s epicentrom u okolici Petrinje magnitude 6,2 prema Richteru. Potres u Petrinji razorio je Glinu i Petrinju, ostavljajući stotine bez domova i krova nad glavom. Dodatno, novi je potres oštetio nesanirane zgrade u Gradu Zagrebu koje su bile oštećene u potresu 22. ožujka. Medijski i društveni život se

opet okrenuo potresima, traumama, izgubljenim životima i tragičnim pričama. Na društvenim mrežama, pa tako i na Twitteru, pratila su se dnevna događanja te su novinari i korisnici društvenih mreža reflektirali na pojave vezane za potres.

5. Izrada anotiranoga Twitter korpusa

5.1. Metodologija

Za izradu nekoga jezičnoga alata nužno je imati jezične resurse s pomoću kojih alate možemo učiti, testirati i primijeniti. Jezični se resursi, ponavljamo, mogu izraditi na mnoštvo načina. Postupak može biti automatski, poluautomatski ili u nekim slučajevima i posve ručni. Svaki od navedenih pristupa ima svoje prednosti i mane. Automatskom se izradom u kratkom vremenskom roku dobije veliki jezični resurs. Ovakvi resursi, međutim, mogu biti *prljavi*, odnosno u njemu se nalaze nefiltrirani i neprobrani materijali. Ovaj tip izrade jezičnoga resursa može se na primjer koristiti pri izradi web-korpusa neke vršne internetske domene²⁹. Ovakav se pristup koristi i pri izradi jezičnoga resursa za analizu sentimenta koji se sastoji od jezičnih jedinica koje je autor označio s emotikonima ili ključnim riječima³⁰. Jezični resursi izrađeni poluautomatskom metodom automatiziraju neke korake izrade jezičnoga resursa, ali istovremeno zadržavaju prednosti ljudskoga nadzora i direktne intervencije. Jezični resursi izrađeni ovakvom metodom najčešće odlikuju balansiranim omjerom količine podataka i kvalitete sadržaja. Pri izradi korpusa *tvitova* s označenim sentimentom najčešće se koristi ovaj pristup³¹. Treća metoda izrade jezičnoga resursa je ručna. Izrada ovakvih korpusa isključuje bilo kakvu automatizaciju. Ručnom se obradom mogu dobiti vrlo kvalitetni jezični resursi, ali je izrada ovakvih jezičnih resursa dugotrajna.

U ovome radu je za izradu korpusa *tvitova* s označenim sentimentom korištena druga, poluautomatska metoda. Ona nam se činila najprikladnijom jer je omogućila nadgledanje procesa prikupljanja i filtriranja podataka, a istovremeno je na ovaj način bilo moguće izraditi jezični resurs primjerene veličine. Suprotno od automatske izrade, poluautomatski pristup se uvijek dijeli na nekoliko etapa, stoga se i izrada ovoga resursa može podijeliti na nekoliko etapa. Prvo su se preuzeli *tvitovi*, potom su se preuzeti *tvitovi* očistili, odnosno napravila se selekcija i normalizacija *tvitova* za anotaciju, zatim su anotatori označili *tvitove*. Četvrti korak se sastojao od provjeravanja suglasnosti među anotatorima te se naposljetku izradio reprezentativni korpus za učenje.

²⁹ Usp. Jakubiček et. al (2013) i Ljubešić i Erjavec (2011).

³⁰ Usp. Hogenboom et al. (2013) i Kouloumpis, Wilson i Moore (2011).

³¹ Usp. Krajcar (2014) i Raguzin (2018).

U ovome radu smo za preuzimanje *tvitova* odabrali knjižnicu *twint*³² programskoga jezika Python. *Tvitovi* preuzeti s *twint* knjižnicom i dalje sadrže sve metapodatke koji se obično javljaju uz *tvitove*. Ovo nam je omogućilo sortiranje *tvitova* prema njihovome jeziku i datumu nastanka. Uvidom u korpus, uočeno je da se među preuzetim *tvitovima* nalazi velika količina *tvitova* na drugim jezicima, većinom engleskome. Korpus je zahtijevao dodatno filtriranje *tvitova* na temelju jezika. Koristeći *langid*³³ Python knjižnicu napravljena je dodatna jezična identifikacija pri kojoj su se iz postojećeg korpusa izdvojili samo *tvitovi* pisani na hrvatskome jeziku. Za dodatnu analizu jezika odabrali smo *langid* knjižnicu, jer je ona u odnosu na druge popularne i ažurne knjižnice za detekciju jezika (*LangDetect*³⁴, *TextCat*³⁵, *CLD*³⁶) najtočnija (Lui i Baldwin 2012: 27–29).

Pri preuzimanju *tvitova* vodili smo se pretpostavkom da su *tviteraši* Twitter koristili kao mikroblogging platformu te su u skladu s time proizveli veliku količinu *tvitova* koji su, s obzirom na kontekst u kojemu su nastali, nabijeni emocijama. Korpus je sastavljen od *tvitova* čija je selekcija ovisila o nekoliko parametara. U korpus su uključeni *tvitovi* nastali od 22.3.2020. (potres u Zagrebu) do 3.3.2021. S Twitterove aplikacije su preuzeti *tvitovi* koji sadržavaju niz znakova (engl. *string*) *potres*, *petrinja*, *glina*, *banija*³⁷. Tijelo (engl. *body*) *tvitova* je u JSON objektu definiran kao niz znakova, stoga je pretraživanje spomenutih nizova uključilo i *tvitove* koji su označeni ključnim riječima *#potres*, *#petrinja*, *#glina*, *#banija*. Preuzeti su samo *autorski tvitovi*, odnosno *ritvitovi* su isključeni iz korpusa.

O količinskome minimumu podataka potrebnih za izradu alata za analizu sentimenta ne postoje čvrste upute. Neki autori govore o početnome korpusu od 5000 označenih jezičnih jedinica koje se onda po potrebi proširuju s dodatnih 2000 do 5000 jedinica dok se ne postigne željeni rezultat³⁸. Usporedbom radova sličnih parametra, mogućnosti i opsega, ustanovljeno je da su istraživači koji su izrađivali korpuse s označenim sentimentom anotirali između 3 000 i 4 000 jezičnih jedinica (usp. Krajcar 2014). Na temelju uvida u literaturu i preporuke stručnjaka za razvoj alata za analizu

³² <https://github.com/twintproject/twint>

³³ <https://github.com/saffsd/langid.py>

³⁴ <https://pypi.org/project/langdetect/>

³⁵ https://www.nltk.org/_modules/nltk/classify/textcat.html

³⁶ <https://pypi.org/project/pyclld2/>

³⁷ Pretraživanje nije bilo osjetljivo na razliku između velikih i malih slova.

³⁸ O raspravi o minimumu podataka vidi *What is the minimum training dataset size to train a sentiment analysis classifier (social media corpus)?* (Haroutunian: 2016).

sentimenta, odnosno uzimajući u obzir mogućnosti anotatora, odlučeno je da će se anotirati korpus od 5 000 *tvitova*.

S obzirom na to da se za izradu jezičnoga resursa odlučila koristiti poluautomatska izrada, anotacija se radila ručno. Korpus od 5000 *tvitova* podijelili smo na skupove od 500 *tvitova*. Korpus je ovakve veličine moguće označiti u periodu od dva do tri sata. Vrijeme nužno za obradu bio je važan faktor, jer anotatori nisu bili plaćeni, već su anotaciju radili volonterski³⁹. Anotatori su označavali jezične jedinice odabirom jedne od četiriju kategorija: „pozitivno“, „negativno“, „neutralno“ i „ne može se odrediti“. Odlučili smo se za trojnu sentimentnu podjelu (s jednom dodatnom kategorijom u slučaju da se sentiment ne može odrediti) kako bismo dobili jezični resurs prikladan za izradu alata za određivanje sentimenta. Skale za određivanja sentimenta od pet ili sedam stavki rezultiraju u korpusima s više odudaranja, te su sustavi za označavanje sentimenta trenirani na njima manje točni (usp. Gräbner et. al 2012, Tripto i Ali 2018). Mnogobrojnije podjele također zahtijevaju veću količinu podataka i veću angažiranost anotatora. Iako smo odabrali trojnu sentimentnu podjelu, anotatori su mogli odabrati jednu od četiriju kategorija. Omogućavanjem opcije “ne može se odrediti”, zaobišli smo probleme koje bi mogli prouzrokovati potencijalni neoznačeni *tvitovi* u korpusu. Anotatori su prije anotacije pročitati detaljne upute za anotaciju. U uputama se sugeriralo da se anotatori oslone na svoju intuiciju i da se usredotoče na sentiment (ili dominantnu emociju) jezične jedinice. Anotatori su upozoreni na moguće dvoumice u slučajevima ironije, sarkazma i drugih *osjećajno* složenih jezičnih iskaza. Osim navedenoga, nikakve se sugestije nisu dale anotatorima, već su se u uputama nalazile dva ogledna primjera za svaku od mogućih kategorija, izuzevši kategoriju „ne može se odrediti“. U uputama se također jasno izrazila mogućnost kontaktiranja autora rada u slučaju bilo kakve nedoumice. Neki od anotatora su kontaktirali autora rada dvoumeći se oko sentimenta nekoga *tvita*. Mahom u svim slučajevima tu je bilo riječi o *tvitovima* s ironijom i sarkazmom. Kako bismo dobili jasniju sliku o jezičnom materijalu koji se označava, odlučili smo se za dvostruku anotaciju. Sukladno tome, svaku je jezičnu jedinicu sentimentom označilo dva anotatora.

U računalnoj se lingvistici od 1990-ih godina može uočiti težnja k empirizaciji, te se kao dio toga nakon postupka anotacije počela računati i interpretirati mjera međuanotatorske pouzdanosti

³⁹ Ovim putem zahvaljujem svim anotatorima. Poimence: Anja Džeko, Ana Acalin, Anja Delić, Bernadett Turul, Branka Blazsetin, Dorotea Cakić, István Blazsetin, Iva Žura, Josipa Erceg, Lucija Bojić, Luka Džeko, Vjekoslav Blazsetin, Maja Matijević, Marin Lisjak, Marta Brkić, Mišo Džeko, Suzana Džeko, Sven Krakar, Tamás Turul.

(engl. *inter-rater reliability*). Ova se mjera razlikuje od jednostavnoga postotka podudarnosti, jer uzima u obzir i mogućnost da se sporazum među ocjenjivačima dogodi slučajno. Cilj je međuanotatorske pouzdanosti pokazati da su ručno anotirani jezični resursi *pouzdati*, odnosno da su kategorizacijske sheme i anotatori prikladni te oni potkrepljuju *valjanost* („*validity*“) i *istinitost* („*truth*“) istraživanja (Artstein i Poesio 2008: 555–557). Postoje razne mjere međuanotatorske podudarnosti, a najčešće se koriste Cohenova kappa, Krippendorfova alpha i Fleissova kappa (Artstein i Poesio 2008: 590). U ovome smo se radu odlučili na računanje Cohenove kappe za međuanotatorsku pouzdanost. U radovima s područja obrade prirodnoga jezika i računalne lingvistike istraživači najčešće koriste Cohenovu kappu, te je ona postala svojevrsnim standardom (Hsu i Field 2003: 206). Cohenova kappa definira se na sljedeći način, pri čemu ρ_0 predstavlja relativno slaganje ocjenjivača, a ρ_e pretpostavljenu vjerojatnost slučajnog slaganja na temelju podataka:

$$\kappa = \frac{\rho_0 - \rho_e}{1 - \rho_e}$$

Formula 2: Cohenova kappa

Iako se Cohenova kappa smatra jednom od najpopularnijih mjera međuanotatorske pouzdanosti, Krippendorf ju kritizira i formira svoju alphu kao zamjenu za Cohenovu kappu. Postoji nekoliko istraživanja koja pokazuju da je Krippendorfova alpha skalabilnija od Cohenove kappe i da bi se ona trebala preferirati⁴⁰, međutim usporedbom Cohenove kappe i Krippendorfove alphe na podacima iz ovoga rada, zaključili smo da se razlika između dvije mjere, u ovome slučaju, može zanemariti. Za Cohenovu smo se kappu odlučili da bismo rad lakše smjestili u kontekst. Kako bi se Cohenova kappa mogla primijeniti na neki skup podataka, on mora zadovoljiti nekoliko preduvjeta ili pretpostavki. Cohenova kappa za mjerenje međuanotatorske pouzdanosti može se koristiti samo kada mjerimo pouzdanost između dvaju anotatora (za razliku od Krippendorfove alphe). Anotatori među kojima se računa pouzdanost moraju označiti isti skup podataka te je važno da su kategorije za označavanje isključive, tj. da se ne preklapaju. Dva anotatora uvijek moraju koristiti identičnu podjelu i u anotiranim podacima ne smiju postojati prazni (neoznačeni) podaci. Podaci predstavljeni u ovome radu ispunjavaju sve navedene uvjete.

⁴⁰ Usp. Friese (2020).

Naposljetku, kako bismo dobili jezični resurs koji je prikladan za izradu jezičnoga alata, izradili smo zlatni standard koji će služiti za učenje alata za analizu sentimenta. Prvo smo pogledali omjer zastupljenosti odgovora svakoga pojedinačnog anotatora te potom izračunali prosjek zastupljenosti svakoga odgovora ili kategorije, neovisno o drugim anotatorima. Zatim, stvorili smo potkorpus od *tvitova* oko čijih su se sentimenta anotatori složili. U ovome smo potkorpusu također izračunali omjer kategorija. Nakon njihove usporedbe, zaključili smo da je razlika između prosječne vrijednosti omjera kategorija i omjera kategorija u potkorpusu zanemariv, te smo ovako dobili jezični resurs za treniranje alata za analizu sentimenta.

5.2. Rezultati

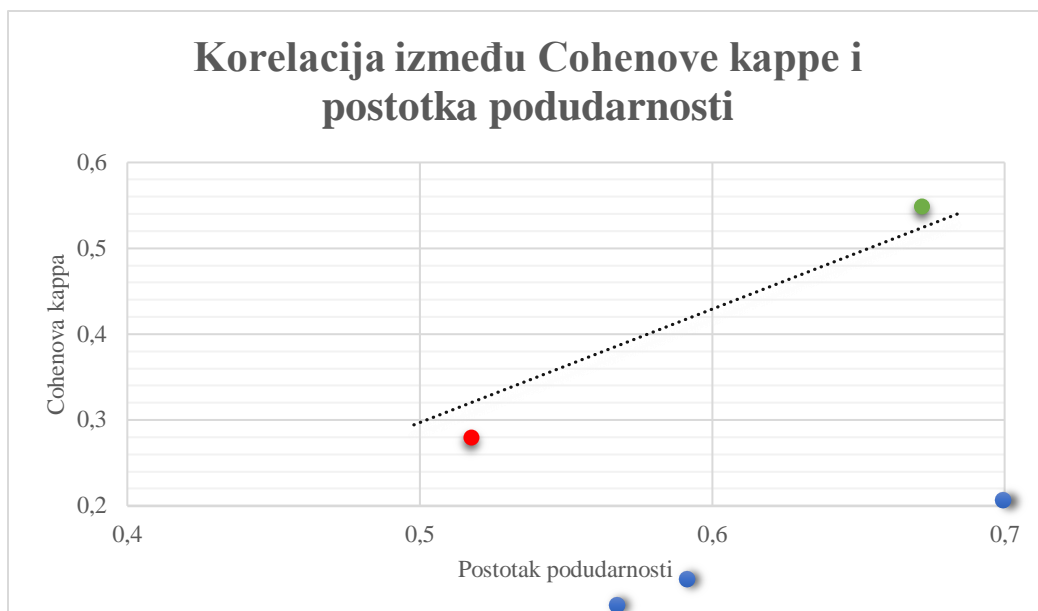
Pri preuzimanju *tvitova*, naš se inicijalni korpus, filtriran Twitterovim jezičnim filtriranjem, sastojao od 30 321 *tvita*. Nakon dodatne provjere jezika *tvitova* s pomoću *langid* knjižnice, dobiven je korpus od 14 745 *tvitova*. Od toga je sačinjen nasumičan potkorpus od 5000 *tvitova* koji je podijeljen u dvadeset potkorpusa od 500 *tvitova* te su oni dodijeljeni anotatorima. Kao što je već spomenuto, svaki je potkorpus anotiralo dvoje anotatora.

U tablici 1. vidimo postotak međuanotatorske podudarnosti i vrijednost Cohenove kappe u potkorpusima i njihove prosječne, minimalne i maksimalne vrijednosti.

	Postotak podudarnosti	Cohenova kapa (κ)
1. potkorpus	0,68	0,513
2. potkorpus	0,684	0,529
3. potkorpus	0,642	0,512
4. potkorpus	0,518	0,279
5. potkorpus	0,586	0,396
6. potkorpus	0,514	0,335
7. potkorpus	0,672	0,548
8. potkorpus	0,616	0,449
9. potkorpus	0,498	0,315
10. potkorpus	0,641	0,483
minimalna vrijednost	0,498	0,279
maksimalna vrijednost	0,684	0,548
prosječna vrijednost	0,605	0,436⁴¹

Tablica 1: Međuanotatorska podudarnost u potkorpusima

⁴¹ Prosječna vrijednost Krippendorfove alphe je 0,4285.



Graf 1: Korelacija između Cohenove kappe i postotka podudarnosti

Graf 1. zorno prikazuje korelaciju između postotka podudarnosti i Cohenove kappe za svaki anotirani potkorpus. Zelena točka predstavlja 7. potkorpus, dok crvena 4. potkorpus.

Izračun Cohenove kappe zahtijeva nekolicinu informacija i izračuna o *navikama* pojedinih anotatora, stoga smo za izračun međuanotatorske pouzdanost formirali tablice. Tablica 2. prikazuje anotaciju 7. potkorpusa (potkorpus s najvećom međuanotatorskom pouzdanosti).

		Anotator A					zastupljenost
		pozitivno	neutralno	negativno	ne može se odrediti	ukupno	
Anotator B	pozitivno	73	25	29	40	167	0,334
	neutralno	2	119	5	3	129	0,258
	negativno	4	33	130	11	178	0,356
	ne može se odrediti	2	7	3	14	26	0,052
	ukupno	81	184	167	68	336	1
	zastupljenost	0,162	0,368	0,334	0,136	1	

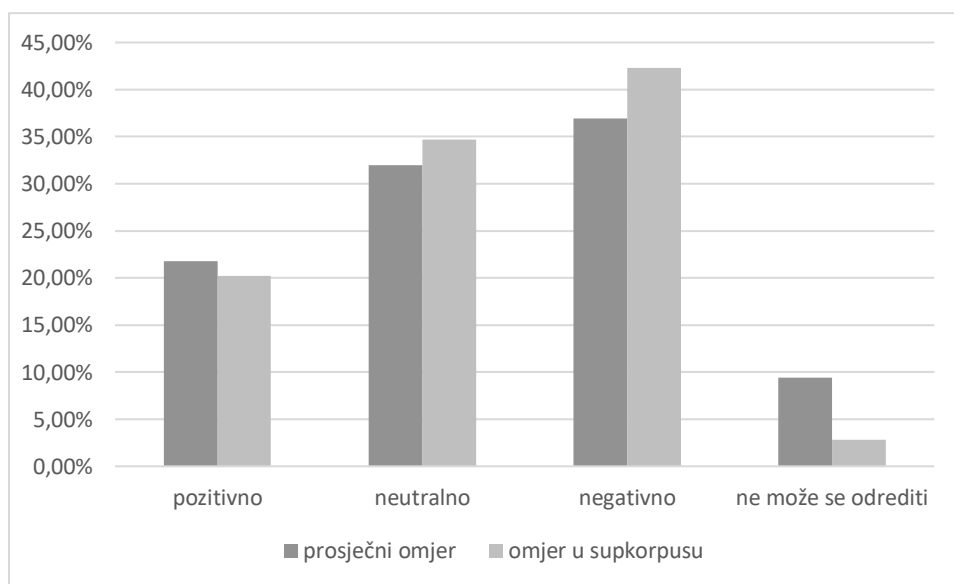
Tablica 2: Podaci o anotaciji 7. potkorpusa

Potkorpus sačinjen od onih *tvitova* u kojima su oba anotatora odabrali istu kategoriju, ima 3026 *tvitova*. Ovaj potkorpus čini 65% sveukupnog korpusa koji ima 5000 *tvitova*. Nakon dodatne analize, odlučili smo da će ovaj korpus poslužiti kao skup za učenje za razvoj alata za analizu sentimenta.

Kako bismo mogli odrediti je li izdvojen potkorpus od 3026 tvitova reprezentativan, usporedili smo omjer kategorija⁴² u tome korpusu s prosječnom vrijednosti omjera kategorija svih anotatora. Brojčano, u potkorpusu od 3026 tvitova nalazi se 611 pozitivnih, 1050 neutralnih i 1280 negativnih tvitova. *Tvitova* s kategorijom *ne može se odrediti* ima svega 85. Prikaz usporedbe omjera potkorpusa i prosječne vrijednosti omjera kategorija izoliranih anotatora možemo vidjeti u tablici 3 i na grafu 2.

	pozitivno	neutralno	negativno	ne može se odrediti
prosječni omjer	0,218	0,32	0,369	0,094
omjer u potkorpusu	0,202	0,347	0,423	0,028
razlika	0,016	-0,027	-0,054	0,066

Tablica 3: Prosječni omjer kategorija svih anotatora, omjer kategorija u potkorpusu i njihova razlika u postotnim bodovima



Graf 2: Prosječni omjer kategorija svih anotatora, omjer kategorija u potkorpusu

Iako smo u ovoj usporedbi koristili samo prosječni omjer kategorija, svakako je zanimljivo pogledati i neke različitosti među anotatorima u omjeru kategorija. U tablici 4. donosimo anotatora s najvećim i najmanjim omjerom kategorije *ne može se odrediti*.

⁴² Vidi više u poglavlju 6.3.3. Omjer kategorija

Anotator	Udio kategorije ne može se odrediti
anotator 1	0,008
anotator 18	0,244

Tablica 4: Anotator s najvećim i najmanjim udjelom kategorije ne može se odrediti

5.3. Interpretacija rezultata

5.3.1. Kontekstualizacija Cohenove kappe

U tablici 1. prikazane su vrijednosti međuanotatorske pozdanosti. Možemo reći da rezultati nisu izrazito dobri, odnosno da se anotatori u puno slučajeva nisu složili. Postotak podudaranja u prosjeku iznosi 0,605, što nam govori o tome da su se anotatori u više nego pola slučajeva složili. Postotak podudarnosti nije posve mjerodavan u računanju međuanotatorske pouzdanosti stoga ćemo se usredotočiti na analizu Cohenove kappe.

Cohenovu kappu dobivenu u ovome radu kontekstualizirat ćemo pomoću podjele koju predlažu Landis i Koch (1977:165):

Kappa	Mjera podudarnosti
<0.00	loša
0.00–0.20	neznatna
0.21–0.40	prihvatljiva
0.41–0.60	umjerena
0.61–0.80	znatna
0.81–1.00	skoro savršena

Tablica 5: Razredi podudarnosti prema Landis i Koch (1977:165)

Prema ovoj klasifikaciji, prosječna je međuanotatorska pouzdanost u ovome radu umjerena. S najnižom kappa vrijednosti od 0,279, niti jedno izračunato međuanotatorsko podudaranja ne ulazi u kategorije „loša“ i „neznatna“ podudarnost, ali isto tako, niti jedna kappa vrijednost ne doseže kategoriju „znatne“ podudarnosti. Međutim, kao što Sim i Wright (2005) ističu ovakve podjele na kategorije u većini slučajeva nisu mjerodavne, one su „neizbježno arbitrarne“. Pri prikazu Cohenove kappe uz pojedinačne vrijednosti i prosječnu vrijednost, radi kontekstualizacije, valja kao što smo to i napravili, istaknuti maksimalnu i minimalnu kappa vrijednost. Pri kontekstualizaciji također moramo obratiti pozornost na prirodu anotiranoga materijala.

Za interpretaciju Cohenove kappe među anotatorima moramo uzeti u obzir da je označavanje sentimenta posve subjektivan zadatak. Odrediti sentiment nekoga jezičnoga iskaza, kao što smo u

prvome dijelu rada već zaključili, nije uopće lagan zadatak. Iako se možda čini da proučavanje međuanotatorske pouzdanosti na polju analize sentimenta nije mjerodavno, ovaj rad nije jedini koji računa međuanotatorsku pouzdanost. Bobicev i Sokolova (2017: 100) donose kappa vrijednost anotacije jezičnih iskaza četirima emocijama. Prosječne kappa vrijednosti nalaze se između 0,38 i 0,50. Takala et al. (2014: 2156) u radu o označavanju jezičnih iskaza sentimentom (koristeći trojnu sentimentnu podjelu) prikazuju Cohenove kappe od 0,682 do 0,728, dakle među njihovim je anotatorima znatno veća podudarnost.

Postoji nekoliko mogućih uzroka tome da je kappa vrijednost anotiranoga korpusa ovoga rada *samo* umjerena, a da je podudarnost u drugim prikazanim radovima veća. Takala et al. (2014) su usporedbu radili na svega 42 anotirana dokumenta (blogova) iz domene financija. Usporedimo li tu količinu anotiranih podataka s podacima ovoga rada, možemo zaključiti da je anotirani korpus ovoga rada višestruko veći, točnije 119 puta. Drugo, u ovome radu, uz tri kategorije sentimenta, anotatorima je bila ponuđena i opcija “ne može se odrediti”, koju Takala et. al (2014) nisu imali. Jasno je da se povećavanjem mogućih odabira, povećava i mogućnost da se anotatori neće složiti. Čini nam se da bi razlika u veličini korpusa i mogućim odgovorima ponuđenim anotatorima mogla utjecati na kappa vrijednost.

5.3.2. Dvosmisleni tvitovi

Pri kontekstualizaciji Cohenove kappe, moramo uzeti u obzir i prirodu jezičnoga iskaza koji se anotirao. Dok su anotatori u Takala et al. (2004) označavali čitave blog objave, anotatori ovoga rada označavali su *tvitove* koji su ograničeni opsegom, dopuštaju posve neformalni jezični diskurs i često sadrže emotikone. Ovakvi se jezični iskazi u većini slučajeva, zbog manjka informacije, teže analiziraju. Često je teško odrediti značenje nekoga *tvita*, kamoli njegov sentiment. Uvidom u korpus ustanovili smo nekoliko obrazaca u kojima je često došlo do nesporazuma pri anotaciji.

Kao što se i moglo očekivati, određivanje sentimenta sarkastičnih i ironičnih *tvitova* podijelilo je anotatore. O tome, je li sarkazam pozitivan ili negativan iskaz, odnosno o razlikama ironije i sarkazma, nećemo raspravljati jer ono otvara bezbroj pitanja i mnogobrojne interpretacije. Sarkazam ćemo, po uzoru na Maynarda i Greenwooda (2014), izjednačiti s ironijom i promatrati ga kao „izjava u kojoj iskazivač misli na suprotno od onoga što kaže“ (2014: 3). Iako Maynard i Greenwood pokazuju da je u njihovome korpusu većina sarkastičnih *tvitova* negativnoga sentimenta, u našem bi slučaju to bilo neodgovorno reći. Čini se da su anotatori u nekim

slučajevima sarkazam smjestili u područje *humorističnoga* i prvenstveno mu pripisali pozitivni sentiment.

Donosimo nekoliko primjera sarkastičnih ili sarkastično-humorističnih tvitova:

- a) Nova jedinica osjetljivosti na #potres: 1 Purger = 0,1 Richtera
- b) Škola, naravno, već godinama ne odgovara niti pedagoškom standardu Tadžikistana (neka se ne uvrijede moji dragi prijatelji Tadžikistanci), a kamoli standardima 21. stoljeća, no svakako možemo biti zadovoljni i sretni što se potres dogodio na neradni dan.
- c) Vizionari - nisu skroz obnovili Petrinju nakon rata jer su znali da će za 25 godina biti potres.

U svim navedenim slučajevima jedan je od anotatora *tvit* označio pozitivnim, a drugi negativnim sentimentom. Teško je reći tko je u pravu, teško je reći što iskaz čini sarkastičnim. Možda možemo pronaći neke razlike. Vidimo da je *tvit a* suptilniji od *tvita b*. Oba ismijavaju, međutim, dok za razumijevanje *tvita a* treba biti upoznat s podtekstom „Zagrepčani su s vremenom postali dosadni jer su na svaki mali trzaj mislili da je potres te svoja opažanja objavljivali na društvenim mrežama“, za razumijevanje je *tvita b* dovoljno pročitati *tvit* jer se u njemu imenuje *objekt* ismijavanja, školski sustav. Međutim, neupitno, oba su *tvita* sarkastična i čini nam se da je razlika izvan njih, isključivo u percepciji anotatora koji ove *tvitove* mogu smatrati ili ruganjem ili humorom. Svakako valja istaknuti, da ovaj problem nije jedinstven ovome radu, već je simptom analize sentimenta⁴³.

Druga situacija u kojoj su se anotatori razilazili je ona u kojoj se negativno događanje prikazuje u humorističnome kontekstu. Ovi su iskazi često podrugivanja i ismijavanja, ali isto tako, u dosta se slučajeva temelje na samoironiji. Donosimo nekoliko primjera ovakvih *tvitova*:

- a) Korona, potres, snijeg, HDZ - nikad kraja...
- b) jeste ga osjetili? nije bio potres, nego sam pokušao svima izmanifestirati orgazam
- c) Korona, potres, potop, jel su javili kada eruptira Sljeme? #ZG2020

Jasno je podrijetlo nesporazuma među anotatorima. U percepciji jednoga anotatora težište autora izjave je na prouzrokovanju smijeha u čitateljima, dakle sentiment je pozitivan, dok u percepciji drugoga, težište je autora na traumatičnom iskustvu, posljedično, sentiment je negativan. Očekivati

⁴³ Uostalom, pitanje je višestruko, interdisciplinarno. Rasprava o humoru i sarkazmu je prisutna i u filozofiji od Platona do Žižeka.

sustavnu podudarnost u ovim situacijama nerealistično je jer odluka i u ovome slučaju ovisi isključivo o percepciji humora.

Treći slučaj u kojemu se anotatori nisu slagali je prenošenje vijesti ili obavijesti na Twitteru. S jedne strane, vijesti i obavijesti se same po sebi predstavljaju kao objektivne istine, a kao takve, one su neutralne. S druge strane, vijesti i obavijesti o potresu često su puni tragičnih događaja pa se tako doimaju negativnim. Slijedi nekoliko primjera vijesti koje su anotatori označili različitim sentimentom:

- a) [01:20:41 UTC] #potres otkriven na 6 km od #Zagreb, #Hrvatska. Preuzmite aplikaciju Earthquake Network s <https://t.co/N9Zyr5BxyY> da biste u stvarnom vremenu primali upozorenja na svom pametnom telefonu. <https://t.co/lcfIUPxqV8>
- b) Danas je skinut neupotrebljiv dio krova i bor je platio svoj danak, ali ne u cjelini. #zagreb#potres
- c) Budite pripremljeni za potres: Što sve trebate imati u ruksaku <https://t.co/YLWTC08RSd>
<https://t.co/NOGN0zymOP>

Za razliku od prijašnjih situacija, u ovim slučajevima, anotatori nisu tvitovima pridali suprotne sentimente, već su ih neki smatrali negativnim, dok drugi neutralnim. Sentiment ovih tvitova ovisi o tome u kojoj mjeri određeni anotator uzima u obzir kontekst u kojemu se *tweet* javlja. Promatramo li ove *tweetove* prvenstveno u njihovome užem kontekstu, odnosno njih promatramo kao izjavne rečenice koje prenose neku obavijest o nečemu što *se objektivno dogodilo*, *tweet* ćemo kategorizirati neutralnim. Gledamo li ove *tweetove* u širem kontekstu (koji nam je često potreban za analizu sentimenta, a to smo pokazali na primjeru sarkazma) *tweet* će biti negativan jer govori o traumama i nesretnim događajima, u ovome slučaju o potresima, o načinima *obrane* protiv potresa ili često o izjavama stradalih ljudi. Jasno je stoga da se anotatori u ovim slučajevima nisu složili zbog toga što su davali prednost različitim aspektima iskaza.

5.3.3. Omjer kategorija

U tablici 3. prikazani su podaci o omjerima kategorija. Usporedili smo prosječni omjer kategorija anatora s omjerom kategorija u potkorpusu. Iako omjer nije posve istovjetan, činilo nam se da je potkorpus dovoljno reprezentativan, s obzirom na to da razlika u omjeru ni u jednoj kategoriji ne prelazi 10 postotnih bodova. Iako smo ovakvim formiranjem korpusa za učenje odbacili skoro 40% cjelokupnog korpusa, smatramo da i dalje imamo dovoljno velik korpus za treniranje alata za određivanje sentimenta. U obama korpusima najzastupljenija je *negativna* kategorija, što smo s

obzirom na temu korpusa i mogli očekivati. Najveći se raskorak između dva korpusa događa u kategoriji *ne može se odrediti*. Doista je malo *tvitova* u kojima su se oba anotatora složili da se sentiment ne može odrediti. Donosimo nekoliko takvih primjera:

- a) Zamislim nebeski potres jacine dueta Zigon - Crnogorcevic, a pravoslavlje prsti na sve strane ko kad je bog stvaru svet, pa strnuo u prazninu
- b) @marina_b612 @Luka_Duvnjak @the_simulacra @NandinoLoncar @LoncaricSa umisto sto mu u pluca zrak upuhavaju i time ga salju u grob-zaboravili su na mix koke/keta...a mozda i malo whiskya...i sad je napeto ko mujin treger-boze scuvaj bozzera...i fluka ne izdaji me stalno jer cu ti razbit advokaturu vise neg potres
- c) Kad je bio potres u Zagrebu moja sestra nije bila tama na fakultetu 3gofinu uspješno završena bit će najbolji doktor u Americi

Iz ovih je *tvitova* vidljivo da se podudaranje anotatora u ovoj kategoriji dogodilo kada je nemoguće razumjeti neki *tweet*. Moguće je da ovi *tweetovi* u kontekstu interakcije dvoje ljudi imaju suvislo značenje, ali izolirano, njihovu je poruku vrlo teško shvatiti, a tako se ni njihov sentiment ne može odrediti.

5.3.4. Razlike u neodlučnosti anotatora

Iako je slučajeva u kojima su se anotatori složili oko toga da se sentiment ne može odrediti malo, svega 2,8%, na individualnoj su razini anotatori prosječno u 9,4% slučajeva osjećali da je sentiment neodrediv. Ipak, važnije je pogledati minimalne i maksimalne vrijednosti omjera ove kategorije. Kao što se u tablici 4. vidi velika je razlika između maksimalne i minimalne vrijednosti. Minimalna vrijednost je 0,8% (Anotator A), a maksimalna 24,4% (Anotator B). Kako bismo mogli kontekstualizirati ove brojke valjalo bi pogledati podatke anotatora s kojima su oni bili upareni. Anotator koji je označavao isti podskup kao *Anotator A*, *ne može se odrediti* dodijelio je 3,4% podskupa. Anotator koji je označavao isti podskup kao *Anotator B*, *ne može se odrediti* dodijelio je 6% podskupa. Možemo zaključiti da je *Anotator B* premašio *normalnu* zastupljenost kategorije *ne može se odrediti*, a to je potvrđeno činjenicom da je udio omjera te kategorije kod suanotatora višestruko manja. Uvidom u podatke, možemo reći da je *Anotator B* bio neodlučan te je često *tweetovima* kojima je *Anotator A* dodijelio neutralni sentiment on odredio kategoriju *ne može se odrediti*. Donosimo nekoliko primjera kako bismo prikazali ovaj problem:

- a) #potres Osjetio i ja ovaj mali potres
- b) Ovaj potres je tridesetak puta jači od onoga u ožujku u Zagrebu. 🙏
- c) @IKopric Dušom i mislima uz vas 🙏🙏🙏🙏🙏🙏 HR #potres #ljudi #love #ljubav #solidarnost #coronavirus #djeca

Uvidom u korpus, prepoznali smo nekoliko aspekata karakterističnih za analizu sentimenta općenito, ali i za analizu sentimenta na Twitteru. Možemo zaključiti da je *tvitu* kao tekstnoj vrsti često teško odrediti sentiment zbog njegove kratkoće i oslanjanja na kontekst. Emotikoni mogu pomoći pri identifikaciji sentimenta jer oni služe kao pojačivači emocije. Uočili smo također da problem predstavljaju sarkastični, ironični i humoristični *tvitovi* koji mogu biti pozitivni ili negativni ovisno o točki gledišta, ovisno o anotatoru. Analiza razlike između cjelokupnog korpusa i potkorpusa pokazala je obrasce u omjeru kategorija što bi moglo predstavljati problem i u budućim istraživanjima sentimenta na Twitteru. Uz sve navedene i analizirane probleme, smatramo da je anotirani korpus vrijedan jezični materijal i da može poslužiti raznim istraživanjima, na primjer, leksičkoj analizi *tvitova* ili za izradu alata za analizu sentimenta.

6. Izrada alata za određivanje sentimenta

6.1. Metodologija

Za izradu alata za analizu sentimenta koristit ćemo se metodom povratnih neuronskih mreža. Ovu smo metodu odabrali, jer je posljednjih godina znanost o dubokom učenju izrazila veliku zainteresiranost za sustave ovakve strukture, nadajući se da će oni ostvariti bolje rezultate od sustava izgrađenih tradicionalnim metodama (Sarker 2021: 1–5). Dodatno, mali je broj hrvatskih stručnih i znanstvenih radova koji se bave izradom sustava za analizu sentimenta koji se temelje na povratnoj neuronskoj mreži⁴⁴.

U ovome smo se radu koristili knjižnicom TensorFlow koju razvija Google. Nju smo odabrali, jer je vrlo popularna, ima vrlo veliku potporu razvojnih programera te se često ažurira. Uz TensorFlow, te kao njegova nadogradnja, dostupna je i Python knjižnica Keras⁴⁵. Budući da se temelji na TensorFlow-u, knjižnica je vrlo ažurna, intuitivna i omogućuje izradu sustava s dubokim učenjem u svega nekoliko koraka. Keras knjižnica nudi sve alate nužne za normaliziranje podataka prije izrade sustava. Također, sama izrada sustava, točnije definiranje arhitekture nekoga sustava za duboko učenje, vrlo je intuitivna. U Kerasu se arhitektura određuje definiranjem slojeva sustava kojima se može odrediti veličina, njegov način pristupa neuronima, razne funkcije koje sloj primjenjuje i drugo. Za izradu alata za određivanje sentimenta koristili smo isključivo korpus za učenje koji je opisan u prethodnom poglavlju (vidi poglavlje 5.).

Za izradu alata bilo je nužno očistiti postojeće podatke. Iz *tvitova* su uklonjene poveznice, oznake korisnika („@korisnik“), sva su velika slova zamijenjena malim slovima i redosljed se *tvitova* još jednom nasumično izmiješao. Uklonjeni su *titovi* koji su sadržavali samo poveznice ili oznake drugih korisnika. Dobiven korpus i dalje sadrži četiri kategorije („pozitivno“, „neutralno“, „negativno“, „ne može se odrediti“), no krajnji je cilj alata da odredi sentiment nekoga *tvita*, stoga, kako bismo olakšali proces klasifikacije, sve smo *titove* označene kategorijom „ne može se odrediti“ uklonili iz korpusa. Ovako dobiven korpus sastoji se od 2941 *tit*.

Izrađen je i sustav koji se učio na korpusu s četirima kategorijama, međutim, njegova je točnost u usporedbi s alatom izrađenim na korpusu s trojnom podjelom bila niža (vidi tablicu 6.). Dodatno,

⁴⁴ Usp. Krajcar (2014), Šikač (2020), Raguzin (2018), Jakopović i Mikelić Preradović (2016).

⁴⁵ <https://keras.io/>

slični radovi nemaju četiri kategorije, odnosno autori niti spominju da je anotatorima pri izradi jezičnoga resursa omogućen odabir četvrte kategorije „ne može se odrediti“, niti spominju podatke o jezičnim segmentima koji nisu bili označeni, a koji zapravo implicitno čine kategoriju „ne može se odrediti“. Uvidom u slične radove, možemo zaključiti, da je standard struke trojna podjela sentimenta („pozitivno“, „neutralno“, „negativno“) te je dodatna kategorija neodlučnosti, po pravilu, prešućena⁴⁶.

Korpus za izradu alata razdvojen je u dvije cjeline, u tzv. korpus za učenje i korpus za provjeru. Korpus za provjeru iznosi 10% cjelokupnog korpusa i on služi za provjeru ili testiranje sustava. Oba su potkorpusa tokenizirana, te se potom na oba potkorpusa primijenila metoda *padding* koja podrazumijeva da se ravnajući prema najdužem unosnom podatku izjednače duljine svih unosa. Jednostavnije, ako je vektorski unos [1 3 14 356 1] sa svojom duljinom od 5 vrijednosti najduži unosni podatak, onda će vektor [34 86 6] nakon što se primijeni *padding* funkcija postati vektor [34 86 6 0 0]. Produljivanje se vrši dodavanjem ničega (tj. 0) na kraj vektora. Ovaj je korak nužan jer se Keras može trenirati tek ako su svi podaci ujednačeni.

Sentiment je svakoga tvita također konvertiran u jedinstvene vektore sačinjene od 0 i 1.

	negativno	neutralno	pozitivno
337	0	1	0
2128	1	0	0
917	0	0	1
2522	0	1	0
847	1	0	0
...
2660	0	1	0
1965	0	0	1
1395	0	0	1
1578	0	0	1
2428	0	1	0

Slika 4: Konvertiran sentiment

Negativni sentiment u našem slučaju predstavlja vektor [1 0 0], neutralni [0 1 0], a pozitivni [0 0 1].

⁴⁶ Usp. Lu et al (2017), Liu (2020), Kandharo et al. (2019).

Potom smo izradili nekolicinu sustava kako bismo ustanovili najuspješniju arhitekturu. Za upute o definiranju arhitekture sustava koristili smo Kerasovu dokumentaciju⁴⁷, te smo u skladu sa sugestijama iz dokumentacije definirali tri arhitekture. Pri izradi sustava koristili smo *Embedding*, *LSTM*, *Dense*, *Dropout*, *Flatten* i *GlobalMaxPooling1D* slojeve. *Embedding* sloj konvertira unose u jedinstvene vektore i možemo reći da priprema podatke, u ovome slučaju tekst *tvitova*, za učenje sustava. *LSTM* sloj je ključni sloj povratne neuronske mreže zahvaljujući kojemu sustav nastoji zapamtiti pravilnosti i obrasce u dužim sekvencama podataka⁴⁸. *Dense* sloj je najobičniji sloj neuronske mreže koji je čvrsto povezan sa slojem koji mu prethodi. Neuroni ovoga sloja povezani su sa svim neuronima prethodnoga. Svi neuroni *Dense* sloja daju jednu izlaznu vrijednost što znači da će sloj *Dense(10)* primiti sve izlazne vrijednosti prijašnjega i dati 10 izlaznih vrijednosti sljedećem sloju. *Dropout* sloj je zapravo funkcija kojom se nasumično izbacuje određeni broj neurona. Ovim se slojem sustav potiče na generalizaciju, sprječavajući mu da tek nauči skup za treniranje, tj. on sprječava pojavu kada sustav prestane biti primjenjiv na nove, neviđene podatke. Ova se pojava naziva prenaučenosť (engl. *overfitting*) i ona se uvijek nastoji izbjeći pri izradi nekoga alata pomoću neuronskih mreža. *GlobalMaxPooling1D* je sloj koji uzima maksimalne vrijednosti prijašnjega sloja i time smanjuje broj unosa za sljedeći sloj.

⁴⁷ <https://keras.io/api/>

⁴⁸ Za više o *LSTM* slojevima vidi poglavlje 3.1.3.

Prvo smo izradili sustav bez *LSTM* sloja. Ovaj se sustav sastoji od *Embedding* sloja, tri *Dense* sloja i *Dropout* sloja. Ovim smo sustavom pogledali kakvi se rezultati postižu bez primjene *LSTM* sloja, točnije izradili smo tradicionalnu unaprijednu neuronsku mrežu kako bismo ju mogli usporediti s povratnom neuronskom mrežom.

```
Model: "sequential_5"
```

Layer (type)	Output Shape	Param #
embedding_4 (Embedding)	(None, 100, 300)	1500000
dense_9 (Dense)	(None, 100, 80)	24080
dense_10 (Dense)	(None, 100, 80)	6480
dropout_4 (Dropout)	(None, 100, 80)	0
flatten (Flatten)	(None, 8000)	0
dense_11 (Dense)	(None, 3)	24003

```

=====
Total params: 1,554,563
Trainable params: 1,554,563
Non-trainable params: 0
=====

```

Slika 5: Arhitektura unaprijedne neuronske mreže

Nakon ovoga sustava izradili smo sustav koji se sastojao od *Embedding* sloja, dva *LSTM* sloja, jednoga *Dropout* i jednoga *Dense* sloja. Ovako je dobivena klasična rekurentna neuronska mreža.

```
Model: "sequential_6"
```

Layer (type)	Output Shape	Param #
embedding_5 (Embedding)	(None, 100, 300)	1500000
lstm_2 (LSTM)	(None, 100, 40)	54560
lstm_3 (LSTM)	(None, 100, 40)	12960
dropout_5 (Dropout)	(None, 100, 40)	0
flatten_1 (Flatten)	(None, 4000)	0
dense_12 (Dense)	(None, 3)	12003

```

=====
Total params: 1,579,523
Trainable params: 1,579,523
Non-trainable params: 0
=====

```

Slika 6: Arhitektura rekurentne neuronske mreže s *LSTM* slojevima

Naposljetku kombinirali smo *LSTM* sustav s *GlobalMaxPooling1D* koji se najčešće koristi pri izradi konvolucijskih neuronskih mreža koje se koriste za prepoznavanje slika i drugih vizualnih materijala. Nova se arhitektura sastojala od *Embedding* sloja, dva *LSTM* sloja, dva *Dense* sloja, jednoga *Dropout* sloja i jednoga *GlobalMaxPooling1D* sloja. Najuspješnijom se arhitekturom pokazala zadnja koja kombinira metode rekurentnih i konvolucijskih neuronskih mreža. Arhitekturu ovoga sustava možemo vidjeti na slici 7.

```

Model: "sequential_42"
-----
Layer (type)                Output Shape                Param #
-----
embedding_42 (Embedding)    (None, 100, 300)          1500000
lstm_66 (LSTM)              (None, 100, 60)           86640
lstm_67 (LSTM)              (None, 100, 60)           29040
dense_54 (Dense)            (None, 100, 40)            2440
dropout_39 (Dropout)        (None, 100, 40)            0
global_max_pooling1d_29 (Gl (None, 40)                  0
obalMaxPooling1D)
dense_55 (Dense)            (None, 3)                   123
-----
Total params: 1,618,243
Trainable params: 1,618,243
Non-trainable params: 0

```

Slika 7: Arhitektura kombinirane povratne neuronske mreže

U tablici 6.⁴⁹ donosimo rezultate navedenih sustava. Također prikazujemo rezultate kombiniranoga sustava treniranoga na korpusu s četirima kategorijama, tj. s kategorijom „ne može se odrediti“.

	Unaprijedna neuronska mreža	Rekurentna neuronska mreža (LSTM)	Kombinirana LSTM mreža	Kombinirana LSTM mreža (4 kategorije)
točnost	0,5797	0,6271	0,6441	0,5016
F1-mjera	0,5844	0,6291	0,6422	0,4160
funkcija gubitka	1,0583	1,9168	0,9633	1,1992

Tablica 6: Usporedba različitih arhitektura

⁴⁹ Detaljnije o prikazanim mjerama vidi u poglavlju 6.2.

Temelj ove mreže, kao što smo već rekli, čine dva *LSTM* sloja pomoću kojih sustav promatra pravilnosti na razini većih segmenata. Također valja obratiti pozornost na zadnji *Dense* sloj čiji je unos tri neurona, a daje jednu vrijednost. Tri neurona, predstavljaju tri sentimenta, a jedan izlazni sentiment je odluka koju donosi sustav. Ovakvo ponašanje omogućuje *softmax* funkcija koja pretvara ulazne vrijednosti u vjerojatnosti. Na slici 7. također možemo vidjeti broj promjenjivih brojeva ili težina (*params*) koji se nalaze u sustavu. Što više slojeva ima jedan sustav, time se više neurona i težina nalazi u njemu. Treba imati na umu da broj slojeva i veličina sustava nisu izravno povezana s njegovom točnošću.

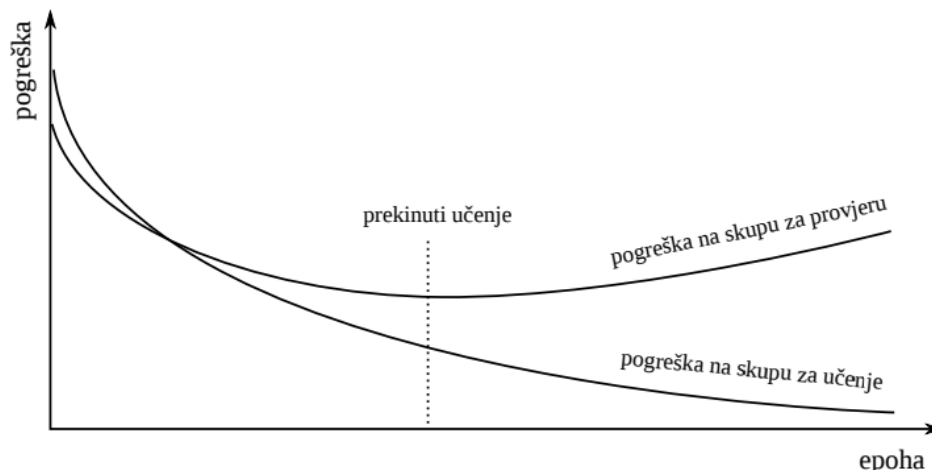
Pri izradi alata za analizu sentimenta pripazili smo i na neravnotežu među trima kategorijama sentimenta. Kako bi se poboljšao sustav težine smo kategorija (omjer) uvažili pri izradi sustava. Pri učenju sustava koristili smo skup podataka od 64 instanci (engl. *batch size*). Ova vrijednost predstavlja broj istovremeno unesenih instanci u sustav. Ovu smo veličinu skupa odabrali s jedne strane zbog hardverskih ograničenja, jer povećanjem skupa podataka za učenje, minimalni hardverski zahtjevi eksponencijalno rastu, a s druge strane, pokazalo se da mali skupovi podataka za učenje pomažu u generalizaciji i daju u većini slučajeva dobre rezultate (Smith et al 2017)⁵⁰. Broj epoha (engl. *epoch*) je 8. Epohe predstavljaju broj iteracija tijekom kojih će se sustav trenirati. Svaka je epoha nadogradnja na prijašnju verziju alata, međutim, ne mora značiti da je uvijek zadnja epoha najuspješnija. Ovaj smo broj epoha odabrali isprva nasumično, međutim prilikom izgradnje sustava uočili smo da se pri učenju sustava već nakon četvrte epohe javlja prenaučenosć, stoga smo implementirali *callback* funkciju koja zaustavlja razvoj sustava ako se njezina točnost pogoršava. Broj epoha nismo promijenili kako bismo mogli prikazati prenaučenosć na *djelu*. Na slici 8. možemo vidjeti proces razvijanja sustava i rezultate za svaku pojedinu epohu do prekidanja razvoja zbog funkcije *callback*.

```
Epoch 1/8
42/42 [=====] - 27s 506ms/step - loss: 1.1133 - accuracy: 0.4539 - f1_score: 0.3656 - val_loss: 0.9360 - val_accuracy: 0.5932 - val_f1_score: 0.3886
Epoch 2/8
42/42 [=====] - 21s 490ms/step - loss: 0.7652 - accuracy: 0.6330 - f1_score: 0.5852 - val_loss: 0.8043 - val_accuracy: 0.6542 - val_f1_score: 0.6053
Epoch 3/8
42/42 [=====] - 21s 489ms/step - loss: 0.5332 - accuracy: 0.7600 - f1_score: 0.7551 - val_loss: 0.7798 - val_accuracy: 0.6542 - val_f1_score: 0.6328
Epoch 4/8
42/42 [=====] - 21s 489ms/step - loss: 0.3390 - accuracy: 0.8617 - f1_score: 0.8603 - val_loss: 0.8785 - val_accuracy: 0.5966 - val_f1_score: 0.6061
Epoch 5/8
42/42 [=====] - 21s 491ms/step - loss: 0.2148 - accuracy: 0.9067 - f1_score: 0.9087 - val_loss: 0.9633 - val_accuracy: 0.6441 - val_f1_score: 0.6422
```

Slika 8: Epohe u razvoju sustava

⁵⁰ Valja napomenuti da o idealnoj veličini skupa podataka za učenje ne postoji konsenzus (usp. Smith 2018). Ono oko čega se većina istraživača slaže je da ne postoji univerzalno pravilo koje se može primijeniti za sve sustave.

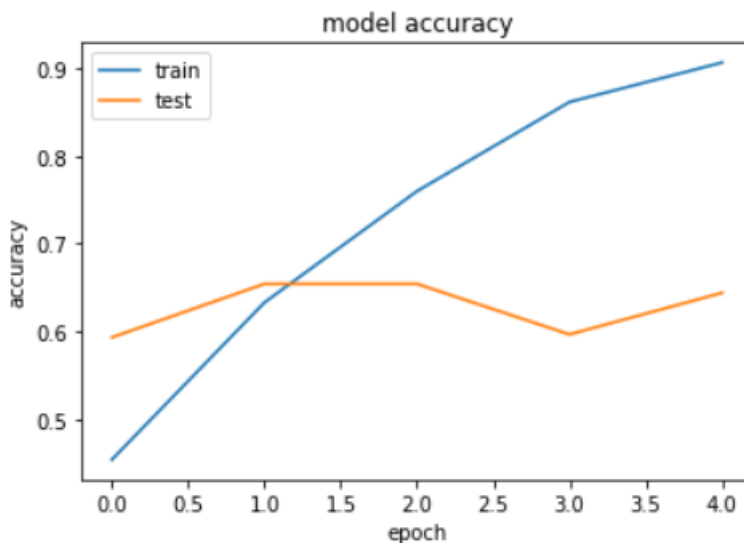
Na slici 9. se prikazuje idealni trenutak za prekid razvoja sustava, odnosno trenutak u kojoj se *callback* funkcija aktivira.



Slika 9: Pogreške pri učenju (Čupić 2016: 10)

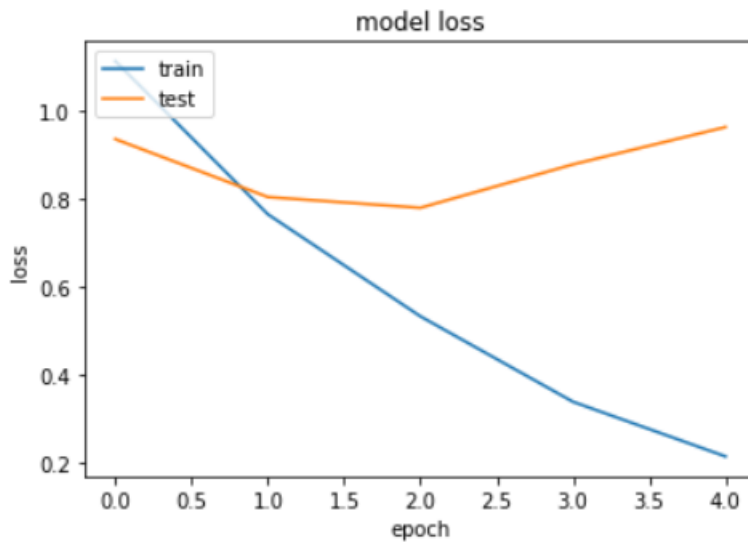
6.2. Rezultati

Izrađeni sustav mjerili smo i vrednovali na nekoliko načina. Sustav je postigao točnost (engl. *accuracy*) od 0,6441 na skupu za vrednovanje (testiranje). Na grafu 3. možemo vidjeti točnost u epohama razvoja.



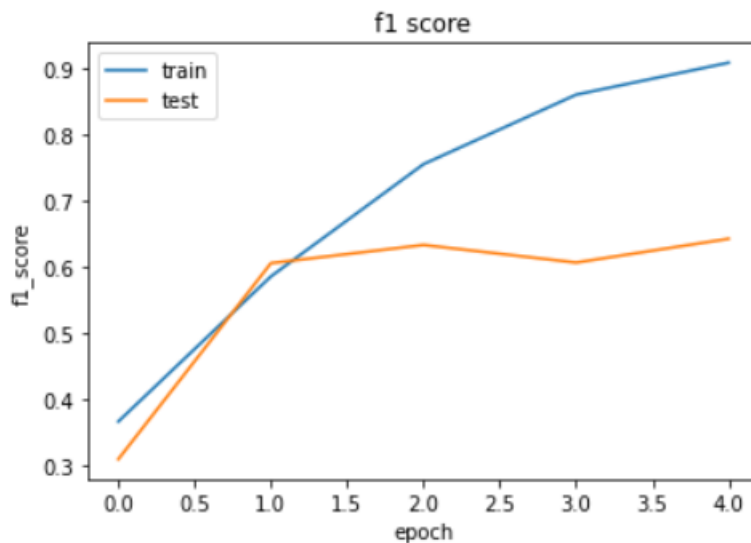
Graf 3: Točnost sustava u epohama razvoja

Na grafu 4. možemo vidjeti funkciju gubitka u epohama razvoja. Ovom se mjerom promatra koliko su grube pogreške sustava. Što je funkcija gubitka manja, to je sustav točniji.



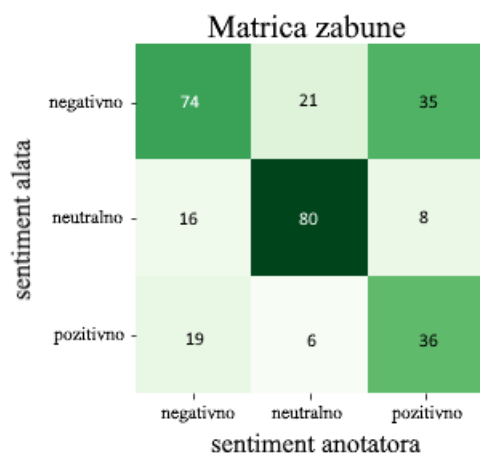
Graf 4: Funkcija gubitka sustava u epohama razvoja

Na grafu 5. možemo vidjeti F1-mjeru sustava u epohama razvoja.



Graf 5: F1-mjera sustava u epohama razvoja

Matrica je zabune (engl. *confusion matrix*) još jedan način s pomoću kojega možemo organizirati podatke, pri čemu se uzimaju u obzir ne samo dobre odluke sustava, već i njegove pogreške. U našem je slučaju riječ o matrici dimenzija 3x3. Matrica se zabune promatra na testnim podacima, odnosno na onim primjerima koje sustav *prvi put vidi*. Na grafu 6. možemo vidjeti matricu zabune sustava.

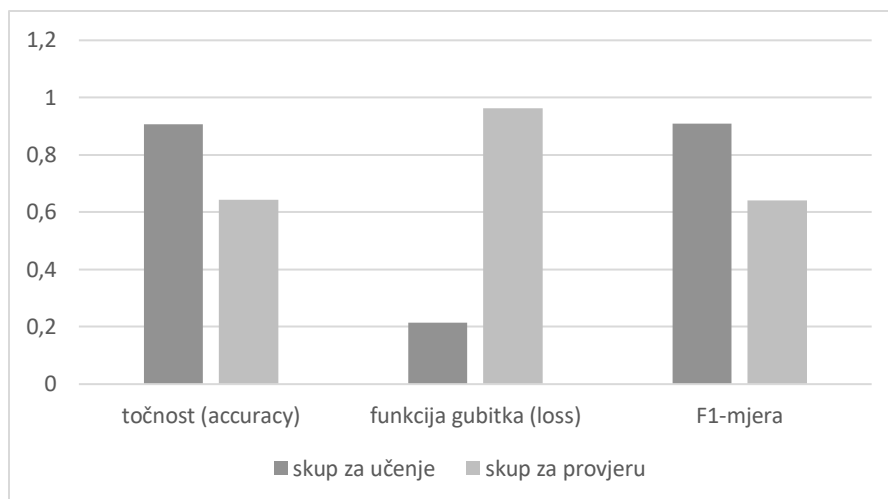


Graf 6: Matrica zabune

Na grafovima koji prikazuju točnost, funkciju gubitka i F1-mjeru vidimo podatke za dva različita skupa. U tablici 7. i na grafu 7. donosimo podatke o mjerama vrednovanja skupova za učenje i provjeru.

	točnost (accuracy)	funkcija gubitka (loss)	F1-mjera
skup za učenje	0,9067	0,2148	0,9087
skup za provjeru	0,6441	0,9633	0,6422

Tablica 7: Usporedba mjera vrednovanja skupovima za učenje i provjeru



Graf 7: Usporedba mjera vrednovanja u skupovima za učenje i provjeru

6.3. Interpretacija rezultata

Promotrimo li rezultate sustava vidimo velike razlike među mjerama vrednovanja između skupa za učenje i vrednovanje. Svaka mjera vrednovanja daje bolje rezultate na skupu za učenje. Svakom epohom sustav sve više pamti skup za učenje te se mjere vrednovanja koje se odnose na skup za učenje stalno poboljšavaju (vidi graf 3, 4, 5). Slično se dogodilo i pri razvoju našega sustava koji je u drugoj epohi imao manju funkciju gubitka, nego na kraju razvoja (vidi graf 3). Na grafovima se također vidi da je zbog *callback* funkcije razvoj sustava prekinut nakon četvrte epohe, jer su se pogoršavali njegovi rezultati postignuti na skupu za provjeru. Za interpretaciju i evaluaciju rezultata se stoga moramo voditi rezultatima postignutim na skupu za provjeru.

Alat je postigao rezultate od 0,644 točnosti i 0,642 F1-mjere koji se možda čine niskim, međutim ove rezultate moramo promatrati u kontekstu drugih istraživanja, ali i u kontekstu određivanja sentimenta općenito. U ovome smo radu razvili alat na temelju jezičnoga resursa čija je izrada također opisana, praćena, evaluirana. Ove nam okolnosti omogućuju usporedbu rezultata alata i anotacije. Podrazumijeva se da ćemo usporediti točnost alata s postotkom međuanotatorske podudarnosti. Na ovaj način alatovu točnost u prepoznavanju sentimenta *tvita* možemo usporediti s podudarnosti između dva čovjeka. Prosječna vrijednost međuanotatorske pouzdanosti bila je 0,605, a maksimalna vrijednost 0,684 (vidi tablicu 1.). Točnost ovoga alata iznosi 0,644. To znači da bi se anotatori s alatom u većoj mjeri složili, nego s nekim drugim anotatorom. Od prosječne međuanotatorske pouzdanosti alat je bio bolji skoro četiri postotna boda, dok je od najuspješnije podudarnosti zaostao samo nešto više od četiri postotna boda. Prema ovim rezultatima, za

anotaciju *tvitova* istraživačima je, uzmemo li u obzir i vremenske i materijalne resurse koje ručna anotacija zahtijeva, svakako bolje i brže anotirati *tvitove* pomoću izrađenoga alata.

Međutim, za potpunu sliku alata, trebamo ga promotriti i u kontekstu drugih sličnih istraživanja. Lu et al. (2017) učili su sustave za klasifikaciju sentimenta koji klasificiraju u tri kategorije. Za izradu su koristili korpus od 1803 tvita, a sustave su izradili raznim metodama dubokog učenja. Klasifikator temeljen na povratnoj neuronskoj mreži ima F1-mjeru od svega 0,417 (sustav izrađen metodom potpornih vektora ima F1-mjeru od 0,588). Liu (2020) prikazuje komparaciju sustava za analizu sentimenta učenih na 350 000 recenzija. Klasifikator izgrađen s pomoću *LSTM* slojeva ima F1-mjeru od 0,701 (sustav izrađen metodom potpornih vektora ima F1-mjeru od 0,757). Heikal, Torki i El-Makky izradili su sustave za analizu sentimenta arapskih *tvitova*. Korpus za učenje im je veličine 10 000 *tvitova*. Klasifikator temeljen na *LSTM* mreži ima F1-mjeru od 0,62. Kandhro et al. (2019) analizirali su studentske recenzije o nastavnicima na fakultetu i na temelju korpusa od 3099 recenzija izradili sustav za automatsku klasifikaciju sentimenta. Sustav izgrađen s pomoću *LSTM* sloja ima F1-mjeru od 0,654. Možemo zaključiti da su rezultati u ovome radu razvijenog sustava u skladu s rezultatima drugih istraživanja.

Pri interpretaciji alata razvijenih dubokim učenjem, moramo veliku važnost dati vrsti podataka na kojima se sustav učio. Kao što se i na primjeru drugih radova vidi, *tweet* kao tekstna vrsta predstavlja problem sustavima za duboko učenje. O problemima određivanja sentimenta *tvitova* svjedoči i činjenica da se ni ljudski anotatori u velikom dijelu primjera nisu mogli složiti (vidi poglavlje 5.). Možemo pretpostaviti da je (ne)uspješnost sustava s jedne strane uzrokovala vrsta podataka na kojima se on učio, a s druge strane manjak podataka. Kako bismo povećali točnost sustava trebali bismo spriječiti prenaučenosť koja se javljala već nakon četvrte epohe. Postoji nekoliko načina kako se sustav koji je u stanju prenaučenosť može popraviti. Kao prvo, može se povećati *Dropout* sloj sustava, međutim, tu smo opciju isprobali i ona nije donijela željene rezultate. Druga je opcija da se poveća količina podataka za učenje sustava, međutim to u sklopu ovoga rada nije izvedivo.

U ovome smo poglavlju već interpretirali mjeru točnosti, funkciju gubitka i F1-mjeru, ali nismo govorili o matrici zabune koja može dati zanimljiv uvid u funkcioniranje sustava. Matrica zabune (vidi graf 6) prikazuje odnos između sentimenta u korpusu i sentimenta koje je odredio alat za određivanje sentimenta. Zanimljivo je uočiti da je alat izrazito uspješan kada radi pretpostavke o *tvitovima* neutralnoga sentimenta. U ovoj kategoriji možemo govoriti o točnosti od 76%. O

uzrocima je krivih klasifikacija nekoga alata za analizu sentimenta uvijek teško sa sigurnošću govoriti, pa ćemo umjesto toga u nastavku prikazati nekoliko *tvitova* kojima je alat krivo odredio sentiment.

- *Tvit* „prvi doživljeni potres u autu #prvisepamte 😞“ alat je označio negativnim. Točnije, predikcija alata za ovaj *twit* bila je „[0.7879019 0.1201734 0.09192467]“. Ovaj nam vektor pokazuje da je alat 78% siguran da je navedeni *twit* negativan, a mogućnost da je pozitivan je tek 9%. Ljudski je anotator prepoznao da se autor *twita* smije svojoj (ne)sreći, a to naglašava ključnom riječju „#prvisepamte“ i emotikonom „plačem od smijeha“, te odredio da je navedeni *twit* pozitivan.
- *Tvitu* „Mrzim kad ga osjetim #potres“ alat je dodijelio vektor „[0.31780416 0.34596387 0.33623198]“, što bi trebalo značiti da je *twit* neutralan, međutim ovaj put alat nije bio toliko odlučan. Ljudski je anotator ovaj *twit* označio negativnim sentimentom.
- „Smještaj za ljude iz #Petrinja dijelite molim!“ je *twit* kojemu je dodijeljen vektor „[0.88665766 0.01254958 0.10079273]“, odnosno negativan sentiment. Ljudski su anotatori prepoznali da u ovome *twitu* nisu dominantni negativni osjećaji (dogodio se potres), već osjećaj nade (da će ljudi dati smještaj) i izražavanje molbe (da se *twit* podijeli) i pripisali *twitu* pozitivni sentiment.
- *Tvitu* „Kad te trese potres iz druge županije.“ alat je odredio vektor „[0.25252122 0.1266648 0.6208139]“, tj. pozitivni sentiment. Ljudski su anotatori za ovaj *twit* rekli da je negativan. Prema alatu je vjerojatnost toga da je ovaj *twit* negativan 25%.
- *Tvitu* „zagreb je imao potres beograd ima poplavu za sad je 1 1“ alat je dodijelio vektor „[0.05982307 0.8457241 0.09445281]“, odnosno neutralni sentiment. Ljudskim je anotatorima jasno da je riječ o vicu te su *twit* označili pozitivnim sentimentom.
- Prema porcijenama alata *twit* „pooo- potres na usnama, nek nocas trese se sve do jadrana. pooo- potres na usnama, nek nocas znaju svi da si me zatresla“ negativnog je sentimenta, jer je pripadajući vektor „[0.635471 0.00680545 0.35772356]“. Ljudski su anotatori *twit* odredili pozitivnim jer je očito riječ o samoironičnom prepjevu hita Miše Kovača, ali to alat, očekivano, nije prepoznao.

Ovi primjeri prije svega pomažu u kontekstualizaciji gore analiziranih mjera vrednovanja i prikazuju na kojim je mjestima pogriješio alat analizu sentimenta. Mogli bismo reći da razvijen

alat za analizu sentimenta ne prepoznaje šalu. *Tvitove* koje su ljudski anotatori svrstali u kategoriju pozitivnih *tvitova*, a često se temelje na ismijavanju nekoga pojedinca ili skupine ljudi, alat je često svrstao među negativne ili neutralne *tvitove*. Na primjerima vidimo da alat doista analizira, a o tome nam govore vektori koji uvijek prikazuju tek vjerojatnost neke kategorije. Dobiveni rezultati u skladu s očekivanjima, štoviše dobri, uzevši u obzir specifičnosti vezane za analizirani jezični materijal.

7. Zaključak

U ovome smo radu dali pregled teorijskih, metodoloških i praktičnih problema koji se tiču izrade *Twitter* korpusa i razvoja alata za analizu sentimenta. Predstavljeni korpus može biti od koristi ne samo za kvantitativnu analizu koju smo predstavili u ovome radu, a koja podrazumijeva metode kojima se koristi obrada prirodnoga jezika, već i za kvalitativnu analizu anotatora, njihove percepcije humora i sarkazma ili možda za analizu shvaćanja pojedinih emotikona. U ovome smo radu, razmišljajući o mogućim uzrocima odluka anotatora tek zagreballi površinu ovakve analize. Iscrpan rad takve tematike vjerojatno čeka nekoga sociologa ili psihologa. Izrađen korpus, uz analizu sentimenta, može poslužiti i za analizu drugih jezičnih razina i drugih jezičnih pojava. Valjalo bi analizirati podatke o čestotnosti riječi, najčešćim sintagmama, najčešćim pridjevima uz vlastita imena i slično, čime se u ovome radu, nismo mogli baviti. Uz sav potencijal što ovakav jezični resurs predstavlja, mora se još jednom istaknuti, da je njegova izrada zahtijevala angažman dvadesetak ljudi. Izrada je korpusa financijski i vremenski zahtjevna radnja, ali je rezultat dragocjen skup podataka nužan za razvoj drugih jezičnih resursa i jezičnih alata koji se u konačnici koriste ne samo u akademske svrhe, već su od pomoći i pri izvršavanju svakodnevnih radnji, poput pretraživanja ili prevođenja.

Izradom smo alata za analizu sentimenta s jedne strane pokazali da su povratne neuronske mreže relativno uspješne u izradi alata ovakvoga tipa, a s druge strane da se izradom alata temeljenih na manjem korpusu mogu postići dobri rezultati, što nam daje nadu smjestimo li ovaj rad u kontekst obrade jezika s malom podrškom, što hrvatski svakako jest. Ovim se radom dokazalo da je bavljenje s hrvatskim jezikom u internetskom okruženju moguće, dapače, može dati i dobre rezultate. Međutim, valja bit kritičan prema rezultatima. Točnost alata koje je 0,644 ne možemo smatrati izvrsnim. Uzrok ovakvoga rezultata se djelomično krije u prirodi građe koju su i anotatori i alat za analizu sentimenta obrađivali. Ukazalo se da su *tvitovi* vrlo nenormalizirani, ali ono što izgleda da je još presudnije pri analizi, *tvitovi* su kratki, stoga je određivanje njihovoga sentimenta, kako ljudima tako i alatu težak zadatak. Iako je točnost alata u odnosu na prosječnu međuanotatorsku pouzdanost bolja, ona bi se mogla unaprijediti. Rezultati bi se alata mogli poboljšati da se on trenira pomoću većega korpusa, u idealnome slučaju, specijaliziranome. Alat bi se također mogao poboljšati dodavanjem sentimentnih rječnika pri njegovom treniranju. Iako

se alat i korpus mogu i trebaju nadograditi, znakovita je činjenica da bi se proširenje korpusa moglo djelomice obavljati i pomoću izrađenoga alata.

Literatura

- Ahel, R., 2003. Postupak klasifikacije teksta temeljen na k-NN metodi i naivnom Bayesovom klasifikatoru. *diplomski rad*.
- Artstein, R. i Poesio, M., 2008. Inter-coder agreement for computational linguistics. *Computational linguistics*, 34(4). 555–596.
- Bobicev, V. i Sokolova, M., 2017. Inter-annotator agreement in sentiment analysis: Machine learning perspective. *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*. 97–102.
- Cambria, E. i White, B., 2014. Jumping NLP curves: A review of natural language processing research. *IEEE Computational intelligence magazine*, 9(2), 48–57.
- Cetinić, E., 2013. Primjena različitih metoda strojnog učenja u problemu klasifikacije slikarskih djela prema autoru. *Sveučilište u Zagrebu – fakultet elektrotehnike i računarstva, Zagreb*.
- Čupić, M. 2016. Umjetna inteligencija: Umjetne neuronske mreže. <http://java.zemris.fer.hr/nastava/ui/ann/ann-20180604.pdf> (25.4.2022.).
- Elbagir, S. i Yang, J., 2019. Twitter sentiment analysis using natural language toolkit and VADER sentiment. *Proceedings of the International MultiConference of Engineers and Computer Scientists*. 122. 16.
- Friese, S. 2020. Measuring Inter-coder Agreement – Why Cohen’s Kappa is not a good choice. <https://atlasti.com/2020/07/12/measuring-inter-coder-agreement/> (07.01.2022).
- Gaspar, R., Pedro, C., Panagiotopoulos, P. i Seibt, B., 2016. Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Computers in Human Behavior*. 56. 179–191.
- Géron, A., 2019. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc.
- Gräbner, D., Zanker, M., Fliedl, G. i Fuchs, M., 2012. Classification of customer reviews based on sentiment analysis. *ENTER*. 460–470.

- Han, H., Zhang, J., Yang, J., Shen, Y. i Zhang, Y., 2018. Generate domain-specific sentiment lexicon for review sentiment analysis. *Multimedia Tools and Applications*, 77(16). 21265–21280.
- Haroutunian L. 2016. What is the minimum training dataset size to train a sentiment analysis classifier (social media corpus). *Quora*. <https://www.quora.com/What-is-the-minimum-training-dataset-size-to-train-a-sentiment-analysis-classifier-social-media-corpus>. (9.3.2022.).
- Hasan, A., Moin, S., Karim, A. i Shamshirband, S., 2018. Machine learning-based sentiment analysis for twitter accounts. *Mathematical and Computational Applications*. 23(1). 11.
- Heikal, M., Torki, M. i El-Makky, N., 2018. Sentiment analysis of Arabic tweets using deep learning. *Procedia Computer Science*. 142. 114–122.
- Hogenboom, A., Bal, D., Frasinca, F., Bal, M., de Jong, F. i Kaymak, U., 2013. Exploiting emoticons in sentiment analysis. *Proceedings of the 28th annual ACM symposium on applied computing*. 703–710.
- Hsu, L.M. i Field, R., 2003. Interrater agreement measures: Comments on Kappan, Cohen's Kappa, Scott's π , and Aickin's α . *Understanding Statistics*, 2(3). 205–219.
- Ivaninić, A. 2021. Analiza sentimenta objava na Twitteru vezanih za koronavirus, *diplomski rad*, <https://urn.nsk.hr/urn:nbn:hr:195:089255> (07.01.2022.).
- Jakopović, H. i Mikelić Preradović, N., 2016. Identifikacija online imidža organizacija temeljem analize sentimenta korisnički generiranog sadržaja na hrvatskim portalima. *Medijska istraživanja: znanstveno-stručni časopis za novinarstvo i medije*, 22(2). 63–82.
- Jakubiček, M., Kilgarriff, A., Kovář, V., Rychlý, P. i Suchomel, V., 2013. The TenTen corpus family. *7th International Corpus Linguistics Conference CL*. 125–127.
- Java, A., Song, X., Finin, T. i Tseng, B., 2007. Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. 56–65.
- Jordan, M.I. i Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science*. 349(6245). 255–260.

- Kandhro, I.A., Wasi, S., Kumar, K., Rind, M. i Ameen, M., 2019. Sentiment analysis of students' comment using long-short term model. *Indian Journal of Science and Technology*. 12(8).1–16.
- Kim, S.M. i Hovy, E., 2004. Determining the sentiment of opinions. *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*. 1367–1373.
- Kouloumpis, E., Wilson, T. i Moore, J., 2011. Twitter sentiment analysis: The good the bad and the omg!. *Proceedings of the international AAAI conference on web and social media*. 5(1). 538–541.
- Krajcar, L. 2014. Analiza sentimenta u tvitovima na hrvatskome jeziku. *diplomski rad*. <https://urn.nsk.hr/urn:nbn:hr:168:908184> (07.01.2022).
- Labille, K., Gauch, S. and Alfarhood, S., 2017. Creating domain-specific sentiment lexicons via text mining. *Proc. Workshop Issues Sentiment Discovery Opinion Mining (WISDOM)*. 1–8.
- Landis, J.R. i Koch, G.G., 1977. The measurement of observer agreement for categorical data. *biometrics*. 159–174.
- LeCun, Y., Bengio, Y. i Hinton, G., 2015. Deep learning. *nature*, 521(7553), 436–444.
- Liu, B., 2011. *Web data mining: exploring hyperlinks, contents, and usage data* (Vol. 1). Berlin: springer.
- Liu, B., 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1). 1–167.
- Liu, B., 2020. *Sentiment analysis: mining opinions, sentiments, and emotions* (Second edition). Cambridge University Press.
- Liu, S., 2020. Sentiment analysis of yelp reviews: a comparison of techniques and models. *arXiv preprint arXiv:2004.13851*.
- Lu, Y., Sakamoto, K., Shibuki, H. i Mori, T., 2017. Are deep learning methods better for twitter sentiment analysis. *Proceedings of the 23rd annual meeting of natural language processing (Japan)*. 787–790.
- Lui, M. i Baldwin, T., 2012. langid. py: An off-the-shelf language identification tool. *Proceedings of the ACL 2012 system demonstrations*. 25–30.

- Ljubešić, N. i Erjavec, T., 2011. hrWaC and slWaC: Compiling web corpora for Croatian and Slovene. *International Conference on Text, Speech and Dialogue*. 395–402. Springer: Berlin.
- Maynard, D.G. i Greenwood, M.A., 2014. Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. *Lrec 2014 proceedings*.
- Mozetič, Igor; Grčar, Miha i Smailović, Jasmina, 2016. Twitter sentiment for 15 European languages. *Slovenian language resource repository CLARIN.SI*. <http://hdl.handle.net/11356/1054> (25.4.2022).
- Neppalli, V.K., Caragea, C. i Caragea, D., 2018. Deep neural networks versus naive bayes classifiers for identifying informative tweets during disasters. *Proceedings of the 15th Annual Conference for Information Systems for Crisis Response and Management (ISCRAM)*.
- Pak, A. i Paroubek, P., 2010. Twitter as a corpus for sentiment analysis and opinion mining. *LREC*. 10(2010). 1320–1326.
- Pelicon, Andraž; Pranjić, Marko; Miljković, Dragana; Škrlić, Blaž i Pollak, Senja, 2020. Sentiment Annotated Dataset of Croatian News. *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1342> (25.4.2022).
- Raguzin, A., 2018. Analiza sentimenta u tekstovima i mikroblogovima o izbjegličkoj krizi. *doktorska disertacija*.
- Rehm, G., Marheinecke, K., Hegele, S., Piperidis, S., Bontcheva, K., Hajič, J., Choukri, K., Vasiljevs, A., Backfried, G., Prinz, C. i Pérez, J.M.G., 2020. The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe. *arXiv preprint arXiv:2003.13833*.
- Rish, I., 2001, August. An empirical study of the naive Bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 3(22), 41–46.
- Sarker, I.H., 2021. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6). 1–20.
- Sarlan, A., Nadam, C. i Basri, S., 2014, November. Twitter sentiment analysis. *Proceedings of the 6th International conference on Information Technology and Multimedia*. 212–216.

- Sim, J. i Wright, C.C., 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy*. 85(3). 257–268.
- Smith, L.N., 2018. A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*. (07.01.2022.).
- Smith, S.L., Kindermans, P.J., Ying, C. i Le, Q.V., 2017. Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*. (07.01.2022.).
- Šikač, P. N. 2020. Računalna analiza sentimenta. *završni rad*. <https://urn.nsk.hr/urn:nbn:hr:211:325795> (07.01.2022.).
- Šnajder, J. 2020. 8. *Stroj potpornih vektora*, Strojno učenje. Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, [https://www.fer.unizg.hr/download/repository/SU-2020-08-StrojPotpornihVektora\[1\].pdf](https://www.fer.unizg.hr/download/repository/SU-2020-08-StrojPotpornihVektora[1].pdf) (09.03.2022.)
- Tadić, M., Brozović-Rončević, D. i Kapetanović, A., 2012. *Hrvatski jezik u digitalnom dobu*. Springer.
- Takala, P., Malo, P., Sinha, A. i Ahlgren, O., 2014. Gold-standard for Topic-specific Sentiment Analysis of Economic Texts. *LREC*. 2014. 2152–2157.
- Tripto, N.I. i Ali, M.E., 2018. Detecting multilabel sentiment and emotions from bangla youtube comments. *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. 1–6.
- Walther, J.B. i D'addario, K.P., 2001. The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review*, 19(3). 324–347.
- Zaki, J., 2020. Catastrophe compassion: Understanding and extending prosociality under crisis. *Trends in cognitive sciences*, 24(8). 587–589.

Popis tablica

Tablica 1: Međuanotatorska podudarnost u potkorpusima	24
Tablica 2: Podaci o anotaciji 7. potkorpusa.....	25
Tablica 3: Prosječni omjer kategorija svih anotatora, omjer kategorija u potkorpusu i njihova...26	
Tablica 4: Anotator s najvećim i najmanjim omjerom kategorije ne može se odrediti	27
Tablica 5: Razredi podudarnosti prema Landis i Koch (1977:165).....	27
Tablica 6: Usporedba različitih arhitektura	37
Tablica 7: Usporedba mjera vrednovanja skupovima za učenje i provjeru	41

Popis slika

Slika 1: Dvodimenzionalan potporni vektor	10
Slika 2: Trodimenzionalni potporni vektor	11
Slika 3: Povratna neuronska mreža i mreže duge kratkoročne memorije.....	12
Slika 4: Konvertiran sentiment	34
Slika 5: Arhitektura unaprijedne neuronske mreže.....	36
Slika 6: Arhitektura rekurentne neuronske mreže s LSTM slojevima.....	36
Slika 7: Arhitektura kombinirane povratne neuronske mreže	37
Slika 8: Epohe u razvoju sustava	38
Slika 9: Pogreške pri učenju (Čupić 2016: 10).....	39

Popis formula

Formula 1: Naivni Bayesov klasifikator	11
Formula 2: Cohenova kappa.....	23

Analiza sentimenta u hrvatskim tvitovima na temu potresa

Sažetak

Analiza sentimenta područje je koje je sve popularnije među istraživačima koji se bave obradom prirodnoga jezika. Ovo područje objedinjuje psihologiju, sociologiju, antropologiju, lingvistiku i računalnu tehnologiju te zahtijeva interdisciplinarni pristup. Zainteresiranost umjetnom inteligencijom i dubokim učenjem koje se može uočiti u svim granama računarstva nije zaobišla ni područje analize sentimenta. Međutim, alati izrađeni za analizu sentimenta u hrvatskome jeziku većinom su izrađeni koristeći tradicionalne i danas već zastarjele metode. U ovome se radu prikazuje postupak izrade alata za analizu sentimenta koji se temelji na suvremenijim metodama, na neuronskim mrežama. U prvome dijelu rada opisana su teorijska polazišta analize sentimenta i strojnoga učenja. Potom se opisuje izrada specijaliziranoga anotiranoga Twitter korpusa na temu potresa. Korpus je anotiralo dvoje autora te kao takav pruža dobar uvid u problematiku same jezične građe. Najčešće poteškoće koje se javljaju pri izradi ovakvoga jezičnoga resursa opisuju se u kratkoj diskusiji. U zadnjem se dijelu rada opisuje izrada alata za analizu sentimenta temeljen na rekurentnim neuronskim mrežama i treniran na već spomenutom korpusu. Rezultati se alata kontekstualiziraju s jedne strane u odnosu na slične radove, a s druge strane u odnosu na međuanotatorsku pouzdanost koja se računala za Twitter korpusa.

Ključne riječi: analiza sentimenta, neuronske mreže, izrada korpusa, tvit, potres

Sentiment Analysis in Croatian Tweets About Earthquakes

Summary

Sentiment analysis is becoming more and more popular in the field of natural language processing. This specific field of study unites researchers with background in psychology, sociology, anthropology, linguistics, and computer sciences, and as such everyone interested in it has to have an interdisciplinary approach to the subject. The growing interest towards artificial intelligence and deep learning seen in every aspect of computer sciences did not bypass sentiment analysis either. However, tools developed to detect sentiment for Croatian language are mostly developed with traditional and outdated methods. In this paper we present the process of the development of an up-to-date sentiment analysis tool based on neural networks. The first part describes the theoretical framework of both sentiment analysis and deep learning, followed by the description of the making a specialized Twitter corpus about earthquakes. The corpus is annotated by two annotators and as such gives a unique perspective regarding the unusuality of the language used on Twitter. The most common difficulties in the process are briefly described as well. The final chapter of the paper describes the making of a sentiment analysis tool built with recurrent neural networks and trained on the mentioned corpus. The results of the sentiment analysis tool are contextualized both by comparing them to similar tools presented in other papers and the inter-annotator reliability calculated for Twitter corpus.

Keywords: sentiment analysis, neural networks, corpus building, tweet, earthquake