

PRIKAZ METODE ANALIZE STAVOVA I MIŠLJENJA KORISNIKA NA PRIMJERU DRUŠTVENE MREŽE TWITTER

PRESENTATION OF THE METHOD FOR ANALYSIS USERS ATTITUDES AND OPINIONS ON EXAMPLE SOCIAL NETWORK TWITTER

Željko Širanović¹, Domagoj Petelinc²

¹Tehničko veleučilište u Zagrebu, Vrbik 8, 10000 Zagreb, Hrvatska

²Tehničko veleučilište u Zagrebu, Vrbik 8, 10000 Zagreb, Hrvatska, student

SAŽETAK

Područje istraživanja ovog rada je rudarenje podataka na društvenim mrežama, analiza stavova i mišljenje korisnika o nekoj temi, koristeći tehnike umjetne inteligencije odnosno procesiranja prirodnog jezika. Prikaz tehnike prikupljanja stavova i mišljenja korisnika o određenoj temi odrađena je na društvenoj mreži Twitter. Tehnikama procesiranja prirodnog jezika obavljena je obrada i analiza prikupljenih tekstualnih informacija, a konačni rezultati stavova korisnika (pozitivni, negativni ili neutralni) prikazani su grafički.

Kao konačni rezultat rada predlažu se alati i model prikupljanja, analize stavova i mišljenja korisnika društvenih mreža o bilo kojoj temi, društvenom događanju, proizvodu i slično.

Ključne riječi: *umjetna inteligencija, strojno učenje, procesiranje prirodnog jezika, analiza osjećaja, društvene mreže*

ABSTRACT

The research area of this paper is data mining on social networks, analysis of attitudes and opinions of users on a topic, using artificial intelligence techniques, i.e. natural language processing. The presentation of the technique of collecting views and opinions of users on a certain topic was done on the social network Twitter. Natural language processing techniques were used to process and analyze the collected textual information, and the final results of user

attitudes (positive, negative or neutral) were presented graphically.

As a final result of the work, tools and a model for collecting, analyzing the attitudes and opinions of users of social networks on any topic, social event, product, etc. are proposed.

Keywords: *artificial intelligence, machine learning, natural language processing, sentiment analysis, social networks*

1. UVOD

1. INTRODUCTION

Veliki broj ljudi na internetu uključen je u društvene mreže i skoro svakodnevno nešto objavljuju, iskazuju svoja mišljenja, stavove, kritike i slično. Zbog toga društvene mreže postaju dobar izvor informacija, koje se mogu prikupiti, obraditi i analizirati i dobiti uvid o tome kako ljudi gledaju na pojedina društvena kretanja, događanja, situacije, proizvode, usluge i slično.

U okviru strojnog učenja i obrade prirodnog jezika, analiza raspoloženja (engl. Sentiment Analysis) može se koristiti kao jedan od načina za otkrivanje mišljenja i stavova korisnika društvenih mreža o nekoj temi. U tu svrhu koriste se metode procesiranja prirodnog jezika (engl. Natural Language Processing – NLP). Metode su obično automatizirani procesi izdvajanja stavova i mišljenja iz pisanog teksta, tweetova i slično. Analiza prikupljeni informacija najčešće uključuje klasificiranje mišljenja u kategorije kao što su "pozitivno",

"negativno" ili "neutralno". Ovaj proces se također često naziva analiza subjektivnosti, istraživanje mišljenja i izdvajanje procjene [1].

U ovom projektu prikazane su metode i tehnike prikupljanja i obrade relevantnih podataka, njihova klasifikacija kroz izračunavanje broja pozitivnih, negativnih i neutralnih osjećaja kako bi se dobilo subjektivno mišljenje o nekoj temi. Za praktični primjer primjene metode analize mišljenja korisnika korišteni su podaci sa društvene mreže Twitter o hrvatskoj nogometnoj reprezentaciji na svjetskom prvenstvu u Kataru 2022. Na Twitteru su od korisnika prikupljena mišljenja vezana za dva pitanja:

1. Da li je Dominik Livaković na ispravan način zaustavio Juliana Alvareza na utakmici Argentina – Hrvatska nakon čega je talijanski sudac Orsato dosudio penal;
2. Općenito mišljenje o igri Hrvatske reprezentaciji na svjetskom prvenstvu u Kataru 2022.

Društvena mreža Twitter korištena je zbog velikog broj korisnika [2] i za nju postoje dobri Python alati koji se mogu koristiti za prikupljanje, analizu i grafički prikaz podataka osjećaja korisnika na osnovu tekstova koji se objavljuju na društvenim mrežama [3].

Svrha i rezultati analiza stavova (osjećaja) i mišljenja koje imaju ljudi na društvenim mrežama o nekoj temi mogu se iskoristiti za različite svrhe kao na primjer: donošenje daljnjih poslovnih odluka, uvođenje poboljšanja ili inovacija, procjena društvenih kretanja, praćenje reakcija na određene događaje, recenzija

kupaca o proizvodima, praćenja konkurentskih proizvoda i slično.

2. OPIS MODELA

2. MODEL DESCRIPTION

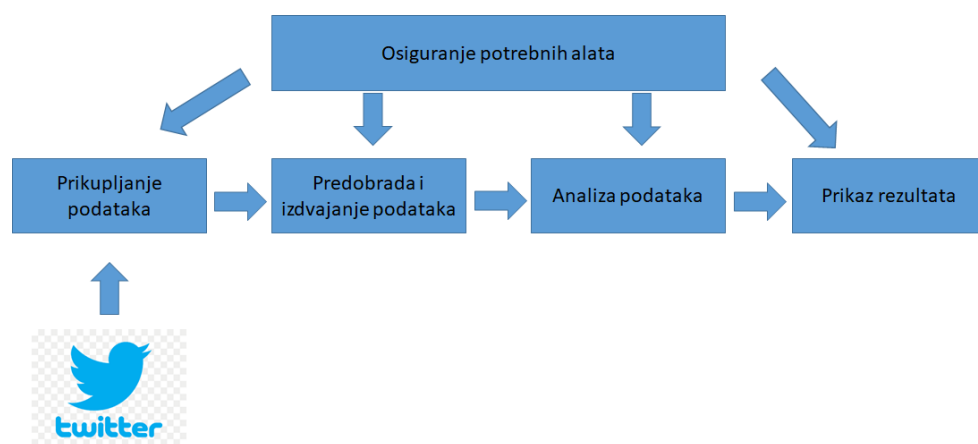
Model prikupljanja, predobrade i analize podataka te prikaz rezultata koji se koristio u ovom istraživanju prikazan je na slici 1. Model se sastoji od nekoliko faza koje slijede jedna iza druge a svaka od njih koristi određene metode i alate. Koraci ovog modela su:

- prikupljanje podataka odnosno ekstrakcija podataka s weba
- predobrada ekstrahiranog teksta a sastoji se od čišćenja tzv. zaustavnih riječi (engl. *stopwords*), interpunkcija, semiotika i ostalih znakova koje imaju nisku semantičku vrijednost, te izdvajanje korisnih riječi koji će ući u daljnju analizu
- analiza tzv. čistih tekstova na osnovu kojih se detektiraju emocionalna stanja tih tekstova u smislu pozitivnih, negativnih ili neutralnih stavova ili mišljenja i vizualizirati učestalost korištenja riječi stvaranjem oblaka riječi.
- Vizualni prikaz rezultata analize osjećaja kroz grafički ili numerički prikaz frekvencija pozitivnih, negativnih ili neutralnih riječi (tweetova).

2.1. PRIKUPLJANJE PODATAKA

2.1. DATA COLLECTION

Proces prikupljanja odnosno ekstrakcija podataka s weba (engl. Web scraping) obično se



Slika 1 Proces analize raspoloženja

Figure 1 The process of sentimental analysis

Tablica 1 Kriterij izbora biblioteke za prikupljanje podataka [6] [7] [8]**Table 1** The tool selection criteria for data collection

KARAKTERISTIKE	Tweepy [Ty]	Snsrape [Se]	Twint [Tt]	IZBOR
Zahtijeva Twitter API vjerodajnice	da	ne	ne	[Se][Tt]
Vremensko ograničenje pristupa prikupljanja tweetova starijih od tjedan dana	da	ne	ne	[Se][Tt]
Ograničenje broja tweetova	da	ne	ne	[Se]
Izvlačenje informacija preko složenijih upita	da	ne	N/a	[Ty]
Primjena na druge društvene mreže (Facebooka, Instagrama, i sl.)	ne	da	ne	[Se]

izvodi pomoću automatiziranih softverskih alata ili prilagođenih skripti. Tehnike Web scrapinga (ponekad se nazivaju rudarenje podataka) koristi se za razne svrhe. Podaci se obično prikupljaju s raznih izvora na mreži kao što su društveni mediji, web stranice konkurencije, tražilice i tržišta e-trgovine. Scraping može pomoći tvrtkama da pridobiju potencijalne klijente, poboljšaju svoju marketinšku strategiju ili može dati neke ideje za mogućnosti ulaganja [4].

Prikupljanje odnosno ekstrakcija može se izvršiti pomoću biblioteka koje služe u tu svrhu (Requests, BeautifulSoup, Cheerio), posebnih API alata (Scrapy, Selenium, ScrapingBee, Smartproxy) ili gotovih alata (ParseHub, Octoparse). Jedan od najpopularnijih programskih alata koji se koristi u tu svrhu je Python i većina web skrapera temelji se na njemu.[5]

Također treba imati u vidu da dobar dio web stranica ne dozvoljavati tzv. scraping te može blokirati ovaj proces tako da blokira IP adresu računala koje to pokušava raditi. Kako bi se neutraliziralo blokiranje IP adrese koriste se proxy poslužitelji koji omogućuje zaobilaznje blokade. Kada se primijeti prigušenje ili blokada preko proxy poslužitelja se postojeća IP adresa zamjenjuje novom, prikrivajući tako izvornu IP adresu i lokacija s koje se scrapira [5]

Uglavnom za prikupljanja podataka s Twittera koristi se API modul ili neka od Python biblioteka, dok se proxy koristi samo da bi se

maskirala IP adresa računala koje prikuplja podatke. U ovom projektu testirana su tri načina (tablica 1) prikupljanja podataka.

Tweepy biblioteka zahtijeva API ključeve i ima ograničenje za prikupljanje tweetova samo na tjedan dana, dok Snsrape nije vremenski ograničen i ne zahtijeva API vjerodajnice (ključeve) ali omogućuje izvlačenje osnovnih informacija kao što su korisnički profili, sadržaji tweeta, izvor i slično.[9]

**Slika 2** Twitter API**Figure 2** Twitter API

Twitter API je interakcijski posrednik između korisnika, aplikacija i drugih API-ja i serverske Twitter pozadine preko koje se pretražuju tweetovi na osnovu tzv. hash tagova i ključnih riječi (slika 2). Osim toga preko Twitter API-ja se mogu obavljati i mnoge druge akcije koje bi se mogle ručno raditi preko sučelja. Da bi se mogao koristiti Twitter API prvo se treba registrirati kao Twitter programer i izraditi Twitter aplikaciju preko koje treba postaviti nekoliko vjerodajnica koje će kasnije koristiti Python Tweepy biblioteka (slika 3).[8]

```

import tweepy
# Tweepy API vjerodajnice
consumer_key = "XXXX" #Your API/Consumer key
consumer_secret = "XXXX" #Your API/Consumer Secret Key
access_token = "XXXX" #Your Access token key
access_token_secret = "XXXX" #Your Access token Secret key

#Prosljeđivanje Twitter API ključeva na provjeru autentičnosti
auth = tweepy.OAuth1UserHandler(
    consumer_key, consumer_secret,
    access_token, access_token_secret
)

# Instanciranje tweepy API
api = tweepy.API(auth, wait_on_rate_limit=True)

```

Slika 3 Prikaz Tweepy API vjerodajnice

Figure 3 Tweepy API credentials

S druge strane besplatni alat za izdvajanje podataka koji ne koristi API vjerodajnice Snsrape, omogućuje izdvajanje podataka osim za Tweeter i druge društvene mreže kao što su: Facebook, Instagram, Reddit, Telegram, VKontakte, Weibo, Mastodon. Ova biblioteka koristi *search_hashtag* funkciju za pretraživanje tweetova koji sadrže određenu riječ ili hashtag koji se nalazi između dva datuma i na jeziku koji

```

import snsrape.modules.twitter as sntwitter
import pandas as pd

def search_hashtag(searchterm, dt_until, dt_since, lang, limit="brojkom
postaviti max broj tweetova"):
# sintaksa upita
query = "({searchterm}) lang:{lang} until:{until}
since:{since}".format(searchterm=searchterm, lang=lang, until=dt_until,
since=dt_since)
# Pretraga tweetova o HR reprezentacij
query = 'Croatian OR #CroatiaWorldCup OR CroatiaFootbal OR
#CroatiaWorldCup OR #Livakovic OR #Gvardiol OR #Modric OR #Reprezentacija
OR #Hrvatska lang:en since:2022-11-20 until:2022-12-18'

#pretraga tweetova Livakovic - penal
query = '#Livaković OR #Alvarez OR #WorldCupQatar2022 OR #Penal OR
#ArgentinaVSCroatia OR Penalty OR #Penalty OR Misjudged Penalty OR
Argentina Penalty kick OR Penalty kick OR Misjudged lang:en since:2022-
11-20 until:2022-12-18'
tweets = []
limit = limit

q = sntwitter.TwitterSearchScrapper(query)
for tweet in q.get_items():
    if len(tweets) == limit:
        break
    else:
        tweets.append([tweet.date, tweet.content])

df = pd.DataFrame(tweets, columns=['Date', 'Tweet'])
csv_name = "data/CroatianFootballTeam/" +
searchterm.replace("#", "hashtag ").replace("\", "").replace(" ", "_") +
"_" + str(since) + "_" + str(until) + ".csv"

df.to_csv(csv_name, index=False)
print("success")

```

odredi korisnik. Funkcija obavlja pretraživanje uzimajući sljedeće parametre (tablica 2) i sprema datum (*Date*) i tekst tweeta (*Tweet*) u rezultatu ovog pretraživanja kao *Pandas DataFrame* u *.csv* formatu.

Tablica 2 Osnovni parametri search_hashtag funkcije

Table 2 The Basic parameters of the search_hashtag function

NAZIV PARAMETRA	ZNAČENJE
<i>searchterm</i>	tražena riječ ili hashtag
<i>dt_since</i>	od datuma u formatu %Y-%M-%D
<i>dt_until</i>	do datuma u formatu %Y-%M-%D
<i>lang</i>	kod jezika
<i>limit</i>	najveći broj tweetova

Preko *search_hashtag* funkcije moguće je izvući tweetove objavljene na engleskom jeziku između datuma 20.11. i 18.12.2022. koji sadrže hashtag #CroatianFootballTeam (slika 4).

Slika 4 Primjer koda za izvlačenje tweetova preko search_hashtag funkcije

Figure 4 Code example for extracting tweets via the search_hashtag function

```

def clean_tweet(tweet):
    if type(tweet) == np.float:
        return ""
    r = tweet.lower()
    r = profanity.censor(r)
    r = re.sub("''", "", r)
    r = re.sub("@[A-Za-z0-9_]+", "", r)
    r = re.sub("#[A-Za-z0-9_]+", "", r)
    r = re.sub(r'http\S+', '', r)
    r = re.sub('[()!?!]', ' ', r)
    r = re.sub('\.[.*?\\]', ' ', r)
    r = re.sub("[^a-z0-9]", " ", r)
    r = r.split()
    stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to",
"from"]
    r = [w for w in r if not w in stopwords]
    r = " ".join(word for word in r)

    return r

```

Slika 5 Dio koda s funkcijom za čišćenja tweetova

Figure 5 A part of code with a tweet cleaning function

2.2. PREDOBRADA I ČIŠĆENJE PODATAKA

2.2. PREPROCESSING AND CLEANING DATA

Predobrada je postupak koji se sastoji od čišćenja suvišnih semiotičkih znakova, tekstova, interpunkcija i ostalih znakova koje imaju nisku semantičku vrijednost, te izdvajanje korisnih riječi koji će ući u daljnju analizu. Objave korisnika na društvenim mrežama kao što je Twitter uključuju posebne oznake (engl. *hashtags*), citiranja, razne poveznice i sl., smatraju se elementima koji imaju vrlo visoku razinu smetnji (engl. *noise*). Zbog tog razloga osim prethodno navedenih predprocesa potrebno je u fazu predobrade uključiti [9][10]:

- pretvaranje svih znakova u mala slova
- čišćenje brojeva
- čišćenje URL-ova
- čišćenje oznaka
- čišćenje hashtagova
- zamjena više od 2 točke razmakom
- ograničavanje više uzastopnih razmaka na jedan razmak
- čišćenje nekih posebnih znakova
- uklanjanje nepristojnih riječi.

Primjer koda za uklanjanje nebitnih i nepristojnih riječi, razmaka, nepotrebnih znakova prikazan je slikom 5.

2.3. ANALIZA PODATAKA

2.3. DATA ANALYSIS

Nakon faze predobrade izdvojenih podataka dobije se očišćeni tekstovi koji su spremni za klasifikaciju njihovih emocionalnih stanja tj. izražavaju li pozitivno ili negativno mišljenje.

Klasifikacija sentimenta može biti binarna klasifikacija (pozitivna ili negativna), klasifikacija u više klasa (negativna, neutralna, pozitivna), regresija ili rangiranje [1].

Unaprijed obučeni modeli koji se koriste za analizu osjećaja dostupni su u biblioteci *Hugging Face Transformers* (instalira se naredbom `!pip install transformers`). Ova biblioteka ima već unaprijed naučene modele za različite zadatke kao što su analiza osjećaja, generiranje teksta, sažimanje itd. U stvari Transformers je cjevovod koji pruža API i alate za jednostavno preuzimanje i korištenje unaprijed obučanih modela za različite namjene kao što su [11]:

- obrada prirodnog jezika (klasifikacija teksta, prepoznavanje imenovanih entiteta, odgovaranje na pitanja, jezično modeliranje, sažimanje, prijevod, višestruki izbor i generiranje teksta),
- računalni vid (klasifikacija slika, detekcija objekata i segmentacija),
- audio (automatsko prepoznavanje govora i audio klasifikacija),
- multimedija (stolno odgovaranje na pitanja, optičko prepoznavanje znakova, izvlačenje informacija iz skeniranih dokumenata, video klasifikacija i vizualno odgovaranje na pitanja).

Tablica 3 Karakteristike klasifikacijskih modela**Table 3** The characteristics of classification models

NAZIV MODELA	OBUČENE KLASSE MODELA		
	POZITIVNA	NEGATIVNA	NEUTRALNA
<i>DistilBERT</i>	DA	DA	NE
<i>Twitter-roBERTa-base</i>	DA	DA	DA

Za potrebe ovog istraživanja korištena su dva obučena modela (tablica 3) DistilBERT i Twitter-roBERTa-base for Sentiment Analysis [11].

Analize sentimenta po modelu *DistilBert* podešena je sa skupom podataka Stanford Sentiment Treebank v2 (SST2). Kada se ovaj obučeni model pokrene njegov tokenizator preuzima i pohranjuje rezultate u objekt koji se zove `sentiment_classifier` (slika 6). Prema zadanim postavkama analize sentimenta koristi istrenirani model klasificirajući tweetove samo kao pozitivne ili negativne, nedostaje neutralni i to je jedan od nedostataka jer je prethodno model obučen samo za dvije klase pozitivnu i negativnu [12].

```
from transformers import pipeline
sentiment_classifier = pipeline('sentiment-analysis')
```

Slika 6 Otvaranje veze prema objektu za sentiment analizu po modelu DistilBERT [11]**Figure 6** Opening a link to the object for sentiment analysis according to the DistilBERT model

Twitter-roBERTa-baza za analizu raspoloženja (slika 7) je jedan od baznih modela obučen na ~58 milijuna tweetova i podešen za analizu raspoloženja pomoću referentne vrijednosti TweetEval za tekstove odnosno riječi za engleski jezik (za više jezični model pogledati XLM-T) [13].

```
from transformers import AutoModelForSequenceClassification, AutoTokenizer,
pipeline
# load model
model = AutoModelForSequenceClassification.from_pretrained("cardiffnlp/twitter-
roberta-base-sentiment")
tokenizer = AutoTokenizer.from_pretrained("cardiffnlp/twitter-roberta-base-
sentiment")
# create pipeline
sa = pipeline("sentiment-analysis", tokenizer=tokenizer, model=model)
```

2.4. PRIKAZ REZULTATA

2.4. DISPLAY OF RESULTS

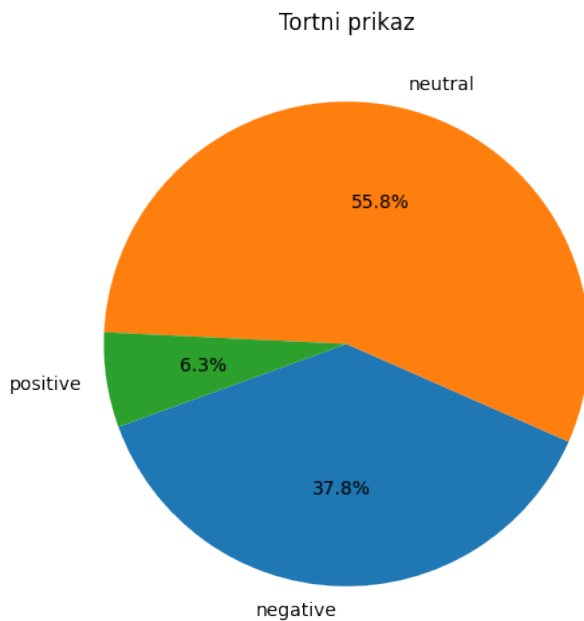
Izlazni rezultati analize osjećaja, stavova ili mišljenja korisnika najčešće se klasificira kroz frekvenciju pozitivnih, negativnih ili neutralnih riječi (tweetova). Vizualno ih je moguće prikazati numerički, grafički ili kao kombinacija oboje. Tako na pitanje da li je Dominik Livaković na ispravan način zaustavio Juliana Alvareza na utakmici Argentina – Hrvatska nakon čega je talijanski sudac Orsato dosudio penal prema iskazanim osjećajima korisnika Twittera klasificirana podaci iskazani su broičano (slika 8).

```
import matplotlib.pyplot as plt
sentiment_counts = df.groupby(['Sentiment']).size()
print(sentiment_counts)
Izlaz:
Sentiment
negative    599
neutral     747
positive    155
dtype: int64
```

Slika 8 Broičani prikaz osjećaja prema broju tweetova**Figure 8** Numerical view of sentiment by number of tweets

Vizualni prikaz klasifikacije osjećaja korisnika također se mogu prikazati u obliku kružnog grafa (slika 9).

Slika 7 Otvaranje veze na sentiment analizu po modelu Twitter-roBERTa-base**Figure 7** Opening a link to the object for sentiment analysis according to the Twitter-roBERTa-base model

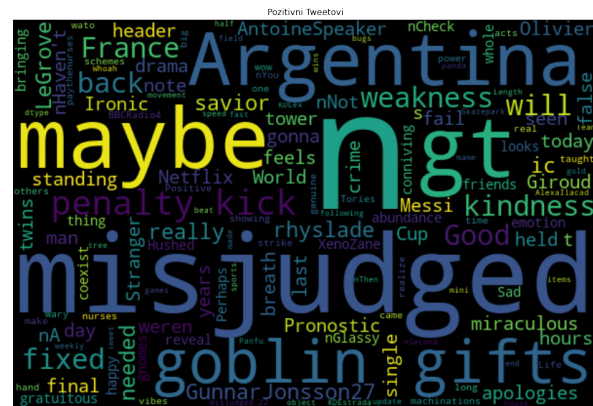


Slika 9 Kružni grafički prikaz klasificiranih osjećaja korisnika izražen u postocima

Figure 9 The pie graphic display of classified users sentiments expressed in percentage

Još jedna tehnika koja se koristi je grafički prikaz frekvencije pozitivnih ili negativnih riječi (tweetova) je pomoću tzv. oblaka riječi (engl. *WordCloud*). U primjeru na slici (slika 10) dan je vizualni prikaz frekvencije pojave pozitivno klasificiranih riječi. Prema prikazu po najveće napisane riječi pojavljuju se češće u tweetovima koji su klasificirani kao pozitivni, kao što je engleska riječ *misjudged*, što dopunjuje prethodni kružni grafički prikaz (slika 9). Iz navedenog se daje zaključiti da je 37,8%

analizirani tweetova bilo iskazano kao loša procijenjen nogometnog suca.



Slika 10 Grafički prikaz frekvencije riječi koje se najviše pojavljuju u tweetovima

Figure 10 The graphic WordCloud representation of the word frequency

Kod za izradu oblaka pozitivno iskazanih riječi dan je slikom (slika 11). U slučaju da se želi dobiti prikaz frekvencije pojave riječi koje se javljaju u negativno klasificiranim tweetovima potrebno je `positive_tweets=df['Tweet'][df["Sentiment"]=='positive']`

zamijeniti sa

`negative_tweets=df['Tweet'][df["Sentiment"]=='negative']`.

Na općenito mišljenje o igri Hrvatske reprezentaciji na svjetskom prvenstvu u Kataru 2022. dobiveni su rezultati kao na slici (Slika 12).

```

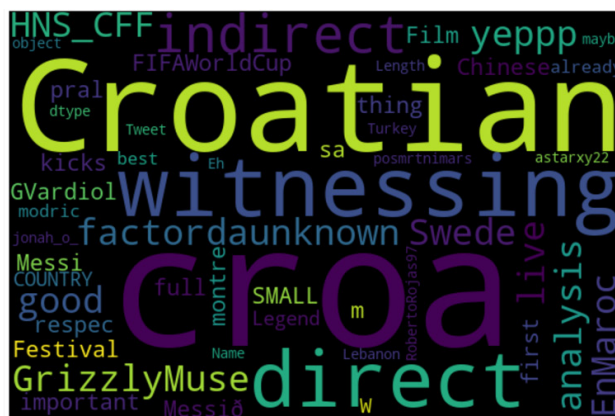
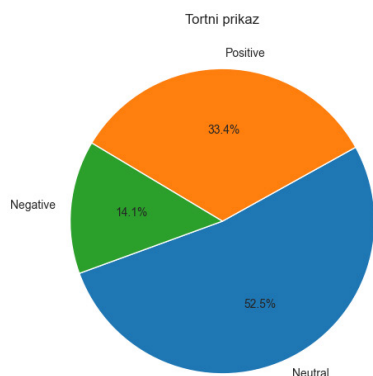
From wordcloud import WordCloud
From wordcloud import STOPWORDS

#Oblak pozitivnih riječi
positive_tweets = df['Tweet'][df["Sentiment"] == 'positive']
stop_words = ["judge", "penal"] + list(STOPWORDS)
positive_wordcloud =
WordCloud(width=600,height=400,background_color="black",stopwords=stop_
words).generate(str(positive_tweets))

plt.figure(figsize=[20,10])
plt.title("Pozitivni Tweetovi")
plt.imshow(positive_wordcloud,interpolation="bilinear")
plt.axis("off")
plt.show()
    
```

Slika 11 Primjer koda za generiranje oblaka riječi

Figure 11 Example of code for generating a wordcloud



Slika 12 Grafički prikazi mišljenja o hrvatskoj reprezentaciji

Figure 12 Graphic representations of opinions about the Croatian national soccer team

2.5. OSTALI ALATI I PREDUVJETI

2.5. OTHER TOOLS AND PREREQUISITES

Za analizu raspoloženju pomoću programskog jezika Python potrebna je verzija Python 3.x i neko razvojno okruženje poput Jupyter Notebook ili Visual studio code i Twitter „delveloper“ račun. Također potrebne su sljedeće Python biblioteke:

- NumPy – osnovna biblioteka za numeričko računanje i rad n-dimenzionalnim nizovima
- Regular Expression – biblioteka s funkcijama za provjeru ključnih riječi u tekstualnim nizovima
- Tweepy – biblioteka sa skupom klasa i metoda za komunikaciju s Twitter API
- TextBlob – biblioteka za obradu tekstualnih podataka
- Matplotlib – vizualizacija podataka i grafičko crtanje
- Pandas – za analizu i manipulaciju podataka
- Wordcloud – vizualizacija učestalosti riječi u tekstualnom sadržaju Better-Profanity – za cenzuriranje loših i vulgarnih riječi u tekstu.

3. ZAKLJUČAK

3. CONCLUSION

U današnje vrijeme strojno učenje kao dio umjetne inteligencije sve više se primjenjuju u mnogim područjima ljudske djelatnosti. U okviru strojnog učenja, metode i tehnike procesiranje prirodnog jezika našle su primjenu i području istraživanja emocionalnih stanja, stavova i mišljenja korisnika društvenih mreža.

U ovom radu objašnjeni su osnovni metodološki postupci, određene tehnike i mogućnosti analize raspoloženja na jednoj od vodećih društvenih mreža Twitter, poznatoj po velikom broju korisnika koji svakodnevno i učestalo objavljuju svoje stavove i diskusije o različitim temama. Uz radu je dan osnovni model postupka analize stavova i mišljenja korisnika društvene mreže koji se sastoji od faza prikupljanja podataka sa društvenih mreža, njihove predobrade i analize te njihovog grafičkog prikaza.

Za fazu prikupljanja podataka sa društvenih preporuča se koristiti Python Snsrape biblioteku jer se može koristiti za preuzimanje podataka sa većine današnjih društvenih mreža (Twitter, Facebook, Instagram, Reddit, Telegram, VKontakte, Weibo, Mastodon). Za analizu i detekciju emocionalnih tekstova na osnovu kojih se detektiraju stanja u smislu pozitivnih, negativnih ili neutralnih stavova ili mišljenja preporuča se koristiti neku biblioteku iz skupa Hugging Face transformera koji sadrže unaprijed obučene modele za klasifikaciju riječi na pozitivne, negativne ili eventualno neutralne. Za vizualno prikazivanje rezultata analize predlaže se numerički i grafički stupčasti ili kružni prikazi, te kao njima kao dopuna prikaz frekvencija pojave pojedinih riječi preko oblaka riječi.

Za testiranje predložene metodologije koristili su se podaci sa društvene mreže Twitter gdje se tražilo mišljenje korisnika oko dva pitanja: da li je ispravno dosuđen kazneni udarac na utakmici Argentina – Hrvatska, i drugo općenitije o tome kakvo je mišljenje o hrvatskoj nogometnoj reprezentaciji.

4. REFERENCE

4. REFERENCE

- [1.] Vishal.A.Kharde, S.Sonawane, Sentiment Analysis of Twitter Data: A Survey of Techniques, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016. ISBN:978-1-5386-8125-1 (<https://arxiv.org/abs/1601.06971>) (5.12.2022.)
- [2.] S. Dixon, Leading countries based on number of Twitter users as of January 2022, Published by Statista, Nov 22, 2022. (<https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>) (5.12.2022.)
- [3.] N. Silaparasetty, What is Twitter Sentiment Analysis in Data Science?. Medium, Aug 18, 2019. (<https://medium.com/ai-for-women/what-is-twitter-sentiment-analysis-in-data-science-c8839e925769>) (6.12.2022.)
- [4.] I. Rivera, Main Uses of Web Scraping: 9 Reasons to Start Gathering Data, Proxyway, December 19, 2022. (<https://proxyway.com/guides/main-uses-of-web-scraping>) (6.12.2022.)
- [5.] I. Rivera, Best Websites to Practice Your Web Scraping Skills. PROXYWAY, May 4, 2022., (<https://proxyway.com/guides/best-websites-to-practice-your-web-scraping-skills>) (6.12.2022.)
- [6.] Massachusetts Institute of Technology, TWINT - Twitter Twitter scraping tool (<https://pypi.org/project/twint/>) (6.12.2022.)
- [7.] JustAnotherArchivist, snsrape -scraper for social networking services (<https://pypi.org/project/snsrape/>)
- [8.] J. Roesslein, Tweepy: Twitter API for Python, MIT (<https://pypi.org/project/tweepy/>) (6.12.2022.)
- [9.] A. Besbes, How To Extract Data From The Twitter API Using Python, Medium, May 16, 2022. (<https://towardsdatascience.com/how-to-extract-data-from-the-twitter-api-using-python-b6fbd7129a33>) (7.12.2022.)
- [10.] B. B. Kömeçoğlu, Part 1: Tweet Scraping and Data Preprocessing, Medium, Nov 21, 2022. (<https://heartbeat.comet.ml/twitter-sentiment-analysis-part-1-6063442c06f3>) (7.12.2022.)
- [11.] Hugging Face community, Hugging Face, (<https://huggingface.co/docs/transformers/index>)
- [12.] D. Subramanian, Scrape Tweets using snsrape and Build a Sentiment Classifier, Medium, Jul 31, 2021. (<https://pub.towardsai.net/scraping-tweets-using-snsrape-and-building-sentiment-classifier-13811dadd11d>) (7.12.2022.)
- [13.] B. B. Kömeçoğlu, Part 2: Sentiment Analysis and Word Cloud Generation, Medium, Nov 28, 2022. (<https://heartbeat.comet.ml/twitter-sentiment-analysis-part-2-5d7fab4e8ef>) (7.12.2022.)

AUTORI · AUTHORS

• **Željko Širanović** - Nepromijenjena biografija nalazi se u časopisu Polytechnic & Design Vol 4 No 4, 2016.

Korespondencija · Correspondence

zsiranovic@tvz.hr

• **Domagoj Petelinc** - rođen je 12.7.1997. u Zagrebu. Završio je XV. Prirodoslovno-matematičku gimnaziju u Zagrebu. Na Tehničkom veleučilištu u Zagrebu 2021.godine završava trogodišnji studij programsko inženjerstvo. Iste godine upisuje na Tehničkom veleučilištu u Zagrebu Stručni diplomski studij informatike smjer računarstvo, gdje i danas studira. Trenutno radi u Ericssonu Nikola Tesla u Zagrebu, u području istraživanja i razvoja 5G sustava. Za sada raspolaže znanjima u programskim jezicima HTML, CSS, Java, Python, C++ i Angular.