

Sentiment tvitov z uporabo leksikona čustveno obarvanih besed, prevedenega s samodejnim prevajalnikom

Marko Meža

Univerza v Ljubljani, Fakulteta za elektrotehniko, Tržaška c. 25, 1000 Ljubljana, Slovenija
E-pošta: marko.meza@fe.uni-lj.si

Povzetek. Twitter je družbeno omrežje s širokim dosegom. Objave, tj. tviti, o aktualnih temah se na omrežju pojavijo hitro, zato je omrežje zanimiv vir javnega mnenja. Del merjenja javnega mnenja je ugotavljanje, ali je določena tema sprejeta pozitivno ali negativno. V ta namen tvite razvrščamo na pozitivne in negativne. V prispevku je opisan pristop razvrščanja tvitov v slovenskem jeziku na pozitivne in negativne na osnovi leksikona afektivnih besed. Kot izhodišče je služil leksikon AFFIN-165 v angleškem jeziku z uporabo samodejnega prevajalnika preveden v slovenski jezik. Delovanje razvrščanja je bilo preverjeno na bazi z afektom označenih tvitov v slovenskem jeziku. Uporabljenih je bilo 5.806 tvitov. Opisana metoda je tvite razvrščala na pozitivne in negativne s 62,4-odstotno točnostjo, kar se ne razlikuje veliko od točnosti ujemanja ocen tvitov s strani človeških ocenjevalcev. V primerih, ko je isti tvit ocenjevalo več ocenjevalcev, so se namreč ocene razvrščanja tvitov na pozitivne in negativne ujemale v 64,4 % primerih.

Ključne besede: tweet, leksikon, sentiment, AFFIN

Analysing sentiment in tweets using an automatically translated affective lexicon

Twitter is a social network with a wide reach. Tweets on specific topics quickly appear on the network, making it an interesting source of public opinion. Part of measuring public opinion is to determine whether a particular topic is perceived positively or negatively by observing whether tweets about the topic are positive or negative. This paper describes an approach to classifying tweets in Slovenian as positive or negative based on a lexicon of affective words. The starting point is the English lexicon AFFIN-165, which was translated into Slovenian using automatic translation. The classification performance is tested using a database of affectively labeled tweets in Slovenian. 5806 tweets were used. The described method classifies tweets as positive or negative with an accuracy of 62.4 %, which is not significantly different from the accuracy of ratings made by human raters. When the same tweet was rated by multiple raters, the ratings used to classify tweets as positive or negative matched 64.4 % of the time.

Keywords: tweet, lexicon, sentiment, AFFIN

1 UVOD

Obstaja mnogo metod spremljanja javnega mnenja. Ena od elegantnih metod je analiza javnih objav na družbenih omrežjih [1]. Twitter* je razširjeno in široko uporabljano družbeno omrežje. V aprilu leta 2022 je imel 465,1 milijona uporabnikov, kar predstavlja 5,9 % celotne populacije planeta†.

Prejet 6. julij, 2022

Odobren 8. avgust, 2022

*<https://twitter.com/>

†<https://datareportal.com/essential-twitter-stats>

Na družbenem omrežju Twitter uporabniki objavljajo mikrobloge – tvite v obliki kratkih tekstovnih zapisov. V času pisanja tega članka je omejitev dolžine teksta 280 znakov. Način objave tvitov z uporabo lojtre (angl. hashtag) pred ključnimi besedami v tvidu omogoča iskanje po ključnih besedah oz. iskanje tvitov glede na izbrano temo. Uporabniki dnevno objavijo 550 milijonov tvitov [1].

Tviti so torej aktualen vir javnega mnenja o izbranih temah. Zaradi preprostosti objavljanja so objave o aktualni temi ali dogodku hitro na voljo, kar pomeni, da je mogoče razmeroma hitro začeti meriti javno mnenje o temi ali dogodku.

Zaradi velikega števila objav je analizo javnega mnenja smiselno opraviti avtomatsko ali vsaj delno avtomatsko. Eno od podpornih orodij za analizo javnega mnenja je analiza sentimenta v tvidih in njihovo razvrščanje na pozitivne, nevtralne in negativne.

Pričujoči prispevek opisuje preizkus delovanja metode analize sentimenta tvitov za slovenski jezik na osnovi leksikona afektivnih besed. Leksikon afektivnih besed smo povzeli po avtorju [2]. Izhodišče za analizo je bil leksikon AFFIN-165 v angleškem jeziku. Za analizo tvitov v slovenskem jeziku smo leksikon z uporabo samodejnega prevajalnika‡ prevedli v slovenski jezik. Delovanje analize sentimenta z uporabo prevedenega leksikona smo preizkusili na slovenski bazi z afektom označenih tvitov [3]. Baza je zajemala ID-številke tvitov, s strani človeškega ocenjevalca ocenjeni sentiment tvita

‡<http://translate.google.com>

v obliki diskretne ocene: [pozitiven, nevtralen, negativen] in ID-številko ocenjevalca.

Članek je razdeljen na 5 poglavij. V sledečem poglavju so predstavljena povezana dela. Sledita poglavje z opisom uporabljenih orodij in postopkov za izdelavo modela za razvrščanje tvitov na pozitivne in negativne ter opis metodologije preverjanja delovanja modela. V naslednjem poglavju so predstavljeni rezultati. V zaključku je delovanje predstavljene metode postavljeno v kontekst točnosti človeških razvrstitev tvitov. Predlagane so tudi ideje za nadaljnje raziskave.

2 POVEZANA DELA

Analiza sentimenta tvitov je popularen problem, ki ga rešuje mnogo avtorjev. Načini analize se v grobem delijo na pristope na osnovi leksikona in pristope na osnovi strojnega učenja. Pristopi na osnovi strojnega učenja se nadalje delijo na pristope z nadzorovanim učenjem in pristope z nenadzorovanim učenjem. Pristopi na osnovi leksikona pa se delijo na pristope na osnovi slovarja in na pristope na osnovi korpusa besed, ki se nadalje deli na statistične in semantične pristope [4]. Kot metrika uspešnosti delovanja pristopov v primeru razvrščanja se običajno uporablja točnost (angl. accuracy). Pogosto se podajata tudi natančnost (angl. precision) in občutljivost (angl. recall) za posamezne razrede.

Ocena sentimenta v tekstu je naloga za analizo naravnega jezika (angl. Natural Language Processing, NLP). V te namene je razvitih mnogo orodij, veliko jih je razvitih posebej za analizo tvitov.

V prispevku [4] je predstavljen pristop na osnovi leksikona, z uporabo preklapljanja polaritete. Avtorji so identificirali pridevnike, ki negirajo pozitivno valenco besed iz leksikona. Kot primer avtorji podajajo obravnavo niza "less delicious", kar pretvorijo v valenco $-1 \times 1 = -1$. Avtorji ugotavljajo, da so pristopi na osnovi leksikona manj učinkoviti kot pristopi na osnovi strojnega učenja. V študiji je uporabljen mali podatkovni nabor, in sicer nekaj 100 tvitov.

V prispevku [1] avtorji za razvrščanje sentimenta tvitov med sabo primerjajo različne pristope nadzorovanega strojnega učenja. Ugotavljajo, da najboljše rezultate na uporabljenem korpusu dosežejo z naivnim bayesovim klasifikatorjem, in sicer točnost 86 %.

V prispevku [5] je predstavljena študija analize sentimenta tvitov o filmih na podlagi v okviru študije zbranih in s strani avtorjev prispevka ročno razvrščenih tvitov. Avtorji so za gradnjo modela uporabili strojno učenje in dosegli točnost od 61 % do 84 %.

Avtorji prispevka [6] so uporabili naivni bayesov klasifikator za razvrščanje tvitov, ki se nanašajo na predsedniške volitve, na pozitivne in negativne. Bazo 433 tvitov so na pozitivne in negativne razvrstili ročno in na podatkih zgradili klasifikacijski model. Z uporabo naivnega bayesovega klasifikatorja so dosegli točnost 76 %.

Prispevek [7] opisuje analizo sentimenta tvitov med zaprtjem države zaradi pandemije covid-19 na Novi Zelandiji. V ta namen so avtorji pridobili 1.162 glede na lokacijo in temo aktualnih tvitov. Analizo so izvedli na osnovi zelo razširjenega leksikona AFFIN in ugotovili, da je v naboru uporabljenih tvitov 71 % negativnih in 29 % pozitivnih.

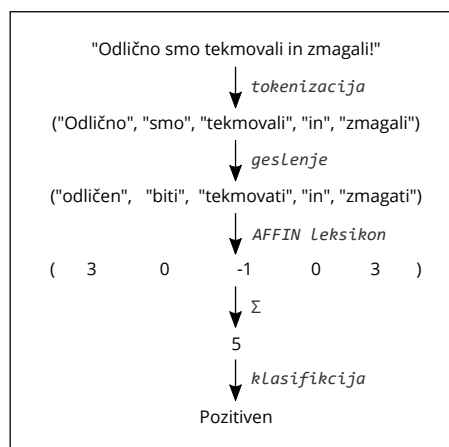
Analiza sentimenta je opravljena tudi za slovenska besedila. Tako je na voljo korpus spletne slovenščine Janes [8], [9], ki vsebuje tvite, spletne forume, novice in uporabniške komentarje nanje, blogovske zapise in komentarje nanje ter uporabniške in pogovorne strani na Wikipediji. Prispevek opisuje postopek zajema besedil za vsakega od vključenih virov podaja kvantitativno analizo zgrajenega korpusa. Sledi predstavitev avtomatskih in ročnih postopkov za obogatitev korpusa s koristnimi metapodatki, kot so tip, spol in regija avtorja ter sentiment in stopnja tehnične in jezikovne standardnosti posameznega besedila.

3 UPORABLJENA ORODJA IN POSTOPKI

Poglavje opisuje uporabljene metode za analizo sentimenta v tvitih, za testiranje uporabljeno podatkovno bazo in validacijski postopek.

3.1 Postopek določanja sentimenta teksta tvita

V študiji uporabljena metoda temelji na leksikonu afektivnih besed. V leksikonu je besedam pripisana valenca, ki jo beseda nosi. Algoritem v tekstu analiziranega tvita poišče besede iz leksikona afektivnih besed ter sešteje njihove valence in tvit razvrsti glede na doseženo vsoto. Postopek je orisan na sliki 1, kjer je ilustrirana klasifikacija teksta "Odlično smo tekmovali in zmagali" v razred *Pozitiven*.



Slika 1: Ilustracija delovanja postopka razvrščanja tweetov.

Za izračun ocene sentimenta posameznega tvita smo kot izhodišče uporabili odprto kodo projekta AFFIN, kjer je poleg leksikona na voljo tudi izvorna koda v programskem jeziku python, ki za dani niz poišče besede iz leksikona in sešteje vrednosti valence posameznih

najdenih besed. Besede v analiziranem tvitu so zapisane s slovnično obliko, koda pa išče popolno ujemanje besed iz zapisa in leksikona, zato je treba tvite predobdelati. Prvi korak predobdelave je tokenizacija.

3.1.1 Tokenizacija: V postopku določanja sentimenta teksta tvita tega najprej razbijemo na posamezne besede. Postopek se imenuje tokenizacija (angl. tokenization). Tokenizacija tekst razbije na dele na osnovi znakov, ki besede ločijo med sabo. To so presledki in ločila. Po tokenizaciji dobimo sekvenco posameznih besed, ki predstavlja vhod v naslednji korak.

3.1.2 Geslenje: Ker imajo v analiziranem tekstu besede mnogo slovničnih oblik, v leksikonu pa so besede v osnovni obliki, bi bila analiza teksta v izvorno zapisani obliki neučinkovita. Če bi na primer analizirali besedo *hodim*, v leksikonu pa je zapisana beseda *hoditi*, metoda za to besedo ne bi delovala. Zato iz slovnično pravilne oblike posameznih besed te pretvorimo v njihove osnovne in med sabo lažje primerljive oblike. Preprost način za pretvorbo, ki ni vedno uspešen, je krnjenje (angl. stemming). Krnjenje besedam odreže končnico besede. Primer krnjenja besede *hodim* je *hodi*, primer krnjena besede *biti* je *bi*.

Boljši, a za izvedbo zahtevnejši postopek je geslenje teksta (angl. lemmatization). Geslenje besedo v poljubni slovnični obliki pretvori v njeno osnovno obliko. Primer geslenja besede *hodim* je *hoditi*, primer geslenja besede *boljšim* je *dober*. V študiji je za geslenje teksta tvitov uporabljena knjižnica Lemmagen* v programskem jeziku python, izdelana na osnovi študije, opisane v [10].

3.1.3 Izračun vrednosti valence tvita: Iz niza besed v osnovni obliki je v nadaljevanju za izračun vrednosti valence uporabljena knjižnica *affin*[†] [2], ki smo jo ustrezno popravili, da je uporabila iz angleškega v slovenski jezik samodejno prevedeno bazo afektivnih besed AFFIN-SI22-165. Knjižnica naloži leksikon besed in vsako besedo iz analiziranega niza poišče v leksikonu. Če jo najde, k skupni vrednosti valence analiziranega tvita prišteje vrednost valence besede.

3.2 Leksikon AFINN

Ocena sentimenta v tekstu je izvedena z uporabo leksikona afektivnih besed AFINN-165 [2]. Obstaja več leksikonov AFFIN. Kot izhodišče je bil uporabljen leksikon AFFIN-165-en, ki vsebuje 3.382 besed v angleščini. Besede v leksikonu so označene z vrednostjo valence v razponu od -5 do 5, pri čemer se -5 ali -4 interpretira kot zelo negativna, 4 ali 5 pa kot zelo pozitivna valenca. Celotna interpretacija je predstavljena v tabeli 1.

Leksikon AFFIN-165-en je bil ustvarjen ročno, z označevanjem zbirke afektivnih besed, ki je bila pridobljena iz tvitov v več iteracijah in preostalih virov, opisanih v prispevku [2], kjer je leksikon tudi podrobneje opisan. Leksikon je na voljo javno[‡]. Poleg leksikona

Tabela 1: Interpretacija vrednosti valence afektivnih besed v leksikonu AFFIN-165 [2].

Valenca	Interpretacija
-5 ali -4	Zelo negativna
-3, -2 ali -1	Negativna
1, 2 ali 3	Pozitivna
4 ali 5	Zelo pozitivna

Tabela 2: Ilustracija nekaterih samodejno prevedenih afektivnih besed iz angleškega leksikona AFFIN-EN-165 v slovenskega AFFIN-SI22-165.

Valenca	Angleški	Slovenski
5	outstanding	... izjemen
4	terrific	... super
3	charming	... očarljiv
2	approval	... odobritev
1	clear	... jasno
-1	cry	... jokati
-2	abandon	... opustiti
-3	abuse	... zloraba
-4	catastrophic	... katastrofalno
-5	bastard	... baraba

v angleščini AFFIN-165-en obstajajo tudi leksikoni v finščini AFFIN-165-fi, poljščini AFFIN-165-pl, turščini AFFIN-165-tr, francoščini AFFIN-165-fr in švedščini AFFIN-165-sv. Ker smo potrebovali orodje za ocenjevanje slovenskega teksta, smo leksikon AFFIN-165-en prevedli v slovenski jezik. Prevajanje smo izvedli s pomočjo avtomatskega prevajalnika Google translate[§]. Za prevedene afektivne besede smo obdržali vrednosti valence iz izvornega leksikona. Primer dela izvornega in prevedenega leksikona je predstavljen v tabeli 2.

V nadaljevanju smo leksikon s prevodi z uporabo postopka geslenja, opisanega v poglavju 3.1.2, preoblikovali tako, da smo vse besede pretvorili v osnovno obliko. Ker je bilo nekaj besed prevedenih v večbesedne izraze, smo te izraze odstranili. V nastalem leksikonu AFINN-SI22-165 je od izvornih 3.382 besed ostalo 3.246 besed. Porazdelitev vrednosti valence besed v nastalem leksikonu je predstavljena v grafu 2 in tabeli 3.

3.3 Razvrščanje

V postopku smo do tega trenutka pridobili skalarno vrednost, ki je opisovala vrednost valence celotnega tvita. V nadaljevanju je bilo treba določiti mejo med pozitivnimi in negativnimi tviti na podlagi vrednosti valence. Za določitev meje smo uporabili naivni bayesov klasifikator, ki je mejo določil na podlagi porazdelitve ocen v učni množici, opisani v poglavju 3.5.

3.4 Vrednotenje rezultatov

Kot osnovno mero za uspešnost smo izbrali točnost razvrščanja (angl. accuracy). Točnost podaja delež pra-

*<https://github.com/izacus/SlovenianLemmatizer-Python>

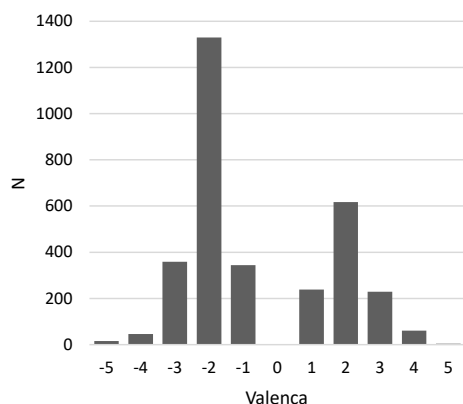
[†]<https://github.com/fnielsen/afinn>

[‡]<https://github.com/fnielsen/afinn>

[§]<https://translate.google.com/>

Tabela 3: Porazdelitev vrednosti valence v leksikonu AFINN-SI22-165.

Valenca	-5	-4	-3	-2	-1	0	1	2	3	4	5
N	16	46	359	1330	344	1	239	617	229	61	4



Slika 2: Porazdelitev vrednosti valence v leksikonu AFINN-SI22-165.

vilno razvrščenih tvitov in je prikazana v enačbi 1. Kot zlati standard smo uporabili s strani človeškega ocenjevalca ocenjene tvite. Zanimalo nas je razvrščanje na pozitivne in negativne tvite. Za to smo nevtralno označene tvite izločili iz postopka. Za večkrat ocenjene tvite smo določili povprečno oceno in uporabili le to.

$$\text{točnost} = \frac{\text{število pravih razvrstitev}}{\text{število vseh razvrstitev}} \quad (1)$$

3.5 Vir podatkov

V raziskavi smo kot učno in testno množico uporabili ročno označeno bazo tvitov 15 evropskih jezikov [3]. Baza vsebuje ID-številke tvitov, ID ocenjevalca in ročno oceno razvrstitve tvita [*Negativna*, *Nevtralna*, *Pozitivna*]. V študiji smo uporabili del baze s tviti v slovenskem jeziku.

Ker so bile v bazi, ki smo jo pridobili poleg ocen, le ID-številke tvitov, je bilo za analizo treba pridobiti tekst tvitov. Twitter omogoča dostop do tvitov preko programskega vmesnika, ki je na voljo razvijalcem aplikacij. Dostop je na voljo različnim namenom. Eden od namenov so tudi raziskave na osnovi podatkov o tvitih. Tako je preko vmesnika mogoče pridobiti posamezne tvite na osnovi različnih iskalnih kriterijev. Za dostop do vmesnika se je treba registrirati kot razvijalec in na spletni strani za razvijalce[¶] pridobiti ključne za dostop do programskega vmesnika API. V študiji smo za dostop do vmesnika in branje teksta tvitov uporabili python knjižnico tweepy^{||}.

V izbrani bazi slovenskih tvitov je bilo 133.936 zapisov. Z uporabo vmesnika nam je uspelo pridobiti tekst 14.739 tvitov. Ker je bila uporabljena baza razmeroma

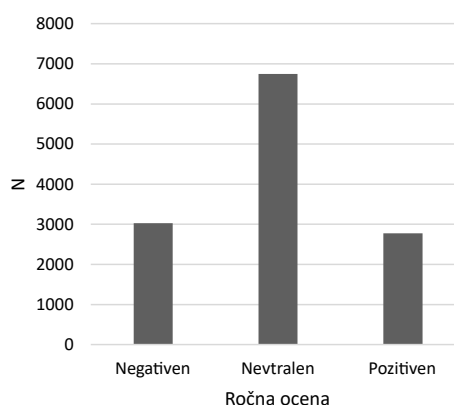
stara, nam preostalih tvitov ni uspelo pridobiti, saj jih API ni vrnil. Twitter namreč stare tvite briše, tvite pa brišejo tudi uporabniki.

Od pridobljenih tvitov je bil delež, 17,5 %, ocenjen večkrat, s strani več ocenjevalcev. Ocene večkrat ocenjenih tvitov se niso vedno ujemale. Zato smo za ocene tvitov, ki so se pojavile večkrat, izračunali povprečno oceno na sledeči način: Negativni oceni smo pripisali vrednost -1, nevtralni 0 in pozitivni +1. Vrednosti smo sešteli. Če je bila vsota manj kot 0, smo tvitu pripisali oceno negativno, če je bila vsota enaka 0, oceno nevtralno, če je bila vsota več kot 0, oceno pozitivno. Preslikava za izračun povprečno ročno določene ocene je predstavljena v tabeli 5. Po združevanju smo imeli na voljo 12.554 tvitov. Porazdelitev povprečnih ročno določenih ocen je prikazana na sliki 3.

Tabela 4: Izračun povprečne ocene ročne razvrstitve tvita, imenovane šibki zlati standard.

Izvorna ocena	Točke	Pogoj (točke)	Ocena
Negativen	-1	< 0	Negativen
Nevtralen	0	= 1	Nevtralen
Pozitiven	1	> 0	Pozitiven

Obdržali smo le tvite, ki so bili ocenjeni s pozitivno ali negativno oceno. Tako nam je ostalo 5.806 tvitov, s katerimi smo preizkusili naš sistem razvrščanja tvitov na pozitivne in negativne.



Slika 3: Številčna porazdelitev povprečne ocene ročne razvrstitve tvitov.

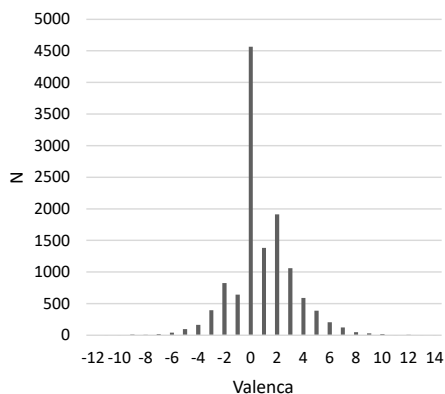
4 REZULTATI

Metodo razvrščanja tvitov na pozitivne in negativne smo preizkusili na bazi slovenskih tvitov [3], pridobljenih na način, opisan v poglavju 3.5. Z uporabo leksikona AFINN-SI22-165 smo izračunali vrednosti valence tvitov. Porazdelitev z našo metodo dobljenih ocen je prika-

[¶]<https://developer.twitter.com/>

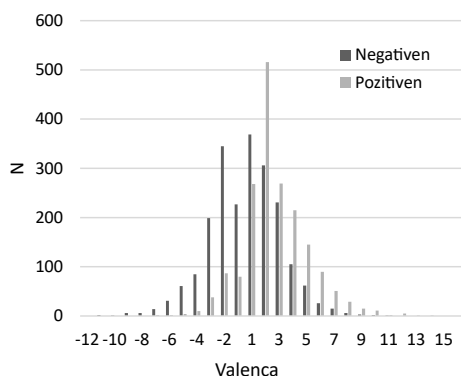
^{||}<https://www.tweepy.org/>

zana na sliki 4. Opazimo, da je znaten delež tvitov (36,4 %) ocenjen z oceno 0. To pomeni, da naša metoda ni našla gesel iz leksikona, ali so bila gesla nevtralna, ali je bila vsota negativnih in pozitivnih ocen enaka 0.



Slika 4: Številčna porazdelitev vrednosti valence tvitov, izračunane z uporabo leksikona AFINN-SI22-165.

Porazdelitvi z našo metodo izračunanih vrednosti valenc za razreda ročno določenih pozitivnih in negativnih ocen tvitov z izpuščenimi vrednostmi valenc 0 sta prikazani na sliki 5. Opazimo, da se porazdelitvi prekrivata, a se njuna vrhova razlikujeta.



Slika 5: Številčna porazdelitev vrednosti valence tvitov, izračunane z uporabo leksikona AFINN-SI22-165. Prikazani so le tviti, katerih ročna ocena je bila negativna ali pozitivna. Izločeni so tudi tviti z vrednostjo valence 0.

Za razvrščanje tvitov na pozitivne ali negativne glede na izračunano valenco smo uporabili naivni bayesov klasifikator. Uporabljena je bila navzkrižna, 10-pregibna validacijska shema.

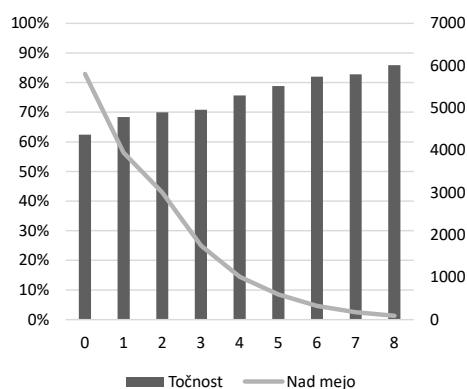
Z vsemi podatki je klasifikator deloval z 62,4-odstotno točnostjo. Matrika razvrščanja (angl. confusion matrix) je prikazana v tabeli 5. V tabeli so zapisani tudi podatki o natančnosti (angl. precision) in občutljivosti (angl. recall) posameznih razredov.

Zanimali so nas ekstremi, zato smo se odločili analizirati le tvite z dovolj visoko absolutno vrednostjo ocene valence. Opazovali smo, kako se obnaša klasifikator, če

Tabela 5: Matrika razvrščanja naivnega bayesovega klasifikatorja na vseh podatkih. Točnost = 62,4 %.

	Poz.	Neg.	Natančnost
p. Poz.	1351	757	64,1 %
p. Neg.	1425	2273	61,5 %
Občutljivost	48,7 %	75,0 %	

v analizo vključimo tvite z različno velikimi absolutnimi vrednostmi valence. Prag smo povečevali in analizirali točnost delovanja klasifikatorja. Rezultati so predstavljeni na grafu 6, kjer je s palicami prikazana dosežena točnost pri izbranem pragu, s črto pa je prikazano število tvitov, ki so dosegli ali presegli mejo in bili uporabljeni v gradnji in preverjanju modela.



Slika 6: Pravilnost razvrščanja tvitov glede na absolutno vrednost ocene.

Če se odločimo za primer, ko je absolutna vrednost valence tvitov več kot 3, je točnost razvrščanja 75,6 %. Matrika razvrščanja je prikazana v tabeli 6.

Tabela 6: Matrika razvrščanja naivnega bayesovega klasifikatorja tvitov z absolutno vrednostjo valence > 3. Točnost = 75,6 %.

	Poz.	Neg.	Natančnost
p. Poz.	558	220	71,7 %
p. Neg.	26	206	88,8 %
Občutljivost	95,6 %	48,4 %	

V ekstremnem, tako rekoč neuporabnem primeru, ko je absolutna vrednost valence tvitov več kot 8, je točnost razvrščanja 86,3 %. Matrika razvrščanja je prikazana v tabeli 7.

Tabela 7: Matrika razvrščanja naivnega bayesovega klasifikatorja tvitov z absolutno vrednostjo valence > 8. Točnost = 86,3 %.

	Poz.	Neg.	Natančnost
p. Poz.	35	6	85,4 %
p. Neg.	1	9	90,0 %
Občutljivost	97,2 %	60,0 %	

5 ZAKLJUČEK

V prispevku je predstavljen poskus samodejne analize sentimenta tvitov v slovenskem jeziku in njihovega razvrščanja na pozitivne in negativne na osnovi leksikona afektivnih besed. Za izhodišče smo uporabili leksikon AFFIN-165 v angleškem jeziku. Leksikon smo z uporabo samodejnega prevajanja prevedli v slovenski jezik in posamezne besede z uporabo geslenja pretvorili v njihove osnovne oblike. Delovanje razvrščanja tvitov smo preverili na testni množici tvitov v slovenskem jeziku, ki so jih človeški ocenjevalci ocenili kot pozitivne nevtralne ali negativne. Kadar je posamezen tweet ocenilo več ocenjevalcev, smo ocene združili in izračunali povprečno razvrstitev. Ker nas je zanimalo razvrščanje na pozitivne in negativne, smo iz testne množice izločili nevtralno ocenjene tvite. Na celotnem preostalem naboru tvitov je bila v prispevku opisana samodejna metoda njihovega razvrščanja na pozitivne in negativne točna v 62,4 % primerih.

Ob tem velja izpostaviti dejstvo, da se ocene človeških ocenjevalcev pogosto ne skladajo. V uporabljeni bazi je 2.185 tvitov ocenilo več ocenjevalcev. Od tega se v 1.454 primerih, kar je 66,5 %, ocene človeških ocenjevalcev niso skladale. Ujemanje človeških razvrstitev na večkrat ocenjenih tvitih je torej le 33,5 %. Neskladanje je izračunano za vse ocene, to je pozitivne, nevtralne in negativne. Če primerjamo zgolj pozitivne in negativne ocene človeških ocenjevalcev, se te pri tvitih ocenjenih več kot enkrat, ne skladajo v 518 primerih, kar pomeni 35,6 % primerov. Če torej rezultat v članku opisane metode primerjamo z rezultatom človeških ocenjevalcev, lahko sklepamo, da so ocene med sabo primerljive. Samodejna metoda, testirana na celotni bazi, je bila 62,4-odstotno uspešna, ocene človeških ocenjevalcev pa se pri razvrstitvi tvitov na pozitivne in negativne skladajo v 64,4 % primerih.

Zapisano kaže, da je ugotavljanje javnega mnenja o posamezni temi na podlagi analize sentimenta iz zgolj enega tvita najverjetneje nezanesljivo. Če za določeno temo zberemo večje število tvitov, je lahko ocena zanesljivejša. Študija zanesljivosti določitve javnega mnenja posamezne teme presega okvir tega prispevka in je lahko izhodišče za nadaljnje raziskave. Prav tako bi bilo zanimivo raziskati pristranskost posameznih ocenjevalcev k pozitivnim ali negativnim razvrstitvam.

Ob povečevanju praga absolutne vrednosti valence tvita opazimo, da je točnost razvrščanja tvitov na pozitivne in negativne vse večja. Večja je absolutna vrednost valence tvita, bolj zanesljivo lahko posamezen tweet razvrstimo med pozitivne ali negativne. Ob povečevanju praga absolutne vrednosti valence tvita sta točnost in zanesljivost napovedi posameznega tvita vse večji, a je na račun tega razvrščenih vse manj tvitov.

Način uporabe samodejnih postopkov za razvrščanje tvitov na pozitivne in negativne je treba uporabiti s pravim poznavanjem ozadja in z jasno določenim ciljem

uporabe. Zmotno je pristop uporabiti brez razumevanja njegovega delovanja in slepo slediti rezultatom.

ZAHVALA

Raziskavo je omogočilo Ministrstvo za visoko šolstvo, znanost in tehnologijo Republike Slovenije v okviru programa P2-0246 – ICT4QoL – Informacijsko komunikacijske tehnologije za kakovostno življenje.

LITERATURA

- [1] L. Mandloi, R. Patel, "Twitter Sentiments Analysis Using Machine Learning Methods", *International Conference for Emerging Technology (INCET)*, pp. 1–5, 2020.
- [2] F.Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs", *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, 2011.
- [3] I. Mozetič, M. Grčar, J. Smailović, "Twitter sentiment for 15 European languages", *Slovenian language resource repository CLARIN.SI*, ISSN 2820-4042, 2016
- [4] Kusrini, M. Mashuri, "Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication", *International Conference of Artificial Intelligence and Information Technology (ICAIT)*, pp. 365–368, 2019.
- [5] A. Blatnik, K. Jarm, M. Meža, "Movie sentiment analysis based on public tweets," *Elektrotehniški vestnik*, vol. 81, no. 4, pp. 160–166, 2014.
- [6] M. Wongkar, A. Angdresy, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter", *Fourth International Conference on Informatics and Computing (ICIC)*, pp. 1–5, 2019.
- [7] O. Baker, J. Liu, M. Gosai, S. Sitoula, "Twitter sentiment analysis using machine learning algorithms for COVID-19 outbreak in New Zealand," *Computer Science and Data Mining, December 2021 in Auckland*, Vol:15, No:12, pp. 649–654, 2021.
- [8] D. Fišer, T. Erjavec, N. Ljubešič, "JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin", *Slovenščina 2.0* vol.4, no 2, pp. 67–100, 2016.
- [9] T. Erjavec, N. Ljubešič, D. Fišer, *Korpus slovenskih spletnih uporabniških vsebin Janes. Viri, orodja in metode za analizo spletne slovenščine*, Znanstvena založba Filozofske fakultete Univerze v Ljubljani, 2018.
- [10] J. Plisson, N. Lavrač, D. Mladenec, "A rule based approach to word lemmatization," *Zbornik C 7. mednarodne multi-konference Informacijska družba IS 2004*, pp. 83–86, 2004.

Marko Meža (Senior member, IEEE) je leta 2001 diplomiral in leta 2007 doktoriral na Fakulteti za elektrotehniko, Univerze v Ljubljani. Od leta 2013 je docent na Fakulteti za elektrotehniko, kjer je trenutno član Laboratorija za osnove elektrotehnike in elektromagnetiko. Raziskovalno se ukvarja z uporabo sodobnih pristopov obdelave, strojnega učenja in podatkovnega rudarjenja na signalih iz tehnike, medicine in socialne interakcije.