Zbornik radova Fakulteta tehničkih nauka, Novi Sad

UDK: 004.85
DOI: https://doi.org/10.24867/27BE20Blagojevic

ANALIZA SENTIMENTA PESAMA

MUSIC SENTIMENT ANALYSIS

Žarko Blagojević, *Fakultet tehničkih nauka, Novi Sad*

Oblast: SOFTVERSKO INŽENJERSTVO I INFORMACIONE TEHNOLOGIJE

Kratak sadržaj – *Rad opisuje poređenje pristupa za klasifikaciju sentimenta pesama. Za potrebe rada kreiran je skup podataka koji sadrži audio karakteristike, tekstove i sentimente pesama. Poređeni su klasični NLP pristupi za klasifikaciju teksta i generativni pristup upotrebom GPT modela..*

Ključne reči: *Analiza sentimenta, RNN, Word2Vec, GloVe, LLM, GPT*

Abstract – *This paper describes a comparison of approaches for classifying the sentiment of songs. For the purposes of the study, a dataset was created that includes audio characteristics, lyrics and song sentiments. Classic NLP approaches for text classification and a generative approach using GPT models were compared*

Keywords: *Sentiment analysis, RNN, Word2Vec, GloVe, LLM, GPT*

1. UVOD

Sistemi za striming muzike, poput Spotify, Deezer, YouTube Music i sl. koriste milijarde ljudi svakodnevno. Pored toga što omogućavaju slušanje gotovo sve muzike ikada zabeležene u digitalnom formatu, od suštinske važnosti za uspeh takvih sistema jeste mogućnost lake pretrage i preporuke muzike korisnicima kako bi što duže vremena proveli na striming platformama. Trenutno raspoloženje korisnika važan je prediktor tipa muzike koju želi da sluša u datom trenutku. Samim tim, sentiment pesme, odnosno raspoloženje koje pesma nosi, predstavlja jedan od ključnih parametara na osnovu kojeg se pretraga i preporuka pesama u ovakvim sistemima može vršiti.

Kako sistemi za strimovanje muzike imaju velik broj pesama, manuelno organizovanje pesama po sentimentu nije moguće izvršiti. Stoga je za rešavanje ovog problema neophodan automatizovan pristup, koji bi sa visokom preciznošću uspeo da svrsta pesme u odgovarajuće klase raspoloženja. U ovom radu prikazano je nekoliko pristupa za klasifikovanje sentimenata na osnovu teksta i audio karakteristika pesama. Najpre su primenjeni različiti pristupi za pretvaranje teksta pesama u vektore (LSTM, GRU, Word2Vec, Glove i BERT) kako bi se zajedno sa audio karakteristikama pesama slali na klasifikaciju. Pored klasičnog NLP pristupa isprobana je i klasifikacija pomoću generativnog modela gpt-3.5-turbo.

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

2. PREGLED STANJA U OBLASTI

Problem klasifikacije sentimenata pesama predstavlja deo oblasti pod imenom Music Information Retrieval. Ova interdisciplinarna oblast bavi se organizacijom, analizom i prikupljanjem raznih informacija vezanih za muziku. To čini kombinujući teoriju muzikologije, procesiranja audio signala, računarskih nauka i od skora mašinskog učenja, sa ciljem automatske analize, indeksiranja, pretrage i preporuke muzike bazirane na različitim atributima. Ovo polje na popularnosti dobija razvojem računara krajem prošlog veka i ranih 2000-ih. U ovom poglavlju biće opisana rešenja relevantna za klasifikaciju sentimenata muzičkih numera.

U 2000-im godinama klasifikacija sentimenta pesama svodila se isključivo na kreiranje audio karakteristika koje su dovođene na klasične klasifikatore mašinskog učenja (SVM, Decision trees, itd.). Porast primene dubokih neuronskih mreža za analizu teksta i klasifikaciju u periodu nakon 2010. omogućio je uključivanje tekstova pesama bogatih informacijama u klasifikaciju sentimenta pesama, značajno poboljšavajući dotadašnje rezultate.

Autor Kano u svojoj doktorskoj disertaciji opisuje više aspekata muzičkih preporuka baziranih na sentimentu [1]. U radu su opisani različiti pristupi kreiranja sistema preporuka koji prate raspoloženje slušalaca (engl. Mood-Aware Music Recommenders). Takođe su istraženi različiti načini reprezentacije reči (Bag-Of-Words CBOW, Skip-Gram i Glove). Pokazano je da tačnost u analizi sentimenta korišćenjem različitih pretreniranih modela zavisi od prirode teksta koji se klasifikuje. Za klasifikaciju tekstova pesama najbolje pokazao GloVe model treniran nad twitter korpusom, iskorišćen u ovom radu, samo nad drugačijim skupom podataka.

U radu [2] takođe su upotrebljeni različiti pristupi dobijanja vektora reči. Prvo su implementirani tzv. baseline modeli: bag-of-words i TF-IDF. Sledeći pristup bio je da se vektori reči dobiju pomoću word2vec modela koji je treniran na tekstovima pesama, zatim pomoću samo google-news-300 modela, da bi na kraju bio iskorišćen google-news-300 model dotreniran tekstovima iz korpusa. Dobijeni vektori pojedinačnih reči su usrednjeni kako bi se dobio vektor celog teksta.

U trećem pristupu iskorišćena je LSTM neuronska mreža za vektorizaciju. Najbolje rezultate postigao je word2vec model treniran nad korpusom, koji je dao bolje rezultate čak i od LSTM modela. Pretpostavka autora je da je do iznenađujućeg rezultata došlo zato što je word2vec model treniran nad veoma velikim korpusom, te je uspeo da napravi veoma dobre reprezentacije reči.

3. METODOLOGIJA

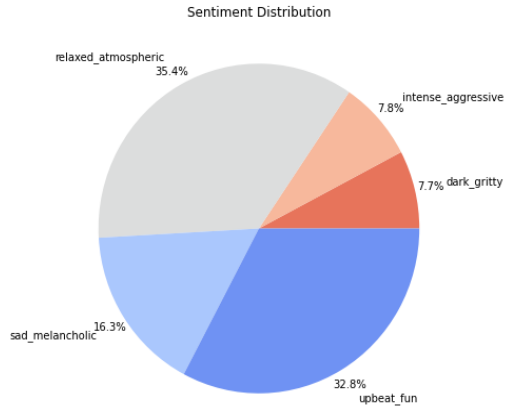
Klasifikacija sentimenta pesme u ovom radu zasniva se na analizi audio karakteristika pesama (tabelarni podaci) i tekstova pesama (prirodni jezik). Pod pretpostavkom da kombinovanje ovih podataka daje širi kontekst o svakoj pesmi, ulazni podaci iskorišćeni su za predviđanje jedne od pet klasa ciljnog obeležja, čineći problem koji se rešava klasifikacijom sa više klasa.

3.1. Skup podataka

Skup podataka korišćen za treniranje i testiranje modela stvoren je za potrebe ovog rada. Rezultirajući skup podataka nastao je kombinacijom dva javno dostupna skupa podataka o muzičkim numerama sa sajta Kaggle i prikupljanjem tekstova pesama (engl. web scraping) sa sajta genius.com.

Prvi skup podataka preuzet sa sajta Kaggle nosi naziv „MuSe: The Musical Sentiment Dataset“ i bio je ključan za ovaj rad jer u sebi sadrži obeležje seeds, koje predstavlja listu osećanja koju svaka pesma nosi. Iako tagovi sami po sebi predstavljaju sentiment, njihov broj bio je preveliki da bi klasifikacija imala smisla. Stoga su, na osnovu analize svih prisutnih osećanja, tagovi grupisani u pet semantičkih grupa: Relaxed/Atmospheric Upbeat/Fun, Intense/Aggressive, Dark/Gritty i Sad/Melancholic, gde je svakoj pesmi dodeljena grupa iz koje takvi tagovi preovlađuju.

Zastupljenost pesama koje nose određeni sentiment prikazuje Slika 1, koja ukazuje da je skup podataka nebalansiran. Ovo značajno utiče na sam proces treninga modela, ali i izbor adekvatnih metrika za evaluaciju modela.



Slika 1. *Količinski udeo pesama svakog sentimenta*

Drugi skup podataka preuzet sa Kaggle-a nosi naziv „Spotify Tracks DB“ i u najvećoj meri nosi informacije o audio karakteristikama pesama. Pored naziva pesama i izvođača, prisutna je i lista žanrova kojima pesma pripada. Kategorički podaci o pesmama, poput moda (durska, molska lestvica) i takta (2/4, 3/4, 4/4, 5/4) kodirani su uz pomoć One Hot Encoding metode. Karakteristike pesama poput akustičnosti, mogućnosti da se pleše uz pesmu (danceability), energije, prisutnosti teksta (speechiness), pozitivnosti pesme (valence) nalazile su se u rasponu od 0-1, dok su tempo pesme (BPM – beats per minute) i glasnoća skalirani na ovaj raspon putem min-maks skaliranja.

Nakon prečišćavanja ova dva skupa podataka, izvršeno je spajanje na osnovu spotify_id kolone. Rezultirajući skup podataka ima približno 6500 redova.

Kako nijedan od pomenutih skupova podataka nije sadržao tekstove pesama, za spojeni skup podataka izvršeno je prikupljanje tekstova (engl. web scraping) sa popularnog sajta za tekstove pesama genius.com. Skripta za prikupljanje tekstova pesama je na osnovu imena izvođača i naziva pesme, oslanjajući se na Google search API, pronašla linkove do stranica tekstova datih pesama na ciljanom sajtu. Zatim su tekstovi pesama izvučeni uz pomoć biblioteke BeautifulSoup. Prikupljanje teksta bilo je uspešno za oko 5500 pesama (od 6500 iz prethodno spojenog skupa podataka), pa su ostale pesme odbačene zbog nemogućnosti da se nad njima primene modeli procesiranja prirodnog jezika. Nakon kreiranja skupa podataka i njegove podele na obučavajući, validacioni i test skup, tablarne audio karakteristike i tekstualni podaci su pretprocesirani sa ciljem njihove pripreme za obučavanje modela i njihovu evaluaciju.

3.2. Klasičan NLP pristup

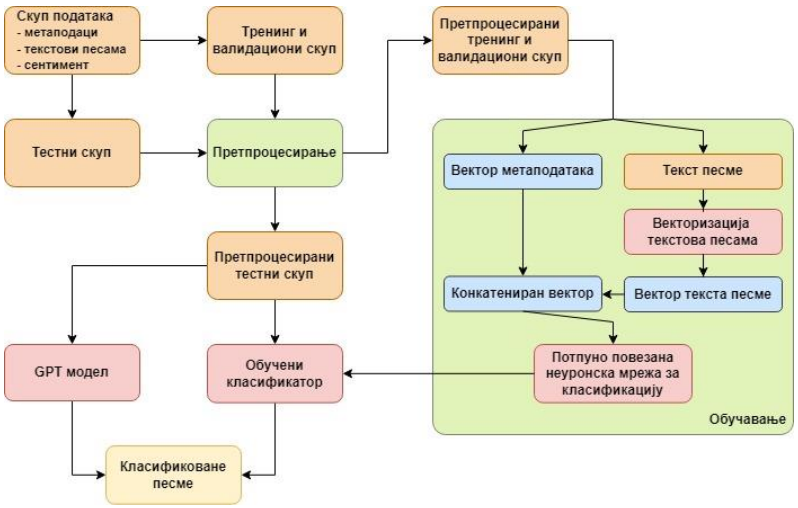
Izdvojeni pretprocesirani trening skup koristi se za obučavanje modela za klasifikaciju sentimenta pesama. Da bi se tekstualni podaci mogli proslediti neuronskoj mreži na klasifikaciju, neophodno ih je pretvoriti u vektore koji bi na odgovarajući način uhvatili semantičku vrednost teksta. Dobijeni vektor teksta pesme se potom konkatenira na vektor audio karakteristika odgovarajuće pesme i šalje na potpuno povezanu neuronsku mrežu koja se obučava za predviđanje sentimenta pesme (Slika 4).

Prvi pristup vektorizacije teksta koristi rekurentne neuronske mreže (RNN), tačnije LSTM/GRU ćelije. Od prerađenog teksta je najpre napravljen vokabular. Za svaku reč u sekvenci teksta, iz vokabulara se dobavlja njen indeks, koji se zatim šalje embedding sloju. Jedna po jedna embedovana reč se zatim šalje RNN sloju sa LSTM ili GRU ćelijama koji od pojedinačnih vektora reči stvara jedan vektor celokupne sekvence.

Drugi pristup podrazumeva upotrebu Word2Vec i GloVe modela gde je pojedinačna reč iz sekvence prvo prevedena u vektor da bi se zatim vektor sekvence dobio usrednjavanjem vektora svake pojedinačne reči. U ovom radu iskorišćene su sledeće modeli iz biblioteke gensim: word2vec-google-news-300, glove-wiki-gigaword-300, glove-twitter-200

Finalni pristup vektorizacije koristi BERT transformer. Sirov tekst svake pesme pretvoren je u vektor pomoću bert-base-uncased modela, implementacije BERT-a iz transformers biblioteke napravljene od strane kompanije Hugging Face.

Za obučavanje potpuno povezane neuronske mreže za klasifikaciju korišćen je optimizator ADAM. Kao funkcija greške iskorišćena je categorical crossentropy. Klasifikator je izgrađen kao kombinacija potpuno povezanih slojeva sa aktivacionom funkcijom ReLU i Dropout slojeva, sve u cilju sprečavanja pretreniranosti (overfitting). Izlazni sloj klasifikatora čini pet neurona (za pet klasa) sa Softmax aktivacionom funkcijom. Rezultat izlaznog sloja je vektor verovatnoća pripadnosti određenoj klasi, gde je predviđena klasa ona čija je verovatnoća najveća.



Slika 2. *Shema sistema implementiranog u radu*

3.3. Generativni pristup

Iako generativni modeli nisu pravljeni sa ciljem da se koriste kao klasifikatori, moguće je iskoristiti formu dijaloga za klasifikaciju. Princip na kojem počiva svaki GPT model jeste predviđanje najverovatnije sledeće reči na osnovu korisničkog unosa. Imajući to u vidu, jedan pristup klasifikaciji pomoću generativnog modela svodi se na sledeće korake:

1. Kroz sistemsku poruku pružiti modelu kontekst zadatka koji treba da izvrši, što obuhvata:
   - Koji atributi i u kom formatu će biti prosleđeni modelu za svaki test primer i koje je značenje tih atributa i njihovih vrednosti.
   - Koje su ponuđene klase koje model može da odabere pri klasifikaciji i šta predstavlja svaka klasa.
   - Bilo koja dodatna ograničenja vezana za konkretni klasifikacioni problem
   - U slučaju predviđanja sa primerom (engl. few-shot learning), modelu treba proslediti i reprezentativne primere svake od klasa.
2. Pružiti test primere koji treba da se klasifikuju.
3. Prikupiti i prečistiti odgovore koje je model dao.

Na osnovu gorenavedenog radnog okvira za pristupe klasifikaciji generativnim modelima, implementirano je nekoliko pristupa.

Klasifikacija bez primera (engl. zero-shot learning) kod generativnih modela se svodi na prosleđivanje samo pravila koja model mora da poštuje pri davanju odgovora i informacije o pesmi. Za ovaj pristup implementirane su dve verzije.

Prva verzija predstavlja najjednostavniju, tzv. „naivnu“ implementaciju klasifikacije generativnim modelima. Sistemska poruka sadrži slabo strukturirane rečenice koje nisu sažete i direktne, što otežava modelu da ih shvati i ispoštuje. Modelu su pri klasifikaciji u slobodnoj formi prosleđeni naziv izvođača pesme i tekst pesme, bez ostalih dostupnih audio karakteristika. Iako ovaj pristup jeste „naivan“, troši najmanje tokena, a samim tim i novca pri pozivima OpenAI API-ja i predstavlja referentnu tačku za ostale modele.

U drugoj tzv. „deskriptivnoj“ verziji predviđanja bez primera, pored teksta i osnovnih informacija o pesmi prosleđeni su ostali bitni metapodaci. Ovaj pristup podrazumeva znatno direktniju i precizno strukturiranu sistemsku poruku, u kojoj je tačno opisan format u kojem će model primiti informacije o pesmi. Odabrani format za opisivanje pesme je JSON objekat, gde je u sistemskoj poruci opisano koje tačno atribute model može da primi i koje je njihovo značenje. Zatim bi se za svaku pesmu modelu prosleđivalo pitanje u zadatom JSON formatu. Bitan detalj je da se tekst pitanja u ovoj verziji završava sa "Sentiment is: ", odnosno započinjanjem odgovora koji treba da finalizuje sam model. Na ovaj način, model je dodatno naveden da svoj odgovor ograniči samo na ime klase koju smatra pogodnom za datu pesmu u odnosu na prethodnu verziju, gde je često znatno divergirao od instrukcija.

Predviđanje sa primerima (eng. few-shot learning) kod klasifikacije generativnim modelom svodi se na to da se modelu nakon sistemske poruke proslede pesme koje najbolje predstavljaju date klase. Informacije svake ove reprezentativne pesme se dodaju u istoriju "dopisivanja" u istom formatu pitanja. Odmah nakon svake pesme se dodaje odgovor modela - klasa koja je unapred poznata - u formi poruke sa ulogom "assistant". Nakon ovih primera, modelu se šalje pitanje za konkretnu pesmu koju treba da klasifikuje i beleži odgovor.

Prva verzija ovog pristupa, one-shot learning, implementirana je tako što su za svaku od klasa odabrani reprezentativni primeri, dok je druga verzija, odnosno two-shot learning, implementirana tako što su, zbog ograničenja koje su unosili troškovi poziva OpenAI API-ja, primeri ograničeni samo na dve klase: relaxed_atmospheric i sad_melancholic, koje su se ispostavile problematičnim zbog sličnosti određenih aspekata pesama koje ih predstavljaju.

4. REZULTATI I DISKUSIJA

Finalna evaluacija je izvršena nad istim test skupom za sve pristupe. Za poređenje su iskorišćene makro usrednjene vrednosti preciznosti, odziva i F1 mere (sve klase podjednako bitne). Zbog nebalansiranosti podataka tačnost nije uzeta u obzir.

Modeli su optimizovani tako da se maksimizuje makro usrednjena preciznost uz obrazloženje da je minimizacija

lažno pozitivnih primera bitno u scenariju korišćenja klasifikovanih pesama za preporuke muzike spram raspoloženja. Ukoliko korisnik jednog takvog sistema odluči da mu se sluša srećna muzika kako bi se oraspoložio, problem bi nastao kada bi sistem ponudio tužnu pesmu koja je označena kao srećna. Pretpostavka u ovom radu je da takva situacija predstavlja mnogo veći problem od toga da sistem bude više rigorozan, te da za pesmu koja zaista jeste srećna kaže da nije srećna ako nije skroz siguran (manje je bitna minimizacija lažnih negativnih, odnosno maksimizacija odziva).

4.1. Poređenje klasičnih NLP pristupa

Rezultate makro usrednjenih vrednosti metrika za klasične NLP pristupe prikazuje Tabela 1. Najveću makro usrednjenu preciznost od 0,68 postigao je pristup koji je za vektorizaciju teksta iskoristio **word2vec model treniran nad Google News korpusom.**

Tabela 1. *Makro prosek mera za sve NLP pristupe*

Макро просек			
Приступ	Прецизност	Одзив	F1 мера
LSTM	0.58	**0.52**	**0.54**
GRU	0.55	0.52	0.53
word2vec-google-news-300	**0.68**	0.44	0.42
glove-wiki-gigaword-300	0.63	0.46	0.49
glove-twitter-200	0.66	0.4	0.4
BERT	0.65	0.41	0.42

Iako word2vec nije treniran nad muzičkim korpusom, pokazao je značajno bolje rezultate od RNN pristupa, što je u skladu s rezultatima iz [2].

RNN su imale lošije rezultate od ostalih pristupa uz pretpostavku da je skup za obučavanje bio mali. RNN bi potencijalno dale bolje rezultate da vokabular nije građen od skupa tekstova pesama već preuzet iz kolokvijalnih word2vec-google-news ili glove-twitter modela treniranih nad mnogo većim korpusom.

Model koji koristi BERT za vektorizaciju se po metrikama iznenađujuće nalazi između RNN i Word2Vec/GloVe, iako je među modelima s najbogatijom reprezentacijom reči. Rezultati s ovim enkoderom potencijalno bi se poboljšali kada bi se dotrenirao nad kolokvijalnim tekstovima pesama.

4.2. Poređenje generativnih pristupa

Rezultate makro usrednjenih vrednosti metrika za generativne pristupe prikazuje Tabela 2. Najveću makro usrednjenu preciznost postigao je „deskriptivni“ pristup bez primera koji prikazuje skok od čak 20% u makro usrednjenoj preciznosti (0,63), što ga čini malo lošijim od BERT pristupa (0,65).

Tabela 2. *Makro prosek mera za sve generativne pristupe*

Макро просек			
Приступ	Прецизност	Одзив	F1 мера
Naive Zero Shot	0.43	0.41	0.41
Descriptive Zero Shot	**0.63**	**0.57**	**0.57**
One shot	0.44	0.25	0.27
Two shots	0.39	0.34	0.33

Pretpostavka jeste da je značajno poboljšanje posledica jasnih i direktnih instrukcija, kao i strukturiranih informacija o tekstu i audio karakteristikama pesama.

Makro usrednjena preciznost generativnog pristupa sa jednim primerom po klasi, iako mnogo skuplja, jedva je bolja od "naivnog" pristupa bez primera, što je suprotno inicijalnoj pretpostavci da će modeli tačnije odrediti klasu pesme ukoliko mu se prikažu primeri.

U skladu sa ovim, generativni pristup sa dva primera je dodatno pogoršao celokupni rezultat, gde je makro usrednjena preciznost najgora od svih pristupa i iznosi 0,39. Još jedna začuđujuća stavka kod ovog pristupa jeste da je za dosta pesama iz test skupa model odlučio da pripadaju "upbeat_fun" klasi, o čemu svedoči mala preciznost od svega 0,5 i izuzetno visok odziv od 0,94.

5. ZAKLJUČAK

U ovom radu pokazano je da je, iako generativni modeli nisu pravljeni sa klasifikacijom u cilju, moguće napraviti klasifikator koji ima slične performanse kao modeli koji su za to pogodniji.

Pretpostavka je da primena generativnih modela za klasifikaciju nad malim skupovima podataka može pokazati dobre rezultate zbog ogromnog opšteg znanja koje ovi modeli poseduju. Međutim i skup od oko 5000 redova podataka pokazao je kako su klasični NLP pristupi moćniji kada postoji dovoljno podataka nad kojima mogu da se istreniraju.

Dalji razvoj rešenja išao bi u pravcu prikupljanja većeg skupa podataka, kao i kombinacije pristupa opisanih u ovom radu (LSTM koji koristi Word2Vec ili GloVe vektore za širi, kolokvijalni vokabular), ili dotreniranja BERT modela nad tekstovima pesama. Pristup sa generativnim modelom koji dobija primere pesama mogao bi se poboljšati sa drugačijim odabirom reprezentativnih pesama i slanjem većeg broja primera svake klase, što bi dodatno povećalo trošak ovog pristupa. Bolje rezultate bi potencijalno donela i drugačija heuristika za kreiranje ciljnog obeležja od tagova sentimenta.

6. LITERATURA

[1] E. Çano, “Text-based Sentiment Analysis and Music Emotion Recognition,” arXiv:1810.03031 [cs], Jun. 2018, doi: https://doi.org/10.6092/polito/porto/2709436.

[2] M. McVicar, B. Di Giorgi, B. Dundar, and M. Mauch, “Lyric document embeddings for music tagging,” arXiv.org, Nov. 29, 2021. https://arxiv.org/abs/2112.11436 (accessed Nov. 07, 2023).

Kratka biografija:

Žarko Blagojević rođen je u Sremskoj Mitrovici 2000. god. Master rad na Fakultetu tehničkih nauka iz oblasti Računarska inteligencija – Softversko inženjerstvo i informacione tehnologije odbranio je 2023. godine.

Kontakt: zarexblage00@gmail.com