Multimodalna analiza sentimenta korištenjem teksta i slike

Krog, Tomislav

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 352

MULTIMODALNA ANALIZA SENTIMENTA KORIŠTENJEM TEKSTA I SLIKE

Tomislav Krog

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 352

MULTIMODALNA ANALIZA SENTIMENTA KORIŠTENJEM TEKSTA I SLIKE

Tomislav Krog

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 4. ožujka 2024.

DIPLOMSKI ZADATAK br. 352

Pristupnik:	**Tomislav Krog (0036519636)**
Studij:	Računarstvo
Profil:	Znanost o podacima
Mentor:	doc. dr. sc. Marko Đurasević

Zadatak: **Multimodalna analiza sentimenta korištenjem teksta i slike**

Opis zadatka:

Tema rada usmjerena je na razvoj i primjenu multimodalne analize sentimenta, specifično na tekstu i slikama. Inicijalna analiza provodit će se na slikama memeova. Memeovi, kao popularan oblik digitalne komunikacije, često prenose složene, suptilne i višeznačne poruke koje mogu biti izazov za analizu korištenjem tradicionalnih metoda fokusiranih samo na tekst ili samo na sliku. Cilj rada je razviti metodologiju koja integrira napredne tehnike obrade prirodnog jezika i računalnog vida za razumijevanje sentimenta izraženog u memeovima. Osnovni koraci uključivat će prikupljanje i pripremu podataka, predobradu i analizu teksta, analizu slika, integraciju i multimodalnu analizu te u konačnici validacija i evaluacija. Diplomski rad istražit će kako kombinacija tekstualnih i vizualnih signala omogućava dublje razumijevanje sentimenta, ironije, satire i humorističnih elemenata često prisutnih u memeovima. Radu priložiti izvorne tekstove programa, dobivene rezultate uz potrebna objašnjenja i korištenu literaturu.

Rok za predaju rada: 28. lipnja 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 352

MULTIMODALNA ANALIZA SENTIMENTA KORIŠTENJEM TEKSTA I SLIKE

Tomislav Krog

Zagreb, lipanj, 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 4. ožujka 2024.

DIPLOMSKI ZADATAK br. 352

Pristupnik:	**Tomislav Krog (0036519636)**
Studij:	Računarstvo
Profil:	Znanost o podacima
Mentor:	doc. dr. sc. Marko Đurasević

Zadatak: **Multimodalna analiza sentimenta korištenjem teksta i slike**

Opis zadatka:

Tema rada usmjerena je na razvoj i primjenu multimodalne analize sentimenta, specifično na tekstu i slikama. Inicijalna analiza provodit će se na slikama memeova. Memeovi, kao popularan oblik digitalne komunikacije, često prenose složene, suptilne i višeznačne poruke koje mogu biti izazov za analizu korištenjem tradicionalnih metoda fokusiranih samo na tekst ili samo na sliku. Cilj rada je razviti metodologiju koja integrira napredne tehnike obrade prirodnog jezika i računalnog vida za razumijevanje sentimenta izraženog u memeovima. Osnovni koraci uključivat će prikupljanje i pripremu podataka, predobradu i analizu teksta, analizu slika, integraciju i multimodalnu analizu te u konačnici validacija i evaluacija. Diplomski rad istražit će kako kombinacija tekstualnih i vizualnih signala omogućava dublje razumijevanje sentimenta, ironije, satire i humorističnih elemenata često prisutnih u memeovima. Radu priložiti izvorne tekstove programa, dobivene rezultate uz potrebna objašnjenja i korištenu literaturu.

Rok za predaju rada: 28. lipnja 2024.

Ovim putem želim izraziti svoju duboku zahvalnost svim osobama koje su me podržavale i pomagale mi tijekom izrade ovog diplomskog rada, kao i tijekom cijelog života.

Prije svega, veliko hvala mom mentoru, prof. dr. sc. Marku Đuraseviću, na prihvaćanju teme rada, pomoći tijekom studija te beskrajnom strpljenju.

Roditeljima, Josipu i Marini, te sestrama Martini, Danijeli i Mirni neizmjerno hvala na njihovoj podršci i razumijevanju. Znam da nije bilo lako, ali vaša vjera u mene učinila je sve ovo mogućim.

Posebna zahvala mojoj mami Marini, čija me neizmjerna ljubav i nesebična podrška uvijek gurala naprijed. Hvala ti što si uvijek bila tu za mene.

Baki Pepi, hvala na svim obrocima i nebrojenim kavama koje su me držale budnim kroz sve one neprospavane noći.

Mojoj djevojci Ivani, hvala što si mi uljepšala život i tolerirala sve moje trenutke stresa i opterećenosti. Tvoja ljubav i podrška znače mi više nego što riječi mogu izraziti.

I naravno, svim mojim prijateljima – hvala što ste me držali prizemljenim, nasmijavali me i pružali bijeg od svijeta znanosti kad god je bilo potrebno. Bez vas, ovaj put bi bio puno teži i manje zabavan.

Hvala svima od srca. Bez vas, ovaj rad ne bi bio moguć, a ja ne bih bio čovjek koji sam danas!

Sadržaj

1. Uvod

1.1. Kontekst i značaj istraživanja

U današnjem digitalnom dobu, internetski sadržaj postaje sve više vizualan, a popularni formati poput memeova igraju ključnu ulogu u komunikaciji na društvenim mrežama. Memeovi su jedinstveni jer kombiniraju tekst i sliku kako bi prenijeli poruke, emocije i humoristične sadržaje. Kombinacija teksta i slike čini ih zanimljivim predmetom za analizu sentimenta, što je postupak razumijevanja emocionalnog tona iza pisanih i vizualnih sadržaja. Multimodalna analiza sentimenta, koja uključuje oba modaliteta - tekstualni i vizualni, predstavlja napredak u analizi osjećaja jer omogućuje dublje i preciznije razumijevanje složenih poruka koje se prenose putem memeova.

Istraživanje u ovoj oblasti ima značaj ne samo za akademsku zajednicu već i za praktične primjene u marketingu, analizi društvenih mreža, detekciji dezinformacija i razumijevanju javnog mnijenja. Razvoj metoda koje mogu učinkovito analizirati sentiment memeova može pomoći tvrtkama i organizacijama da bolje razumiju reakcije korisnika i prilagode svoje strategije u skladu s tim.

1.2. Ciljevi rada

Glavni cilj ovog rada istražiti je i razviti metode za multimodalnu analizu sentimenta koristeći tekst i slike u memeovima. Specifični ciljevi uključuju:

- Identificirati i pregledati postojeće pristupe sentiment analizi za tekstualne i vizualne podatke.
- Razviti metodologiju za analizu memeova kao multimodalnih podataka.

- Implementirati modele za analizu sentimenta koji kombiniraju tekstualne i vizualne informacije.
- Evaluirati performanse razvijenih modela na odabranom skupu podataka.
- Diskutirati rezultate i identificirati potencijalna poboljšanja i buduće smjerove istraživanja.

2. Pregled područja

U radu Sharma et al. (2020) [1] opisuje se model dubokog učenja korišten za zadatak Memotion analize na SemEval natjecanju. Predložena arhitektura koristi prijenosno učenje (*eng. transfer learning*) za ekstrakciju značajki iz slika i teksta, koje se potom spajaju koristeći LSTM i GRU modele s mehanizmom pažnje za konačne predikcije.

Wang et al. (2023) [2] istražuju sposobnosti GPT-3.5 modela u obradi sentimenta internet memeova. Istraživanje je otkrilo snage i ograničenja GPT-3.5 modela u subjektivnim zadacima kao što su klasifikacija sentimenta memeova, određivanje tipa humora i detekcija implicitne mržnje. Rezultati pokazuju da, iako GPT-3.5 ima značajan potencijal, suočava se s izazovima u tumačenju konteksta i implicitnih značenja u memeovima.

U svom radu Bucur et al. (2022) [3] predstavljaju dva pristupa za klasifikaciju memova. Prvi pristup temelji se isključivo na tekstualnoj metodi koristeći BERT, dok drugi pristup koristi Multi-Modal-Multi-Task transformer mrežu koja obrađuje sliku mema i njegov OCR za konačnu klasifikaciju. U oba pristupa autori koriste predtrenirane modele za obradu teksta (BERT, Sentence Transformer) i slike (EfficientNetV4, CLIP).

Pramanick et al. (2021) [4] predlažu višeslojni okvir neuronske mreže zasnovan na mehanizmu pažnje, nazvan MHA-MEME, čiji je glavni cilj iskoristiti prostornu korespondenciju između vizualne modalnosti (slike) i različitih tekstualnih segmenata kako bi se izvukle detaljne značajke za klasifikaciju. Evaluiraju MHA-MEME na Memotion Analysis skupu podataka. Njihova komparativna studija pokazuje da MHA-MEME postiže *state-of-the-art* performanse. Osim toga, potvrđuju generalizaciju MHA-MEME na drugom skupu ručno anotiranih testnih uzoraka i opažaju konzistentnost predikcija.

Xuan et al. (2024) [5] u svom su radu istražili potencijal velikih vizualno-jezičnih modela (LVLM) za detekciju multimodalnih dezinformacija. Otkrili su da, iako LVLM ima

superiorne performanse u usporedbi s LLM-ovima, duboko zaključivanje može pokazati ograničenu snagu bez prisutnosti dokaza. Na temelju tih zapažanja, autori predlažu LEMMA: LVLM-poboljšanu detekciju multimodalnih dezinformacija uz augmentaciju vanjskim znanjem, koja poboljšava točnost detekcije dezinformacija za 7% na Twitter i 13% na Fakeddit skupovima podataka.

3. Teorijska pozadina

3.1. Emocije

3.1.1. Definicija i vrste emocija

Emocija je subjektivna reakcija organizma na vanjske podražaje [6]. Emocija obuhvaća niz fizioloških, kognitivnih i ponašajnih odgovora. Fiziološki odgovori mogu uključivati promjene u otkucajima srca, disanju i hormonalnim razinama. Kognitivni odgovori odnose se na procjenu situacije koja izaziva emociju, dok ponašajni odgovori mogu uključivati izraze lica, gestikulaciju i druge oblike neverbalne komunikacije. Emocije igraju ključnu ulogu u donošenju odluka, socijalnoj interakciji i prilagodbi na okolinu. Također su povezane s mentalnim zdravljem, a njihovo razumijevanje i regulacija važni su za opće blagostanje. Emocije igraju ključnu ulogu u ljudskoj komunikaciji i razumijevanju. One oblikuju naše interakcije, odluke i percepcije. Prema Plutchiku (1980) [7], postoji osam osnovnih emocija: radost, povjerenje, strah, iznenađenje, tuga, očekivanje, ljutnja i gađenje . Svaka od ovih osnovnih emocija može se dalje kombinirati i stvarati složenije emocije.

3.1.2. Teorije emocija

Postoji nekoliko ključnih teorija koje pokušavaju objasniti kako i zašto doživljavamo emocije:

- **Darwinova teorija emocija** sugerira da su emocije evolucijski prilagođene i da služe kao način komunikacije unutar vrste.
- **James-Langeova teorija** tvrdi da emocije proizlaze iz naših fizioloških reakcija na vanjske podražaje.

- **Cannon-Bardova teorija** predlaže da emocionalni i fiziološki odgovori na podražaj nastaju istovremeno, a ne uzročno-posljedično.
- **Schachter-Singerova teorija** sugerira da emocije nastaju iz kombinacije fiziološke pobude i kognitivne interpretacije te pobude.

3.1.3. Emocije u digitalnoj komunikaciji

S porastom digitalne komunikacije, istraživanje emocija u ovom kontekstu postaje sve važnije. Emocije igraju ključnu ulogu u načinu na koji komuniciramo putem digitalnih medija, uključujući društvene mreže, e-poštu i chat aplikacije. Emotikoni, GIF-ovi i memeovi postali su važan alat za izražavanje emocija u digitalnom svijetu . Istraživanja pokazuju da emocije u digitalnoj komunikaciji mogu biti intenzivnije zbog nedostatka neverbalnih znakova, što može dovesti do nesporazuma ili pojačanja emocionalnog izraza. Osim toga, algoritmi većine društvenih mreža često favoriziraju sadržaj koji izaziva snažne emocionalne reakcije, što može utjecati na naše emocionalno stanje i percepciju svijeta oko nas.

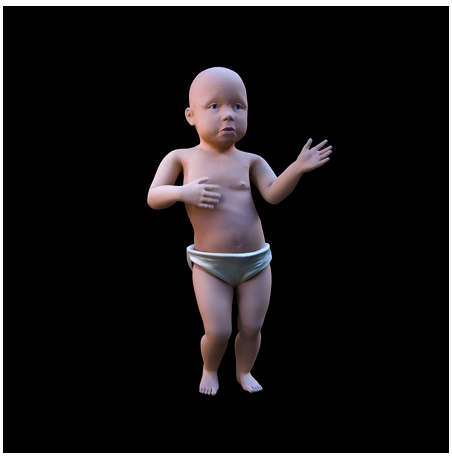
3.2. Memeovi

3.2.1. Definicija i povijest memeova

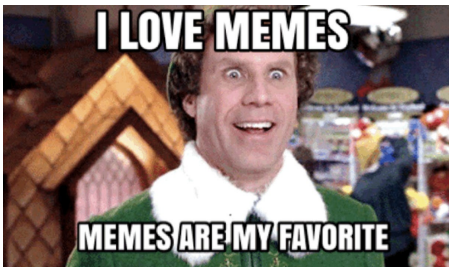
Meme je pristup, koncept, ideja ili stil koji se širi putem društvenih mreža unutar društva, često s ciljem izražavanja trenda, teme ili značaja kojeg sam meme predstavlja [8]. Oni mogu biti u obliku slika, videa, teksta ili kombinacije ovih elemenata. Memeovi su specifični po svojoj sposobnosti da prenesu složene ideje, emocije ili humoristične sadržaje na sažet način. Popularnost memeova leži u njihovoj jednostavnosti, univerzalnosti i sposobnosti da brzo postanu viralni. Zbog svoje prirode, memeovi su idealni za proučavanje multimodalne analize sentimenta jer objedinjuju tekstualne i vizualne elemente, stvarajući bogat kontekst za analizu. Memeovi, kao specifičan oblik komunikacije, zahtijevaju poseban pristup u analizi sentimenta. Kombinacija tekstualnih i vizualnih podataka u memeovima stvara složen kontekst koji može preciznije odražavati emocije i stavove korisnika. Multimodalna analiza sentimenta omogućava bolje razumijevanje ovih kompleksnih izraza, pružajući dublji uvid u kolektivne i individualne

emocionalne reakcije.

Pojam "meme" prvi put je uveo Richard Dawkins u svojoj knjizi Sebični gen (*eng. The Selfish Gene*) iz 1976. godine [9]. Dawkins je definirao memeove kao kulturne ekvivalente gena, koji se repliciraju i evoluiraju putem kulturne transmisije. Digitalni memeovi, kakve danas poznajemo, pojavili su se s razvojem interneta i društvenih mreža. Ovi memeovi često kombiniraju tekst i sliku te se brzo šire putem internetskih platformi, stvarajući nove oblike online komunikacije i humora. Jedan od prvih poznatih internetskih memeova je "Dancing Baby" iz 1996. godine (slika 3.1.), koji je postao viralan putem e-maila i web stranica.



Slika 3.1. Dancing Baby (1996)



Slika 3.2. Primjer memea

3.2.2. Klasifikacija i vrste memeova

Memeovi se mogu klasificirati prema različitim kriterijima, uključujući oblik, sadržaj i namjenu. Neke od glavnih vrsta memeova uključuju:

1. **Image Macro Memes**: Najčešći oblik memeova, koji kombiniraju sliku s tekstom. Primjeri uključuju "Bad Luck Brian" i "Success Kid".

2. **Video Memes**: Kratki video isječci koji se dijele i prepravljaju radi humora ili društvenih komentara. Primjeri uključuju "Rickrolling" i "Gangnam Style".

3. **GIF Memes**: Animirane slike koje se koriste za izražavanje emocija ili reakcija. Primjeri uključuju "Blinking Guy" i "Crying Jordan".

4. **Hashtag Memes**: Memeovi koji se šire putem specifičnih hashtagova na društvenim mrežama. Primjeri uključuju #IceBucketChallenge i #MannequinChallenge".

3.2.3. Utjecaj memeova na društvene mreže

Memeovi imaju značajan utjecaj na društvene mreže, oblikujući načine na koji ljudi komuniciraju, dijele informacije i izražavaju se. Neki od ključnih aspekata utjecaja memeova na društvene mreže uključuju:

- **Viralnost**: Memeovi se brzo šire internetom, često postajući viralni u vrlo kratkom vremenu. Ovo omogućuje brzu distribuciju ideja i kulturoloških referenci.

- **Kreiranje zajednica**: Memeovi često okupljaju ljude sličnih interesa i humora, stvarajući online zajednice koje dijele i kreiraju nove memeove.

- **Politički i društveni utjecaj**: Memeovi se koriste za izražavanje političkih stavova, društvenih komentara i mobilizaciju aktivizma. Primjeri uključuju memeove povezane s pokretima kao što je npr. #BlackLivesMatter.

- **Komercijalna upotreba**: Tvrtke koriste memeove za marketing i komunikaciju s mlađom publikom, često kroz humor i relevantne kulturološke reference.

Memeovi također mogu utjecati na percepciju stvarnosti, oblikujući mišljenja i stavove korisnika društvenih mreža kroz humor i sarkazam.

3.3. Sentiment analiza

3.3.1. Definicija i važnost

Analiza sentimenta proces je korištenja prirodnog jezika i algoritama strojnog učenja za identificiranje i ekstrakciju subjektivnih informacija iz tekstualnih podataka. Cilj je

analize sentimenta odrediti emocionalni ton poruke te ga u najjednostavnijem slučaju svrstati u pozitivan, negativan ili neutralan [10]. Pojam "sentiment analiza" prvi put su spomenuli Nasukawa et. al. (2003) [11], dok se pojam "rudarenje mišljenja" prvi put pojavio u Dave et. al. (2003) [12].

3.3.2. Metode i pristupi sentiment analizi

Postoje tri glavna pristupa sentiment analizi:

1. **Pristup temeljen na pravilima**: Ovaj pristup koristi unaprijed definirane leksikone - skupove riječi s pridruženim sentimentima. Na primjer, riječi poput "sretan", "pristupačan" i "brz" dobivaju pozitivne bodove, dok se riječima poput "loš", "skup" i "težak" dodijeljuju negativni bodovi. Ovaj pristup jednostavan je za postavljanje, ali može biti teško skalabilan i često zahtijeva redovito održavanje leksikona [13].

2. **Strojno učenje**: Ovaj pristup koristi algoritme strojnog učenja za klasifikaciju tekstualnog sentimenta. Model se trenira na velikom skupu podataka kako bi mogao prepoznati emocionalni ton u novim, do sad neviđenim podacima. Prednost ovog pristupa njegova je sposobnost da precizno obrađuje širok raspon tekstualnih informacija, ali modeli često zahtijevaju ponovno treniranje kada se primjenjuju na različite domene [13].

3. **Hibridni pristup**: Kombinira prednosti oba pristupa - pravila i strojnog učenja - kako bi optimizirao brzinu i točnost analize. Iako ovaj pristup pruža najbolje rezultate, zahtijeva više resursa, vremena, tehničkih sposobnosti kao i domenskog znanja [13].

3.4. Multimodalna analiza

3.4.1. Definicija i izazovi

Multimodalna analiza odnosi se na analizu podataka koji uključuju više od jednog modaliteta, kao što su tekst i slike. U kontekstu analize sentimenta, to znači integriranje tekstualnih i vizualnih informacija kako bi se dobio sveobuhvatan uvid u emocionalni

ton poruke. Glavni izazovi u multimodalnoj analizi uključuju složenost integracije različitih vrsta podataka, potrebu za naprednim algoritmima koji mogu obraditi te podatke te potreba za velikim računalnim resursima.

3.4.2. Integracija teksta i slike

Integracija teksta i slike u analizi sentimenta uključuje nekoliko koraka:

1. **Ekstrakcija značajki**: Iz teksta se izvlače značajke kao što su ključne riječi ili fraze, dok se iz slika izdvajaju vizualne značajke poput boja, oblika i tekstura.

2. **Kombiniranje značajki**: Obje vrste značajki kombiniraju se (*eng. fusion*) na različite načine te u različitim fazama. Fuzija se može izvoditi kao fuzija značajki (*eng. feature-based multimodal fusion*) u ranoj fazi, fuzija temeljena na modelu (*eng. model-based multimodal fusion*) u srednjoj fazi, te fuzija temeljena na odlukama (*eng. decision-based multimodal fusion*) u kasnoj fazi.

3. **Modeliranje i analiza**: Kombinirane značajke zatim se koriste za treniranje modela koji može predvidjeti sentiment na temelju obje vrste podataka.

Ova integracija omogućuje preciznije razumijevanje složenih poruka koje sadrže tekstualne i vizualne elemente, što je posebno korisno u analizi memeova.

4. Metodologija

4.1. Opis skupa podataka

4.1.1. Izvor i karakteristike skupa podataka

Skup podataka korišten u ovom istraživanju je "Memotion Dataset", dostupan na platformi Kaggle [14]. Memotion Dataset skup je podataka koji se koristio za potrebe natjecanja "Memotion Analysis". Natjecanje se sastojalo od tri zadatka za detekciju emocija u memeovima, kako je opisano u nastavku:

- **Task A: Sentiment Analysis**: Klasifikacija memea u jednu od tri klase: negativan, neutralan, pozitivan.
- **Task B: Emotion Classification**: Identifikacija emocije izražene u memeu: humor, sarkazam, uvredljivost i motivacija. Meme može izraziti više od jedne emocije.
- **Task C: Scales/Intensity of Emotion Classes**: Kvantifikacija intenziteta izražene emocije u memeu.

U ovom istraživanju fokusirali smo se isključivo na **Task A: Sentiment Analysis**, gdje je cilj identificirati da li je meme pozitivan, negativan ili neutralan. Skup podataka dizajniran za ovaj zadatak sadrži 7,000 memeova prikupljenih s različitih društvenih mreža. Memotion Dataset posebno je dizajniran za istraživanje multimodalne analize sentimenta i emocija u memeovima, kombinirajući tekstualne i vizualne elemente. Skup podataka sadrži slike memeova i odgovarajući tekst izdvojen OCR-om. Memeovi su anotirani pomoću Amazon Mechanical Turka. Dodatan zadatak anotatora bio je ispraviti OCR tekst kada je on bio netočan.

4.1.2. Struktura i format podataka

Skup podataka sadrži slike memeova te CSV datoteku pod nazivom `labels.csv`, koja sadrži sljedeće stupce:

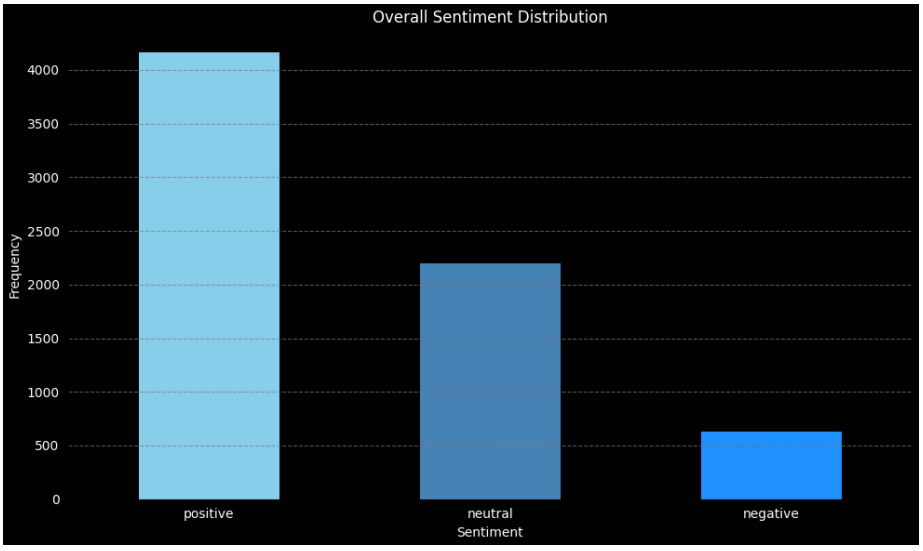
- **image_name**: Naziv slike memea.
- **text_ocr**: Tekstualni sadržaj ekstrahiran iz slike putem optičkog prepoznavanja znakova (OCR).
- **text_corrected**: Ručno ispravljeni tekstualni sadržaj od strane anotatora.
- **overall_sentiment**: Kategorija sentimenta za svaki meme.

Stupac **overall_sentiment** izvorno sadrži oznake *very_negative*, *negative*, *neutral*, *very_positive* te *positive*. Za potrebe analize, ove oznake su pojednostavljene i grupirane u tri kategorije kao što je bilo i predviđeno u originalnom zadatku natjecanja: *negative*, *neutral* i *positive*.

U tablici 4.1. kao i na grafu 4.1. prikazana je distribucija memeova prema kategoriji sentimenta.

Sveukupni sentiment	Distribucija oznaka (%)
Negative and Very Negative	13.9
Neutral	21.7
Positive and Very Positive	64.4

Tablica 4.1. Distribucija oznaka sentimenta

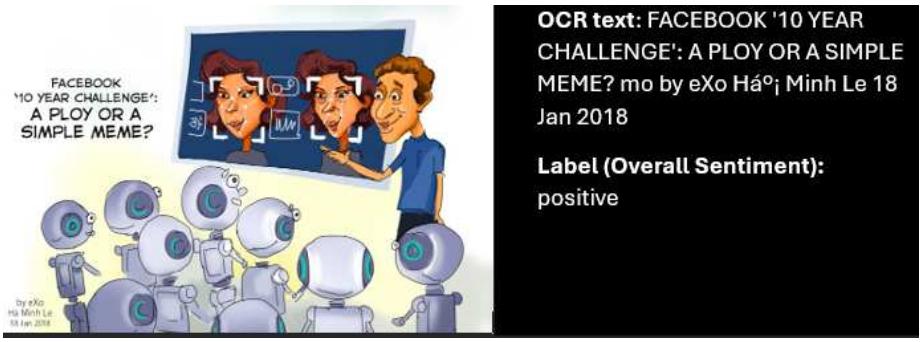


Slika 4.1. Distribucija oznaka sentimenta

4.1.3. Kompleksnost memeova iz Memotion Dataset-a

Memeovi iz "Memotion Dataset" skupa podataka iznimno su izazovni za prepoznavanje sentimenta, čak i za ljude. Kompleksnost dolazi iz kombinacije vizualnih i tekstualnih elemenata koji često sadrže sarkazam, ironiju i kulturološke reference koje je teško automatski obraditi. Dodatno, mnogi memeovi sami su po sebi nesmisleni ili smisleni samo rijetkim pojedincima, koji su možda kreatori tih memeova ili imaju svoju vrlo usku publiku. Ova izazovnost može se vidjeti u primjerima prvih dvaju memeova iz skupa podataka prikazanih na slikama 4.2. i 4.3.

Slika 4.2. Prvi primjer Memotion Dataset-a



Slika 4.3. Drugi primjer Memotion Dataset-a

4.2. Predprocesuiranje podataka

Predprocesuiranje podataka ključni je korak u pripremi podataka za analizu i modeliranje. U ovom istraživanju koristili smo različite tehnike predprocesuiranja kako bismo osigurali kvalitetu i konzistentnost podataka. Predprocesuiranje se provodilo na tekstualnim i vizualnim podacima.

4.2.1. Tekstualni podaci

Efikasna analiza tekstualnih podataka zahtijeva temeljito predprocesuiranje i čišćenje kako bi se osigurala njihova dosljednost i pravilno formatiranje. Proces čišćenja teksta obuhvaća nekoliko ključnih koraka koji su detaljno opisani u nastavku.

Čišćenje i predprocesuiranje tekstualnih podataka

1. **EMOTICONS rječnik**: EMOTICONS je rječnik koji sadrži parove ključ-vrijednost gdje su ključevi različiti emotikoni, a vrijednosti su opisi tih emotikona. Emotikoni su sekvence znakova koje ljudi koriste u digitalnoj komunikaciji za izražavanje emocija. Na primjer, emotikon ":-)" predstavlja "Happy face or smiley", dok emotikon ":-(" označava "Frown, sad, angry or pouting".

2. **Funkcija clean_text**: Funkcija `clean_text` koristi se za čišćenje i predprocesuiranje tekstualnog ulaza kako bi se tekst pripremio za dalju obradu.Funkcija obuhvaća nekoliko koraka:

    - **Pretvaranje u string i mala slova**: Ulazni tekst prvo se konvertira u string, a zatim u mala slova radi dosljednosti. Ovo omogućuje da pretrage ne budu osjetljive na velika i mala slova.

    - **Zamjena emotikona**: Funkcija prolazi kroz svaki emotikon u rječniku EMOTICONS i zamjenjuje ga njegovim tekstualnim opisom u ulaznom tekstu. Na primjer, ako tekst sadrži emotikon ":-)", on će biti zamijenjen s "Happy face or smiley". Ovo omogućuje pravilnu interpretaciju emotikona tokom analize teksta.

    - **Uklanjanje neželjenih znakova**: - Svi znakovi koji nisu slova, brojevi ili razmaci uklanjaju se korištenjem regularnih izraza. Ovo pomaže u čišćenju teksta od interpunkcije i specijalnih znakova koji nisu potrebni za analizu.

    - **Uklanjanje suvišnih razmaka**: Sve višestruke razmake funkcija kombinira u jedan te uklanja one nepotrebne. Ovo osigurava da tekst bude kompaktan bez nepotrebnih praznina koje bi mogle utjecati na analizu.

    - **Uklanjanje URL-ova**: Svi URL-ovi uklanjaju se iz teksta. URL-ovi često

ne nose relevantnu informacijsku vrijednost za većinu tekstualnih analiza te predstavljaju šum.

- **Uklanjanje oznaka korisnika i hashtagova**: Funkcija uklanja oznake korisnika (npr. @korisnik) i hashtagove (npr. #hashtag). Ove oznake specifične su za društvene mreže i obično nisu potrebne u analizi teksta.

- **Tokenizacija**: Nakon uklanjanja neželjenih elemenata, tekst se razbija na pojedinačne riječi ili tokene. Tokenizacija je proces dijeljenja teksta na manje jedinice (riječi), što omogućuje lakšu obradu i analizu teksta. Tokeni koji nisu sastavljeni od slova, kao što su brojevi i specijalni znakovi, eliminiraju se.

- **Stematizacija (*eng. stemming*)**: Svaka riječ svodi se na njen osnovni ili korijenski oblik pomoću Porterovog stemmera. Stematizacija uklanja sufikse iz riječi kako bi se dobio njen osnovni oblik, što pomaže u normalizaciji teksta i smanjenju broja različitih oblika iste riječi.

- **Rekonstruiranje teksta**: Na kraju, funkcija ponovo sastavlja očišćeni tekst iz tokena u jedan string, koji je spreman za daljnju analizu. Rezultat je očišćen, normaliziran i pripremljen tekst koji se može koristiti za razne analize u okviru obrade prirodnog jezika.

Ovi koraci predprocesuiranja osiguravaju da tekstualni podaci budu dosljedni, očišćeni i spremni za daljnju analizu, omogućujući preciznu analizu sentimenta i emocija u memeovima.

4.2.2. Vizualni podaci

Priprema slikovnih podataka za analizu i modeliranje zahtijeva uklanjanje oštećenih slika i primjenu odgovarajućih transformacija na slike koje će se koristiti. Proces predprocesuiranja slikovnih podataka sastoji se od dva ključna dijela: uklanjanje oštećenih slika iz skupa podataka i primjena niza transformacija na slike.

Uklanjanje oštećenih slika

Funkcija `remove_corrupted_images` osmišljena je za identifikaciju i uklanjanje oštećenih slika iz skupa podataka. Ova funkcija prolazi kroz svaki slikovni zapis, pokušava ga otvoriti te provjeriti njegovu ispravnost. Ako slika nije oštećena, njen indeks se pohranjuje za kasniju upotrebu. Ako je slika oštećena, izostavlja se iz skupa podataka za modeliranje.

Transformacije slika

Transformacije slika neophodne su za pripremu slika za modele dubokog učenja. Transformacije koje se koriste uključuju promjenu veličine slike, nasumično horizontalno rotiranje, promjene u boji, nasumično izrezivanje slike, konverziju u tensor i normalizaciju. Ove transformacije definirane su koristeći biblioteku `torchvision.transforms`.

- **Promjena veličine:** Prva transformacija mijenja veličinu slike na 224x224 piksela. Ovo je standardna veličina koja se često koristi u modelima dubokog učenja za slike te ta fiksna veličina omogućuje dosljedan unos u model.
- **Nasumično horizontalno rotiranje:** Ova transformacija nasumično rotira sliku horizontalno. Horizontalno rotiranje pomaže u augmentaciji podataka stvarajući varijacije koje model može koristiti za bolje učenje i generalizaciju.
- **Promjena boja:** Transformacija `ColorJitter` nasumično mijenja osvjetljenost, kontrast, zasićenje i nijansu slike. Ovo dodatno povećava varijabilnost podataka i pomaže modelu da postane otporniji na razlike u osvjetljenju i bojama.
- **Nasumično izrezivanje slike:** Transformacija `RandomResizedCrop` nasumično izrezuje sliku na veličinu 224x224 piksela, koristeći skalu od 90% do 100% originalne veličine slike. Ovo osigurava da model vidi različite dijelove slike tijekom treninga, što poboljšava njegovu sposobnost da prepozna bitne značajke.
- **Konverzija u tensor:** Transformacija `ToTensor` pretvara sliku u tensor, što je standardni format koji koristi PyTorch.
- **Normalizacija:** Posljednja transformacija normalizira vrijednosti piksela slike koristeći prosječne vrijednosti i standardne devijacije za crveni, zeleni i plavi kanal.

Normalizacija pomaže u stabilizaciji i ubrzanju treninga modela.

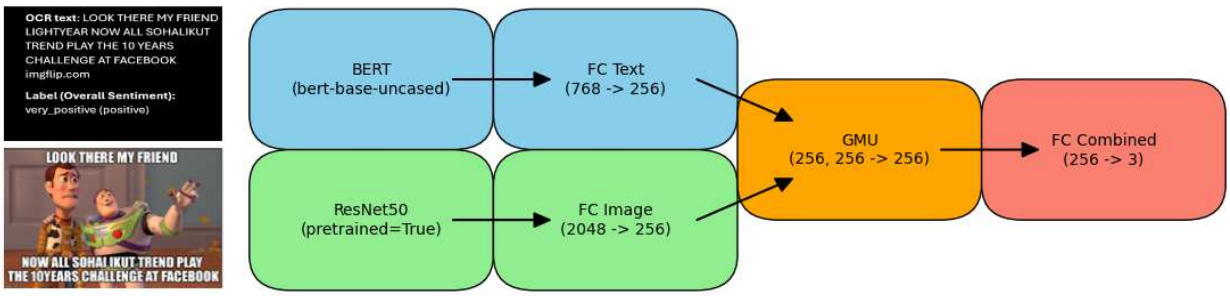
Ove transformacije osiguravaju da slikovni podaci budu dosljedni, očišćeni i pripremljeni za daljnju analizu.

4.3. Modeliranje

Cilj ovog istraživanja bio je ispitati učinkovitost različitih modela u analizi sentimenta u memeovima. Korišteno je pet različitih modela za tekstualne i vizualne podatke, kao i za njihovu fuziju, kako bi se dobio uvid u to koji pristupi najbolje funkcioniraju u ovom kontekstu. U nastavku su detaljno opisani korišteni modeli i njihove karakteristike.

4.3.1. Model 1,2: BERT-ResNet-GMU + BERT-ResNet-AttnLSTM

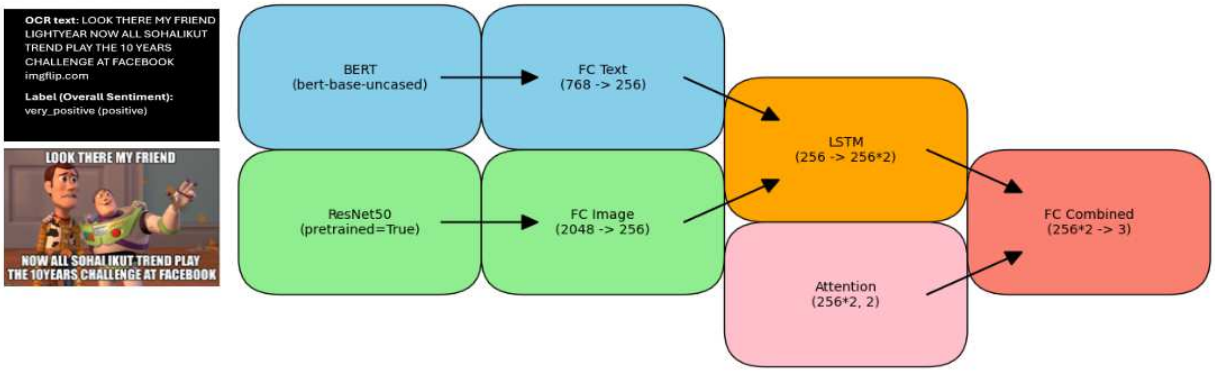
Implementirani modeli BERT-ResNet-GMU (slika 4.4.) te BERT-ResNet-AttnLSTM (slika 4.5.) koriste kombinaciju BERT-a [15] za ekstrakciju tekstualnih značajki i ResNet modela [16] za ekstrakciju vizualnih značajki. Ova dva skupa značajki zatim se spajaju pomoću Gated Multimodal Unit (GMU) [17] u prvom modelu dok se u drugom modelu značajke kombiniraju pomoću LSTM-a temeljenog na mehanizmu pozornosti, omogućavajući dinamičku integraciju informacija iz oba modaliteta.



Slika 4.4. Arhitektura BERT-ResNet-GMU

Ekstrakcija tekstualnih značajki pomoću BERT-a

BERT *(Bidirectional Encoder Representations from Transformers)* jedan je od najnaprednijih modela za obradu prirodnog jezika. U ovom istraživanju, BERT model korišten je za ekstrakciju značajki iz tekstualnih podataka memeova. Proces ekstrakcije značajki pomoću BERT-a uključuje:



Slika 4.5. Arhitektura BERT-ResNet-AttnLSTM

1. **Tokenizacija i unos u model:** Tekstualni podaci prvo su tokenizirani korištenjem BERT tokenizatora, koji pretvara tekst u odgovarajuće tokene, maske pozornosti (attention masks) i segmente.

2. **Prolazak kroz BERT model:** Tokenizirani podaci uneseni su u unaprijed istrenirani BERT model. BERT model generira vektorske reprezentacije (embeddinge) za svaki token u tekstu.

3. **Ekstrakcija značajki:** Značajke su izvučene iz zadnjeg sloja BERT modela. Korišteni su vektori reprezentacije CLS tokena (klasifikacijski token) kao agregirane značajke za cijeli ulazni tekst, što nam daje sveobuhvatan prikaz značenja i konteksta unutar teksta.

Ekstrakcija vizualnih značajki pomoću ResNet-a

ResNet (Residual Networks) napredni je CNN model koji se koristi za ekstrakciju značajki iz slika. U ovom istraživanju, unaprijed istrenirani ResNet model korišten je za dobivanje značajki iz vizualnih podataka memeova. Proces ekstrakcije značajki pomoću ResNet-a uključuje:

1. **Predprocesuiranje slika:** Slike su predprocesuirane uključivanjem promjene veličine, normalizacije i konverzije u tenzore kako bi bile prikladne za unos u ResNet model.

2. **Prolazak kroz ResNet model:** Predprocesuirane slike unesene su u unaprijed istrenirani ResNet model. ResNet model prolazi kroz nekoliko konvolucijskih slo-

jeva kako bi ekstrahirao vizualne značajke.

3. **Ekstrakcija značajki:** Značajke su ekstrahirane iz posljednjeg sloja prije klasifikacijskog sloja (potpuno povezanog sloja). Te značajke predstavljaju apstrakcije vizualnih elemenata slike i koriste se kao ulaz za daljnju fuziju s tekstualnim značajkama.

Fuzija značajki pomoću GMU (Gated Multimodal Unit)

Gated Multimodal Unit (GMU) omogućava dinamičku integraciju tekstualnih i vizualnih značajki koristeći mehanizam vrata (*eng. gates*). GMU arhitektura omogućava modelu da selektivno kombinira informacije iz oba modaliteta, poboljšavajući sposobnost modela da prepozna i razumije složene obrasce u podacima.

Kombinacija značajki: Nakon ekstrakcije značajki iz BERT-a i ResNet-a, kombiniraju se tekstualne i vizualne značajke:

1. **Proširenje dimenzija:** Značajke iz BERT-a i ResNet-a proširuju se tako da imaju istu dimenziju. To se postiže korištenjem linearnih slojeva (*eng. fully connected layers*) za smanjenje dimenzionalnosti značajki na istu veličinu.

2. **Kombinacija značajki:** Tekstualne i vizualne značajke kombiniraju se pomoću GMU, koji koristi mehanizam vrata za dinamičku integraciju informacija iz oba modaliteta.

Prolazak kroz GMU sloj: Kombinirane značajke prolaze kroz GMU sloj:

1. **Izračunavanje vrata (*eng. gates*):** GMU izračunava vrijednosti vrata za tekstualne i vizualne značajke koristeći sigmoidalnu funkciju. To omogućava modelu da dinamički prilagođava težinu svake vrste značajki.

2. **Izračun skrivenih reprezentacija:** Skriveni slojevi izračunavaju se za tekstualne i vizualne značajke koristeći `tanh` (tangens hiperbolni) funkciju. Time dobivamo nelinerarnu transformaciju ulaznih značajki.

3. **Kombinacija skrivenih reprezentacija:** Kombinacija skrivenih reprezentacija tekstualnih i vizualnih značajki postiže se ponderiranjem značajki prema izraču-

natim vrijednostima vrata. Ovo omogućava modelu da selektivno kombinira informacije iz oba modaliteta.

Izlazne značajke i klasifikacija: Nakon primjene GMU, dobivene značajke koriste se za konačnu klasifikaciju:

1. **Agregacija značajki:** Značajke kombinirane pomoću GMU agregiraju se kako bi dobili konačni vektor značajki koji predstavlja ulazne podatke.

2. **Konačna klasifikacija:** Konačni vektor značajki unosi se u potpuno povezani sloj koji provodi konačnu klasifikaciju sentimenta. Izlaz je vektor koji predstavlja predikcije za različite kategorije sentimenta (negativan, neutralan, pozitivan).

Fuzija značajki pomoću Attention LSTM

Long Short-Term Memory (LSTM) mreža s mehanizmom pozornosti koristi se za fuziju tekstualnih i vizualnih značajki. Mehanizam pozornosti omogućuje modelu da se fokusira na relevantne dijelove informacija iz oba modaliteta, poboljšavajući sposobnost modela da prepozna i razumije složene obrasce u podacima.

Kombinacija značajki: Nakon ekstrakcije značajki iz BERT-a i ResNet-a, kombiniraju se tekstualne i vizualne značajke:

1. **Proširenje dimenzija:** Značajke iz BERT-a i ResNet-a proširuju se tako da imaju istu dimenziju. To se postiže korištenjem linearnih slojeva (*eng. fully connected layers*) za smanjenje dimenzionalnosti značajki na istu veličinu.

2. **Kombinacija značajki:** Tekstualne i vizualne značajke kombiniraju se u jedan zajednički vektor značajki. Ovo se postiže konkatenacijom značajki duž nove dimenzije, stvarajući tenzor s dimenzijama (`batch size, sequence length, feature size`).

Prolazak kroz LSTM mrežu s mehanizmom pozornosti: Kombinirane značajke ulaze u LSTM mrežu koja koristi mehanizam pozornosti:

1. **LSTM za obradu sekvencijalnih podataka:** Kombinirane značajke prolaze kroz

LSTM sloj koji obrađuje sekvencijalne podatke. LSTM sloj omogućava modelu da hvata dugoročne zavisnosti i kontekstualne informacije iz kombiniranih značajki.

2. **Mehanizam pozornosti:** Mehanizam pozornosti primjenjuje se na izlaz LSTM sloja kako bi se model fokusirao na relevantne dijelove ulaznih značajki. Pozornost se računa pomoću ponderiranja izlaza LSTM-a, omogućavajući modelu da se fokusira na najvažnije dijelove ulaza.

3. **Izračun pozornosti:** Pozornost se računa koristeći skup težina koji se primjenjuju na izlazne značajke LSTM-a. Težine se normaliziraju kako bi se dobila distribucija pozornosti koja se primjenjuje na ulazne značajke.

Izlazne značajke i klasifikacija: Nakon primjene mehanizma pozornosti, dobivene značajke koriste se za konačnu klasifikaciju:

1. **Agregacija značajki:** Značajke ponderirane pozornošću agregiraju se kako bi se dobio konačni vektor značajki koji predstavlja ulazne podatke.

2. **Konačna klasifikacija:** Konačni vektor značajki unosi se u potpuno povezani sloj koji provodi konačnu klasifikaciju sentimenta. Izlaz je vektor koji predstavlja predikcije za različite kategorije sentimenta (negativan, neutralan, pozitivan).

Treniranje modela: BERT-ResNet-GMU + BERT-ResNet-AttnLSTM

Modeli su trenirani korištenjem ADAM optimizatora, usmjerenog na minimizaciju funkcije gubitka unakrsne entropije. Prilikom treniranja, korištena je veličina grupe od 32 uzoraka, a model je prošao kroz ukupno 20 epoha. Proces treniranja odvijao se na platformi Kaggle [18], koristeći GPU P100 akcelerator.

4.3.2. Model 3: MoodModel v3

MoodModel v3 [1], odnosno treći implementirani model sastoji se od četiri glavne komponente:

Ekstrakcija slikovnih značajki

Slikovni vektori značajki ekstrahirani su korištenjem predtrenirane EfficientNet mreže, trenirane na ImageNet skupu podataka, bez zadnjeg, klasifikacijskog sloja.

Ekstrakcija tekstualnih značajki

Za vektorsku reprezentaciju riječi u tekstualnom sadržaju memeova koristili smo 100-dimenzionalne GloVe ugradbene vektore, bazirane na velikom Twitter korpusu. Nakon toga, ove ugradbene vektore proslijedili smo kroz dvosmjerni LSTM kako bismo dobili bogatiju reprezentaciju značajki sposobnu za razumijevanje ukupnog konteksta rečenice.

Fuzija značajki

Ovaj korak uključuje kombiniranje vektora značajki iz teksta i slike koristeći dvosmjerni LSTM i mehanizam pozornosti. Dimenzija slikovnog vektora značajki smanjuje se na veličinu sloja dvosmjernog LSTM-a. Skriveno stanje i stanje ćelije dvosmjernog LSTM-a inicijaliziraju se slikovnim vektorom značajki. Pozornost se računa za tekstualne vektore značajki u odnosu na slikovni vektor značajki, koji se koristi kao ulaz u LSTM za svaku vremensku oznaku. Ovaj pristup daje veću važnost specifičnim riječima u odnosu na sliku, zanemarujući manje važne informacije. Izlazi iz LSTM-a zatim se prosljeđuju u GRU. Izlaz zadnje vremenske oznake GRU-a normalizira se, a izlaz iz slojeva za normalizaciju predstavlja spojeni vektor značajki.

Regularizacija

Zbog male veličine skupa podataka za obuku, model je podložan prekomjernom prilagođavanju. Kako bismo spriječili ili barem ublažili prenaučenost, koristili smo dropout za nasumično isključivanje neurona u mreži, čime sprječavamo susjedne neurone da uče slične značajke. U završnom gustom sloju također smo primijenili L2 regularizaciju.

Težine klasa

Klase skupa podataka nisu bile jednako uravnotežene što može značajno utjecati na pristranost modela većinskoj klasi. Da bismo to spriječili, koristili smo težine klasa za veće

kažnjavanje modela za predviđanje više zastupljene klase. Neka je $\mathbf{X}$ vektor koji sadrži brojač svake klase X_i gdje je $i \in \mathbf{X}$. Zatim su težine za svaku klasu dane kao:

$$\text{weight}_i = \frac{\max(X)}{X_i + \max(X)} \tag{4.1}$$

Hiperparametri

Ekstrahirani su 1280-dimenzionalni slikovni vektori značajki iz EfficientNet-a, koji su dalje smanjeni na 200 dimenzija koristeći gusti sloj s `relu` aktivacijom. Za reprezentaciju tekstualnog sadržaja memeova korišteni su 100-dimenzionalni GloVe vektori. Kroz cijeli model korišteni su dvosmjerni LSTM-ovi s 200 dimenzija. Korak fuzije značajki koristi GRU s 64 dimenzije. Dropout od 0.2 primijenjen je na ugradbenom sloju, 0.4 nakon prvog LSTM-a, 0.1 nakon koraka fuzije, te konačno 0.2 za povratne veze GRU-a i LSTM-a. Završni gusti sloj koristi L2 regularizaciju od 0.001 i `elu` aktivaciju. Izlaz završnog, klasifikacijskog sloja koristi `softmax` aktivaciju.

Treniranje modela

Model je treniran korištenjem ADAM optimizatora, usmjerenog na minimizaciju funkcije gubitka unakrsne entropije. Prilikom treniranja, korištena je veličina grupe od 200 uzoraka, a model je prošao kroz ukupno 200 epoha. Proces treniranja odvijao se na platformi Kaggle [18], koristeći napredne mogućnosti GPU P100 akceleratora.

4.3.3. Model 4: GPT-4 Vision - GPT-4 Turbo

Četvrti eksperiment koristi kombinaciju GPT-4 Vision modela za generiranje opisa slika memeova i GPT-4 Turbo modela za analizu sentimenta. Proces se sastoji od dva ključna koraka: slanja zahtjeva GPT-4 Vision modelu za generiranje opisa slika, a zatim slanje tih opisa zajedno s OCR tekstom GPT-4 Turbo modelu za određivanje sentimenta.

Generiranje opisa slika pomoću GPT-4 Vision

U prvom koraku, slike memeova šalju se GPT-4 Vision modelu kako bi se generirali opisi fokusirani na emocionalne aspekte slika. Prompt koji se šalje GPT-4 Vision modelu je sljedeći:

> Explain this image! Focus on emotional aspects of this image. Don't do OCR because I already have that extracted. In later stages I will use your generated image caption and OCR text for meme sentiment extraction. So just keep in mind what that caption will be used for.

Primjer slanja zahtjeva GPT-4 Vision modelu prikazan je u nastavku:

```
response = client.chat.completions.create(
model="gpt-4-vision-preview",
messages=[
{
"role": "user",
"content": [
{"type": "text", "text": "Explain this image! Focus on emotional aspects of this
image. Don't do OCR because I already have that extracted. In later stages
I will use your generated image caption and OCR text for meme sentiment
extraction. So just keep in mind what that caption will be used for."},
{"type": "image_url", "image_url": row['image_url']}
]
}
],
max_tokens=100
)
```

GPT-4 Vision model generira opis slike, koji se zatim pohranjuje zajedno s ostalim podacima o slici.

Analiza sentimenta pomoću GPT-4 Turbo

U drugom koraku, kombiniraju se generirani opisi slika i tekst ekstrahiran OCR-om kako bi se odredio sentiment memeova. Ovi podaci šalju se GPT-4-Turbo modelu sa sljedećim promptom:

```
prompt = f"""
Determine the sentiment of the meme image based on the following texts:
```

```
Image Caption: {row['image_caption']}
Corrected Text: {row['text_corrected']}

Analyze the relationship between these texts and provide the overall sentiment
(e.g., positive:2, negative:0, neutral:1) without explanations.
Return only one number (0 for negative, 1 for neutral or 2 for positive)!
"""

response = client.chat.completions.create(
model="gpt-4-turbo",
messages=[
{
"role": "user",
"content": prompt
}
],
max_tokens=100
)
```

GPT-4 Turbo model analizira odnos između generiranog opisa i OCR teksta te vraća sentiment (negativan, neutralan, pozitivan).

Implementacija modela

Implementacija modela uključuje sljedeće ključne komponente:

- **Generiranje opisa slika pomoću GPT-4 Vision**: Slanje zahtjeva GPT-4 Vision modelu kako bi se generirali opisi slika fokusirani na emocionalne aspekte.
- **Analiza sentimenta pomoću GPT-4 Turbo**: Kombinacija generiranih opisa i OCR teksta te slanje tih podataka GPT-4 Turbo modelu za određivanje sentimenta.
- **Integracija rezultata**: Pohrana rezultata analize sentimenta zajedno s originalnim podacima o slikama.

4.3.4. Model 5: GPT-4o

Zadnji eksperiment koristi GPT-4o model za klasifikaciju sentimenta kombiniranjem OCR teksta i URL-a slike memeova. S obzirom na to da se ovaj model trenutno smatra najboljim multimodalnim modelom, njegova primjena na slikama memeova čini se kao idealan izbor. Proces uključuje slanje zahtjeva GPT-4o modelu s promptom koji sadrži ispravljeni OCR tekst i URL slike, te dobivanje klasifikacije sentimenta.

Prompt za GPT-4o

U ovom modelu, svaki meme klasificira se na temelju kombinacije ispravljenog OCR teksta i slike. Prompt koji se šalje GPT-4o modelu sljedeći je:

> Task: Classify the sentiment of the provided meme image using the given OCR text and image URL.
>
> Inputs: Corrected Text: [extracted OCR text] Image URL: [meme image URL]
>
> Output: Determine the overall sentiment of the meme and return a single numerical value: (0 for negative, 1 for neutral, 2 for positive) Please return only the numerical value (0, 1, or 2) without any explanations.

Primjer slanja zahtjeva GPT-4o modelu prikazan je u nastavku:

```
prompt = f"""
Task: Classify the sentiment of the provided meme image
using the given OCR text and image URL.

Inputs:
Corrected Text: [extracted OCR text]
Image URL: [meme image URL]

Output: Determine the overall sentiment of the meme and
return a single numerical value: (0 for negative, 1 for neutral, 2 for positive)
Please return only the numerical value (0, 1, or 2) without any explanations.
"""
```

```
response = client.chat.completions.create(
model="gpt-4o",
messages=[
{
"role": "user",
"content": [
{"type": "text", "text": prompt},
{
"type": "image_url",
"image_url": {
"url": row['image_url'],
},
},
],
}
],
max_tokens=150
)
```

GPT-4o model analizira kombinaciju ispravljenog OCR teksta i slike te vraća sentiment (0 za negativan, 1 za neutralan, 2 za pozitivan).

Implementacija modela

Implementacija modela uključuje sljedeće ključne komponente:

- **Slanje zahtjeva GPT-4o modelu**: Za svaki meme, šalje se zahtjev GPT-4o modelu koji uključuje ispravljeni OCR tekst i URL slike. Model zatim generira klasifikaciju sentimenta.
- **Pohrana rezultata**: Rezultati analize sentimenta pohranjuju se zajedno s originalnim podacima o slikama i ispravljenim OCR tekstom.

5. Rezultati

U ovom poglavlju predstavljamo rezultate eksperimentalnih postavki za svih pet modela. Fokusirali smo se na macro F1 score zbog nebalansiranosti skupa podataka. Macro F1 score daje jednak značaj svakoj klasi, bez obzira na njezinu učestalost, čime osigurava da se performanse modela procjenjuju ravnopravno za sve klase. Ovo je posebno važno u našem slučaju jer želimo da naš model dobro prepoznaje sve vrste sentimenta (pozitivan, neutralan, negativan) i da ne bude pristran prema češćim klasama. Također prikazujemo matrice konfuzije i analiziramo problematiku rezultata, uključujući analizu grešaka.

5.1. Eksperimentalna postavka

Različiti modeli koriste različite skupove podataka i metode podjele kako bi se postigli najbolji rezultati u skladu s njihovim specifičnim zahtjevima i ograničenjima.

5.1.1. Model 1 i 2: BERT-ResNet-GMU i BERT-ResNet-AttnLSTM

Za prva dva modela korišten je isti skup podataka. Podjela skupa podataka izvršena je pomoću stratificirane metode kako bi se osigurali reprezentativni omjeri klasa u trening i validacijskom skupu.

Stratificirana podjela skupa podataka:

- **Veličina validacijskog skupa:** Postotak skupa podataka koji se koristi za validaciju iznosi 10%.
- **Stratificirana K-Fold metoda:** Koristi se `StratifiedKFold` metoda za podjelu skupa podataka. Broj podskupova inverzan je veličini validacijskog skupa.
- **Mapiranje oznaka:** Oznake klasa pretvaraju se u numeričke vrijednosti.

- **Generiranje podskupova:** Na temelju podskupova, određuju se indeksi za trening i validacijski skup.

Ovaj pristup osigurava da oba skupa podataka imaju reprezentativne omjere klasa, što je ključno za pouzdanu procjenu modela.

5.1.2. Model 3: MoodModel v3

Za ovaj model koristili smo običnu podjelu skupa podataka na trening i validacijski skup. Korištenje nasumičnog uzorka omogućuje nam da vidimo kako se model ponaša kada omjeri klasa u trening i validacijskom skupu nisu nužno isti.

Podjela skupa podataka:

- **Veličina validacijskog skupa:** Postotak skupa podataka koji će se koristiti za validaciju iznosi 20%.
- **Nasumična podjela:** Podaci su nasumično podijeljeni u trening i validacijski skup.

Ovaj pristup omogućuje procjenu performansi modela u stvarnim uvjetima, gdje omjeri klasa možda nisu savršeno uravnoteženi između trening i testnog skupa.

5.1.3. Model 4 i 5: GPT-4 Vision - GPT-4 Turbo i GPT-4o

Za ova dva modela koristili smo prvih 500 nasumično odabranih slika koje su prošle predprocesiranje. Razlog za korištenje manjeg skupa slika jest ograničenje API-ja i česti problemi s otvaranjem ili obradom nekih slika.

Podjela skupa podataka:

- **Prvih 500 slika:** Koristimo prvih 500 slika koje su prošle predprocesiranje i bile su ispravne za obradu putem API-ja.
- **Random odabir:** Slike su nasumično odabrane kako bi se smanjila pristranost.

Ovaj pristup omogućuje evaluaciju modela na manjem, ali kvalitetnom podskupu podataka zbog tehničkih ograničenja API-ja.

5.2. Macro F1 Score

5.2.1. Definicija i značaj

Macro F1 score metrika je koja se koristi za ocjenjivanje performansi klasifikacijskih modela, posebno kada se radi o nebalansiranim skupovima podataka. F1 score harmonijska je sredina preciznosti (*eng. precision*) i odziva (*eng. recall*). Dok se F1 score može računati za svaku klasu posebno, macro F1 score dobiva se kao prosjek F1 score-ova svih klasa, bez obzira na njihovu učestalost.

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{5.1}$$

Macro F1 score računa se kao:

$$\text{macro F1 score} = \frac{1}{N} \sum_{i=1}^{N} \text{F1 score}_i \tag{5.2}$$

gdje je N broj klasa.

5.2.2. Referentna vrijednost

Za usporedbu s našim rezultatima, koristimo referentnu vrijednost *(baseline macro f1 score)*. Referentna vrijednost predstavlja rezultat koji bi model postigao kada bi uvijek predviđao najzastupljeniju klasu u skupu podataka. S obzirom na to da je naš skup podataka nebalansiran, model bi postigao najbolji mogući rezultat u pogledu točnosti (*eng. accuracy*) predviđajući uvijek najzastupljeniju klasu. Međutim, takav pristup ne bi bio koristan za prepoznavanje drugih klasa, jer ne bi pružio uvid u performanse modela za manje zastupljene klase. Izračunavanje referentne vrijednosti omogućuje nam da postavimo donju granicu performansi i usporedimo je s našim modelima. Ako naš model postigne bolji macro F1 score u usporedbi s referentnom vrijednošću, to je pokazatelj da model uspješno prepoznaje i manje zastupljene klase, što je cilj našeg istraživanja.

Izračun baseline macro F1 scorea

1. **Identifikacija najzastupljenije klase:** Pretpostavimo da imamo tri klase: C_0, C_1 i C_2. Neka klasa C_m ima najveći broj primjera (najzastupljenija klasa).

2. **Izračun F1 score za svaku klasu:**

   Klasa C_0:

$$P_0 = \frac{TP_0}{TP_0 + FP_0}$$

$$TP_0 = 0 \quad \text{(nikada ne predviđamo klasu } C_0\text{)}$$

$$FP_0 = 0 \quad \text{(nikada ne predviđamo klasu } C_0\text{)}$$

$$P_0 = 0$$

$$R_0 = \frac{TP_0}{TP_0 + FN_0}$$

$$FN_0 = \text{broj stvarnih primjera klase } C_0$$

$$R_0 = 0$$

$$F1_0 = \frac{2 \cdot P_0 \cdot R_0}{P_0 + R_0} = 0$$

   Klasa C_1:

$$P_1 = \frac{TP_1}{TP_1 + FP_1}$$

$$TP_1 = 0 \quad \text{(nikada ne predviđamo klasu } C_1\text{)}$$

$$FP_1 = 0 \quad \text{(nikada ne predviđamo klasu } C_1\text{)}$$

$$P_1 = 0$$

$$R_1 = \frac{TP_1}{TP_1 + FN_1}$$

$$FN_1 = \text{broj stvarnih primjera klase } C_1$$

$$R_1 = 0$$

$$F1_1 = \frac{2 \cdot P_1 \cdot R_1}{P_1 + R_1} = 0$$

Klasa C_m (najzastupljenija klasa):

$$P_m = \frac{TP_m}{TP_m + FP_m}$$

$TP_m =$ broj stvarnih primjera klase C_m

$FP_m =$ broj predviđenih primjera klase C_m koji nisu stvarni primjeri klase C_m

$$P_m = \frac{TP_m}{TP_m + FP_m}$$

$$R_m = \frac{TP_m}{TP_m + FN_m}$$

$FN_m = 0$ (nikada ne propuštamo stvarne primjere klase C_m)

$$R_m = 1$$

$$F1_m = \frac{2 \cdot P_m \cdot R_m}{P_m + R_m} = \frac{2 \cdot P_m \cdot 1}{P_m + 1} = \frac{2 \cdot P_m}{P_m + 1}$$

3. **Izračun macro F1 scorea:**

$$\begin{aligned} F1_{macro} &= \frac{F1_0 + F1_1 + F1_m}{3} \\ &= \frac{0 + 0 + \frac{2 \cdot P_m}{P_m + 1}}{3} \\ &= \frac{\frac{2 \cdot P_m}{P_m + 1}}{3} \\ &= \frac{2 \cdot P_m}{3 \cdot (P_m + 1)} \end{aligned}$$

Dakle, baseline macro F1 score u ovom općenitom scenariju je $\frac{2 \cdot P_m}{3 \cdot (P_m + 1)}$.

5.3. Rezultati za Model 1: BERT-ResNet-GMU

5.3.1. Macro F1 Score

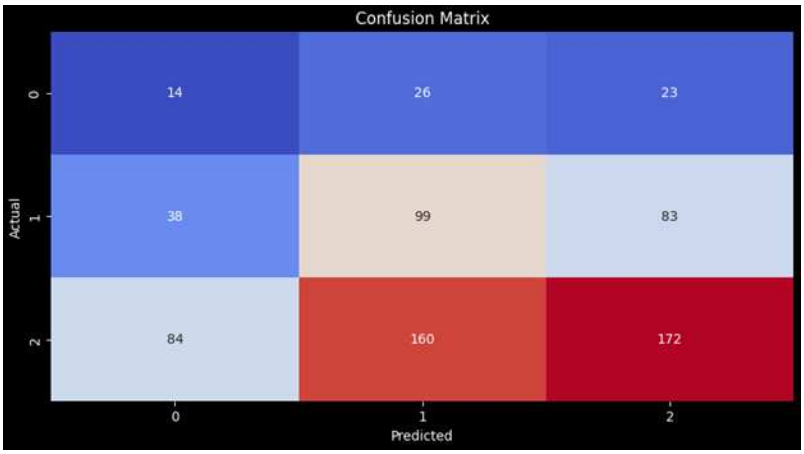
Rezultati na validacijskom skupu za model BERT-ResNet GMU prikazani su u Tablici 5.1.

baseline macro F1	macro F1	accuracy
0.2487	0.3535	0.4220

Tablica 5.1. Macro F1 Score (BERT-ResNet-GMU)

5.3.2. Matrica konfuzije

Matrica konfuzije za model BERT-ResNet-GMU prikazana je na slici 5.1.



Slika 5.1. Matrica konfuzije BERT-ResNet-GMU

5.4. Rezultati za Model 2: BERT-ResNet-AttnLSTM

5.4.1. Macro F1 Score

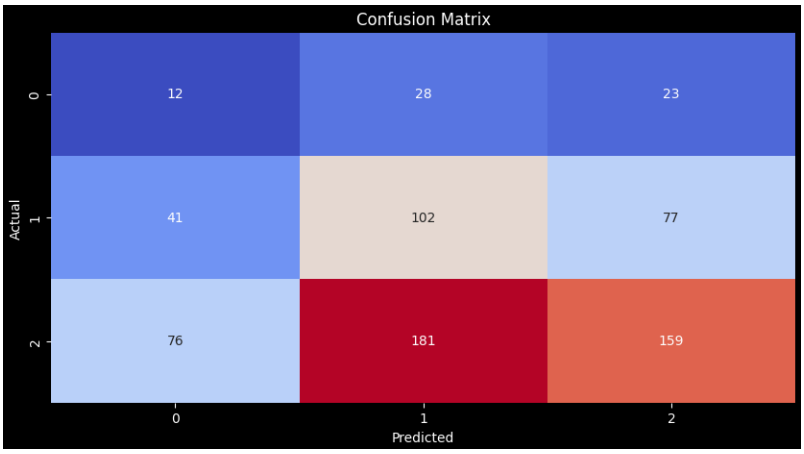
Rezultati na validacijskom skupu za model BERT-ResNet-AttnLSTM prikazani su u Tablici 5.2.

baseline macro F1	macro F1	accuracy
0.2487	0.3216	0.3920

Tablica 5.2. Macro F1 Score (BERT-ResNet-AttnLSTM)

5.4.2. Matrica konfuzije

Matrica konfuzije za model BERT-ResNet-AttnLSTM prikazana je na slici 5.2.



Slika 5.2. Matrica konfuzije BERT-ResNet-AttnLSTM

5.5. Rezultati za Model 3: MoodModel v3

5.5.1. Macro F1 Score

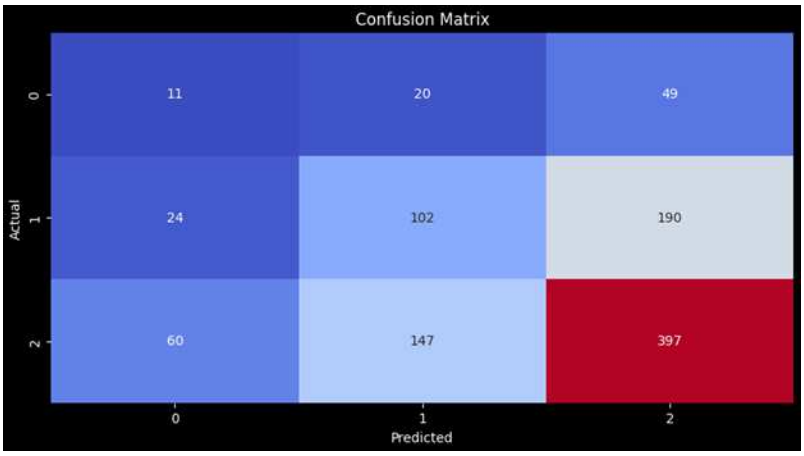
Rezultati na validacijskom skupu za model MoodModel v3 prikazani su u Tablici 5.3.

baseline macro F1	macro F1	accuracy
0.2487	**0.3716**	**0.5100**

Tablica 5.3. Macro F1 Score (MoodModel v3)

5.5.2. Matrica konfuzije

Matrica konfuzije za model MoodModel v3 prikazana je na slici 5.3.



Slika 5.3. Matrica konfuzije MoodModel v3

5.6. Rezultati za Model 4: GPT-4 Vision - GPT-4 Turbo

5.6.1. Macro F1 Score

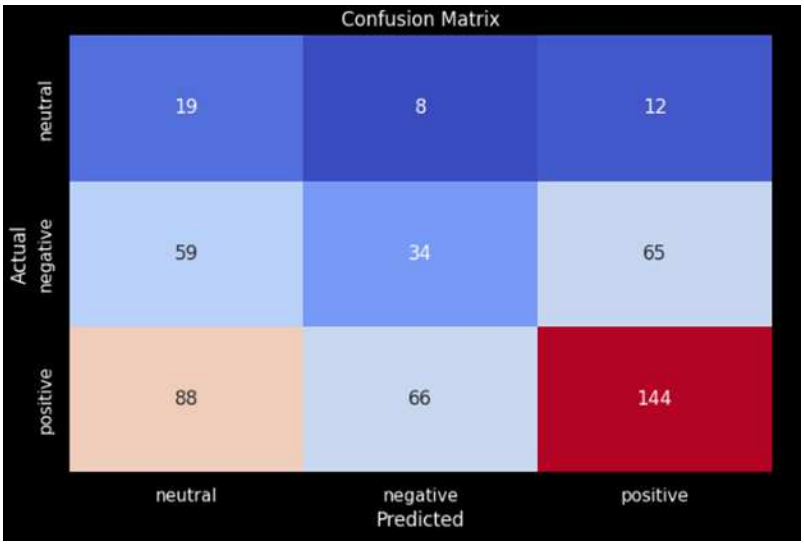
Rezultati na validacijskom skupu za model GPT-4 Vision - GPT-4 Turbo prikazani su u Tablici 5.4.

baseline macro F1	macro F1	accuracy
0.2505	0.3320	0.3980

Tablica 5.4. Macro F1 Score (GPT-4 Vision - GPT-4 Turbo)

5.6.2. Matrica konfuzije

Matrica konfuzije za model GPT-4 Vision - GPT-4 Turbo prikazana je na slici 5.4.



Slika 5.4. Matrica konfuzije GPT-4 Vision - GPT-4 Turbo

5.7. Rezultati za Model 5: GPT-4o

5.7.1. Macro F1 Score

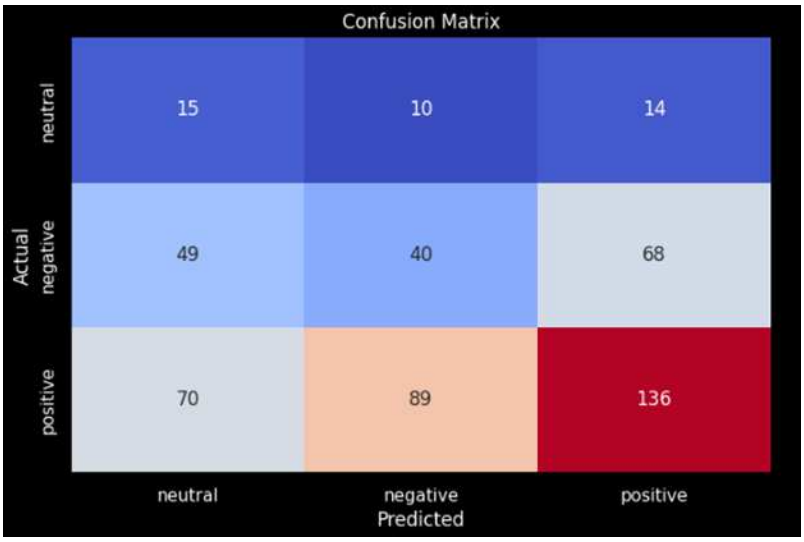
Rezultati na validacijskom skupu za model GPT-4o prikazani su u Tablici 5.5.

baseline macro F1	macro F1	accuracy
0.2505	0.3246	0.3890

Tablica 5.5. Macro F1 Score (GPT-4o)

5.7.2. Matrica konfuzije

Matrica konfuzije za model GPT-4o prikazana je na slici 5.5.



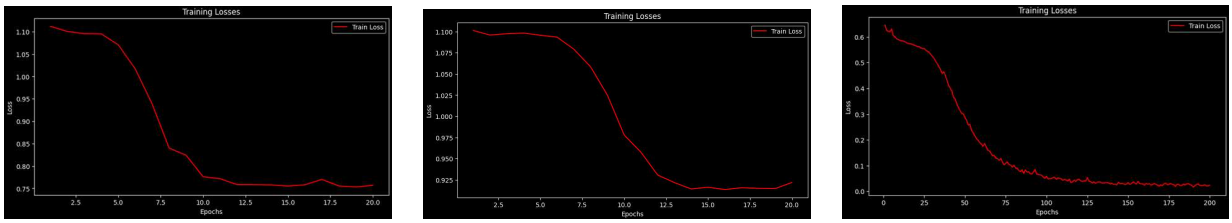
Slika 5.5. Matrica konfuzije GPT-4o

5.8. Grafički prikaz rezultata

Kako bismo bolje ilustrirali performanse modela tijekom treninga, prikazujemo grafove gubitaka na skupu za treniranje i validaciju te macro F1 score-a za sve modele. Svaki graf prikazuje vrijednosti po epohama, što omogućuje detaljan uvid u proces konvergencije i stabilnosti modela.

- **Train i validation gubitak:** Prikazuje gubitak na `train` i `validation` skupu kroz epohe.
- **Validation točnost:** Prikazuje točnost na `validation` skupu kroz epohe.
- **Validation macro F1 score:** Prikazuje macro F1 score `validation` skupu kroz epohe.

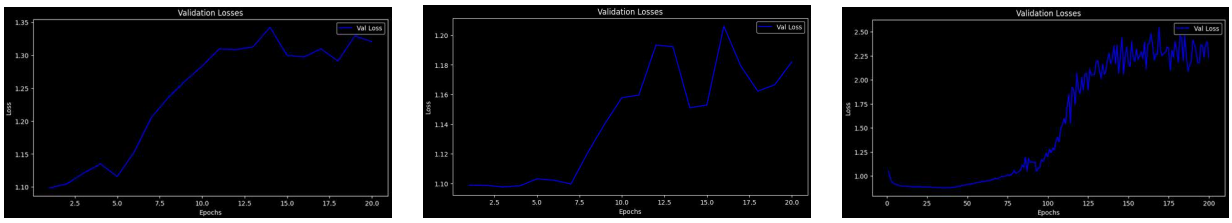
Grafovi gubitaka na skupu za treniranje prikazani su na slici 5.6.



(a) BERT-ResNet-GMU (b) BERT-ResNet-AttnLSTM (c) MoodModel v3

Slika 5.6. Gubitak na skupu za treniranje kroz epohe

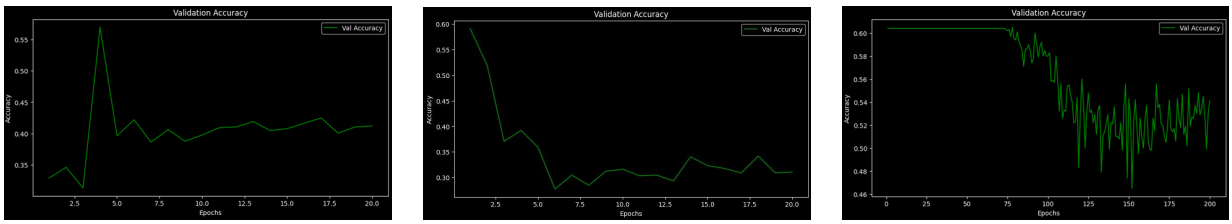
Grafovi gubitaka na skupu za validaciju prikazani su na slici 5.7.



(a) BERT-ResNet-GMU (b) BERT-ResNet-AttnLSTM (c) MoodModel v3

Slika 5.7. Gubitak na skupu za validaciju kroz epohe

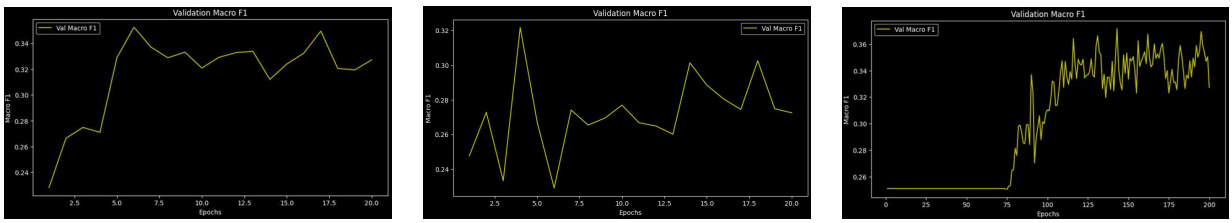
Grafovi točnosti na skupu za validaciju prikazani su na slici 5.8.



(a) BERT-ResNet-GMU (b) BERT-ResNet-AttnLSTM (c) MoodModel v3

Slika 5.8. Točnost na skupu za validaciju kroz epohe

Grafovi macro F1 score-a na skupu za validaciju prikazani su na slici 5.9.



(a) BERT-ResNet-GMU (b) BERT-ResNet-AttnLSTM (c) MoodModel v3

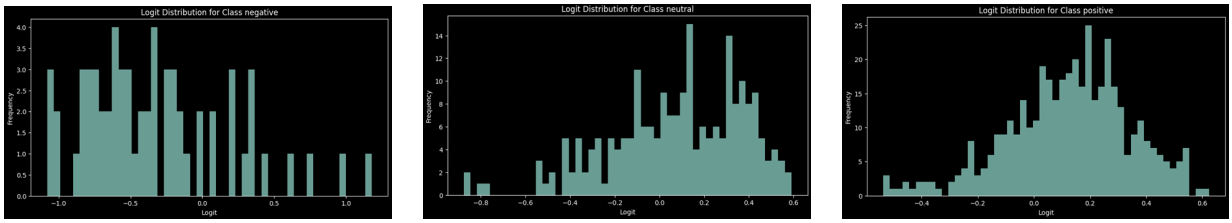
Slika 5.9. Macro F1 na skupu za validaciju kroz epohe

5.9. Analiza grešaka

U ovom poglavlju analiziramo rezultate klasifikacije i utjecaj logita na validacijski gubitak. Logiti predstavljaju neobrađene izlazne vrijednosti modela prije primjene softmax funkcije. Analizom logita možemo bolje razumijeti kako model donosi odluke i zašto dolazi do određenih grešaka.

5.9.1. Analiza logita BERT-ResNet-GMU

Logiti predstavljaju sirove rezultate koje model generira prije primjene softmax funkcije. Na slici 5.10. prikazane su distribucije logita za negativne, neutralne i pozitivne primjere.



(a) Negativna klasa (b) Neutralna klasa (c) Pozitivna klasa

Slika 5.10. Distribucija logita za različite klase

U tablici 5.6. prikazane su srednje vrijednosti i standardne devijacije logita za različite klase.

Klasa	Logiti (srednja vrijednost)	Standardna devijacija
Negativna	[-0.3282608, 0.09023456, 0.11946961]	[0.5035907, 0.3299622, 0.25607046]
Neutralna	[-0.32123443, 0.07480556, 0.12915738]	[0.41204968, 0.2969295, 0.20701788]
Pozitivna	[-0.29442227, 0.05102183, 0.12438601]	[0.42220795, 0.30092323, 0.21938013]

Tablica 5.6. Logiti i standardne devijacije za svaku klasu

Izračun softmax funkcije

Softmax funkcija pretvara logite u vjerojatnosti:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

Za navedene logite negativne klase, vjerojatnosti su:

$$\text{softmax}([-0.3282608, 0.09023456, 0.11946961]) = [p_0, p_1, p_2]$$

Izračun gubitka unakrsne entropije

Unakrsna entropija uspoređuje predviđene vjerojatnosti s pravim oznakama:

$$\text{gubitak} = -\sum_{i} y_i \log(p_i)$$

gdje je y_i prava oznaka (one-hot kodirana) i p_i predviđena vjerojatnost za klasu i.

Utjecaj na validacijski gubitak

- **Pogrešne klasifikacije:** Ako model generira visoke logite za pogrešnu klasu, gubitak će biti visok. Na primjer, ako je prava klasa `negativna`, ali logiti daju veću vjerojatnost za `pozitivnu`, gubitak će biti visok.
- **Ispravne klasifikacije s niskom sigurnošću:** Ako model ispravno klasificira primjer, ali s niskom sigurnošću (logiti su blizu jedan drugome), gubitak će biti viši nego da je model vrlo siguran. Na primjer, ako je prava klasa `negativna`, ali logiti daju samo malo veću vjerojatnost za `negativnu` u odnosu na druge klase, gubitak će i dalje biti relativno visok.
- **Ispravne klasifikacije s visokom sigurnošću:** Ako model ispravno klasificira primjer s visokom sigurnošću (logit za pravu klasu je mnogo viši od ostalih), gubitak će biti nizak. Na primjer, ako je prava klasa `negativna`, a logiti rezultiraju visokom vjerojatnošću za `negativnu` i vrlo niskim vjerojatnostima za ostale klase, gubitak će biti nizak.

Primjer izračuna

Dani logiti:

$$\text{logiti} = [-0.3282608, 0.09023456, 0.11946961]$$

Softmax izračun:

$$p_0 = \frac{e^{-0.3282608}}{e^{-0.3282608} + e^{0.09023456} + e^{0.11946961}}$$
$$p_1 = \frac{e^{0.09023456}}{e^{-0.3282608} + e^{0.09023456} + e^{0.11946961}}$$
$$p_2 = \frac{e^{0.11946961}}{e^{-0.3282608} + e^{0.09023456} + e^{0.11946961}}$$

Unakrsna entropija gubitka: Pretpostavimo da je prava oznaka `negativna`, one-hot kodirana oznaka y bila bi $[1, 0, 0]$ te bi vrijedilo:

$$\text{gubitak} = -[1 \cdot \log(p_0) + 0 \cdot \log(p_1) + 0 \cdot \log(p_2)]$$

Jednostavnije:

$$\text{gubitak} = -\log(p_0)$$

Utjecaj na validacijski gubitak:

- Ako je p_0 mali (tj. model nije siguran u ispravnu klasu), $-\log(p_0)$ će biti velik, povećavajući gubitak.
- Ako je p_0 velik (tj. model je siguran u ispravnu klasu), $-\log(p_0)$ će biti mali, smanjujući gubitak.

5.9.2. Zaključak analize logita

Povećanje validacijskog gubitka zajedno s poboljšanjem macro F1 score-a ukazuje na sljedeće:

- Model se poboljšava u prepoznavanju nekih klasa (što rezultira višim macro F1 score-om).

- Model se pogoršava u prepoznavanju drugih klasa, što dovodi do viših gubitaka za te pogrešne klasifikacije.
- Model može biti ispravniji, ali s nižom sigurnošću, što dovodi do višeg unakrsnog entropijskog gubitka čak i ako su klasifikacije ispravne.

U ovom eksperimentu vidimo da se model najviše muči s negativnom klasom, što je vidljivo iz najniže srednje vrijednosti logita i najveće varijabilnosti. Pozitivna klasa čini se najlakšom za klasificiranje modelu, s obzirom na najvišu srednju vrijednost logita i relativno nižu varijabilnost. Kako se model poboljšava u smislu macro F1 score-a, sve više primjera označava s negativnom i neutralnom klasom s ne toliko velikom vjerojatnošću, za razliku od početnih epoha gdje većini pridijeljuje pozitivnu oznaku s velikom sigurnošću. Iako model dodjeljuje negativne i neutralne klase, nije siguran u njih, što povećava gubitak, ali i poboljšava macro F1 rezultat.

5.10. Sumiranje rezultata

U ovom poglavlju sumiramo rezultate svih pet modela koji su korišteni za analizu sentimenta memeova. Kroz eksperimentalne postavke i analizu performansi, identificirali smo najbolji model, istovremeno prepoznajući izazove s kojima smo se susreli zbog prirode dataset-a.

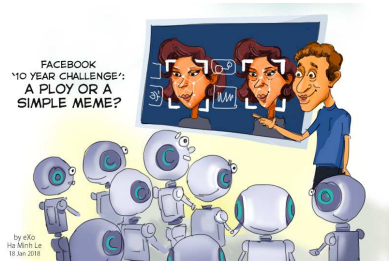
5.10.1. Komparativna analiza performansi

Svi modeli korišteni u ovom istraživanju pokazali su bolje rezultate u usporedbi s baseline modelom. Baseline model, koji predviđa najčešću klasu u datasetu, poslužio je kao donja granica za evaluaciju naših modela.

Izazovi dataset-a

Dataset koji smo koristili za analizu sentimenta memeova pokazao se izuzetno teškim za klasifikaciju. Memeovi često sadrže složene i dvosmislene poruke, što otežava modelima da precizno predvide sentiment. Ovo dodatno naglašava važnost naših rezultata jer su svi modeli nadmašili baseline performanse, unatoč izazovima koje dataset predstavlja. Na slici 5.11. prikazani su primjeri memova koje su svi modeli krivo klasificirali. Prva

dva memea označena su pozitivnim sentimentom u datasetu, dok ih modeli klasificiraju kao negativne ili neutralne. Za prvi meme može se zaključiti da na memeu ne postoje ni tekstualne ni vizualne značajke koje bi mogle otkriti sentiment poruke. Za drugi meme, pretpostavlja se da je došlo do šuma u oznakama, jer bih ga osobno klasificirao kao neutralan ili negativan, a nikako kao pozitivan, kako je označen. Treći meme ima pravu oznaku neutralnog sentimenta, dok svi modeli klasificiraju meme kao pozitivan, što ima smisla jer vizualne značajke (osmijeh na slici) mogu navesti modele na predviđanje pozitivne oznake. S druge strane na slici 5.12. prikazani su primjeri memeova s kojima nijedan model nije imao problema kod klasifikacije.



(a) Težak pozitivan meme br. 1

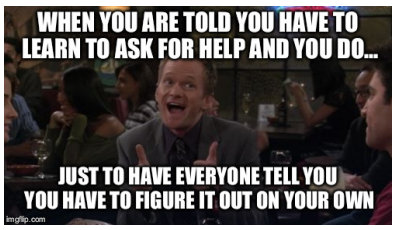
(b) Težak pozitivan meme br. 2

(c) Težak neutralan meme br. 3

Slika 5.11. Teški memeovi

(a) Lagani pozitivan meme

(b) Lagani neutralan meme br. 2



(c) Lagani negativan meme br. 3

Slika 5.12. Lagani memeovi

Najbolji model

Prema macro F1 score-u, najbolji model u našem istraživanju je treći model, odnosno MoodModel v3. Rezultati svih modela prikazani su u tablici 5.7.

Model	Macro F1 Score
Model 1 (BERT-ResNet-GMU)	0.3535
Model 2 (BERT-ResNet-AttnLSTM)	0.3216
Model 3 (MoodModel v3)	**0.3716**
Model 4 (GPT-4 Vision - GPT-4 Turbo)	0.3320
Model 5 (GPT-4o)	0.3246
Baseline model	0.2500

Tablica 5.7. Rezultati macro F1 score-a za različite modele

Performanse modela u odnosu na GPT-4o

Jedan od najvažnijih nalaza našeg istraživanja je da su svi custom implementirani modeli nadmašili GPT-4o model. GPT-4o, koji trenutno slovi kao jedan od najboljih multimodalnih modela na svijetu, poslužio je kao benchmark za našu analizu. Ipak, naši modeli pokazali su superiorne rezultate, što ukazuje na učinkovitost specijaliziranih arhitektura za zadatke analize sentimenta u memeovima.

6. Zaključak

U ovom poglavlju sažeti su ključni nalazi istraživanja, razmatrani su budući smjerovi istraživanja te su iznesene završne misli.

6.1. Pregled ključnih nalaza

Ovo istraživanje imalo je za cilj analizirati sentiment memeova korištenjem multimodalnih modela koji kombiniraju tekstualne i slikovne značajke. Glavni nalazi su sljedeći:

- Svi modeli koje smo razvili nadmašili su baseline model u pogledu macro F1 score-a, što potvrđuje njihovu učinkovitost u analizi sentimenta u nebalansiranom datasetu.
- MoodModel v3 pokazao se najboljim modelom s najvišim macro F1 score-om, što ga čini najučinkovitijim za prepoznavanje sentimenta u memeovima.
- Dataset koji smo koristili izuzetno je težak zbog složenih i dvosmislenih poruka u memeovima, što dodatno naglašava važnost postignutih rezultata.
- Implementirani modeli nadmašili su GPT-4o model, što pokazuje superiornost specijaliziranih arhitektura za zadatak analize sentimenta u memeovima.

6.2. Budući smjerovi istraživanja

Na temelju naših nalaza, identificirali smo nekoliko područja za buduće istraživanje:

- **Poboljšanje modela:** Daljnja optimizacija modela, uključujući dodatan fine-tuning hiperparametara i korištenje naprednijih tehnika fuzije značajki, mogla bi dodatno poboljšati performanse.

- **Raznolikost podataka:** Istraživanje na različitim skupovima podataka, uključujući memeove s različitim temama i stilovima, može pomoći u izradi robusnijih modela koji generaliziraju bolje na nove podatke.
- **Razumijevanje konteksta:** Razvoj modela koji bolje razumiju kontekst memeova, uključujući kulturne i društvene reference, može poboljšati točnost analize sentimenta.
- **Integracija dodatnih modaliteta:** Istraživanje mogućnosti uključivanja dodatnih modaliteta, poput zvuka ili videa, moglo bi proširiti primjenjivost modela na širi spektar multimedijalnih sadržaja.

6.3. Završne misli

Analiza sentimenta u memeovima predstavlja izuzetno težak zadatak zbog složenosti i dvosmislenosti sadržaja. Ipak, ovo istraživanje pokazalo je da je moguće postići značajna poboljšanja u performansama modela korištenjem naprednih tehnika fuzije značajki i specijaliziranih arhitektura. Najvažniji zaključak ovog istraživanja je da custom implementirani modeli mogu nadmašiti čak i najnaprednije generičke multimodalne modele poput GPT-4o kada su prilagođeni specifičnom zadatku.

Rezultati ovog istraživanja imaju značajne implikacije za šire područje analize digitalnih komunikacija i društvenih medija. Korištenje specijaliziranih modela može unaprijediti razumijevanje kompleksnih multimedijalnih sadržaja, što je posebno relevantno u eri gdje društvene mreže igraju ključnu ulogu u svakodnevnoj komunikaciji.

Daljnja istraživanja trebala bi se usmjeriti na dodatnu optimizaciju modela, proširenje raznolikosti podataka i bolju integraciju kontekstualnih informacija. Razumijevanje kulturoloških i društvenih referenci unutar memeova moglo bi značajno unaprijediti točnost modela u prepoznavanju sentimenta. Također, istraživanje mogućnosti uključivanja dodatnih modaliteta, poput zvuka ili videa, moglo bi proširiti primjenjivost modela na širi spektar multimedijalnih sadržaja.

Zaključno, nalazi ovog istraživanja otvaraju nove smjerove za istraživanje i razvoj u području multimodalne analize sentimenta, s potencijalom za značajne primjene u razli-

čitim domenama digitalne komunikacije i analize društvenih medija. Primjena ovih metoda može doprinijeti dubljem razumijevanju složenih i dinamičnih sadržaja koji oblikuju suvremeni digitalni pejzaž.

Literatura

[1] M. Sharma, I. Kandasamy, i W. Vasantha, “Memebusters at SemEval-2020 task 8: Feature fusion model for sentiment analysis on memes using transfer learning”, u *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May, i E. Shutova, Ur. Barcelona (online): International Committee for Computational Linguistics, prosinac 2020., str. 1163–1171. https://doi.org/10.18653/v1/2020.semeval-1.154

[2] J. Wang, J. Luo, G. Yang, A. Hong, i F. Luo, “Is gpt powerful enough to analyze the emotions of memes?” 2023.

[3] A.-M. Bucur, A. Cosma, i I.-B. Iordache, “Blue at memotion 2.0 2022: You have my image, my text and my transformer”, 2022.

[4] S. Pramanick, M. S. Akhtar, i T. Chakraborty, “Exercise? i thought you said ’extra fries’: Leveraging sentence demarcations and multi-hop attention for meme affect analysis”, 2021.

[5] K. Xuan, L. Yi, F. Yang, R. Wu, Y. R. Fung, i H. Ji, “Lemma: Towards lvlm-enhanced multimodal misinformation detection with external knowledge augmentation”, 2024.

[6] S. Lai, X. Hu, H. Xu, Z. Ren, i Z. Liu, “Multimodal sentiment analysis: A survey”, 2023.

[7] Robert Plutchik, “Plutchiks wheel of emotions”, 1980. [Mrežno]. Adresa: https://positivepsychology.com/emotion-wheel/

[8] A. L. P. V i E. M. Tolunay, “Dank learning: Generating memes using deep neural networks”, *arXiv preprint arXiv:1806.04510*, 2018. [Mrežno]. Adresa: https://arxiv.org/abs/1806.04510

[9] R. Dawkins, *The Selfish Gene*. Oxford: Oxford University Press, 1976.

[10] AWS, “What is sentiment analysis? - sentiment analysis explained”, 2024. [Mrežno]. Adresa: https://aws.amazon.com/what-is/sentiment-analysis/

[11] T. Nasukawa i J. Yi, “Sentiment analysis: Capturing favorability using natural language processing”, *Proceedings of the Association for Computational Linguistics*, sv. 1, 2003.

[12] K. Dave, S. Lawrence, i D. M. Pennock, “Mining the web for opinion structure”, *Proceedings of the 12th International Conference on World Wide Web*, sv. 1, 2003.

[13] MonkeyLearn, “Sentiment analysis”, 2024. [Mrežno]. Adresa: https://monkeylearn.com/sentiment-analysis/

[14] C. Sharma, W. Paka, Scott, D. Bhageria, A. Das, S. Poria, T. Chakraborty, i B. Gambäck, “Task Report: Memotion Analysis 1.0 @SemEval 2020: The Visuo-Lingual Metaphor!” u *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020)*. Barcelona, Spain: Association for Computational Linguistics, Sep 2020.

[15] J. Devlin, M.-W. Chang, K. Lee, i K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding”, 2019.

[16] K. He, X. Zhang, S. Ren, i J. Sun, “Deep residual learning for image recognition”, 2015.

[17] J. Arevalo, T. Solorio, M. Montes-y Gómez, i F. A. González, “Gated multimodal units for information fusion”, *arXiv preprint arXiv:1702.01992*, 2017. [Mrežno]. Adresa: https://arxiv.org/pdf/1702.01992

[18] Kaggle, “Kaggle: Your home for data science”, 2024., accessed: 2024-06-19. [Mrežno]. Adresa: https://www.kaggle.com

Sažetak

Multimodalna analiza sentimenta korištenjem teksta i slike

Tomislav Krog

Ovaj diplomski rad bavi se multimodalnom analizom sentimenta primjenom tekstualnih i slikovnih značajki na slikama memeova. Memeovi su poseban izazov za analizu sentimenta zbog svoje složenosti i dvosmislenosti. U radu su razvijeni i evaluirani različiti modeli koji integriraju tekstualne i slikovne značajke radi poboljšanja točnosti predikcije sentimenta. Primijenjene metode uključuju napredne tehnike fuzije značajki kao što su BERT-ResNet-GMU, BERT-ResNet-AttnLSTM, te integracije s modelima GPT-4-Vision i GPT-4. Najbolji model prema macro F1 score-u bio je MoodModel v3, koji koristi fuziju značajki putem dvosmjernog LSTM-a temeljenog na mehanizmu pažnje. Rezultati istraživanja pokazuju da posebno prilagođeni modeli značajno poboljšavaju performanse analize sentimenta u odnosu na osnovne modele. Rad također naglašava izazove rada s nebalansiranim datasetima i predlaže smjernice za buduća istraživanja u ovom području.

Ključne riječi: multimodalna analiza sentimenta; memeovi; emocije; BERT; ResNet; EfficientNet; GMU; LSTM; fuzija značajki; GPT-4; GPT-4o; analiza sentimenta

Abstract

Multimodal sentiment analysis using image and text

Tomislav Krog

This thesis explores the multimodal sentiment analysis of memes by leveraging both textual and visual features. Memes pose a unique challenge for sentiment analysis due to their complexity and ambiguity. In this study, various models that integrate textual and visual features were developed and evaluated to enhance the accuracy of sentiment prediction. The applied methods include advanced feature fusion techniques such as BERT-ResNet-GMU, BERT-ResNet-AttnLSTM, and integrations with GPT-4-Vision and GPT-4o models. The best-performing model, according to the macro F1 score, was Mood-Model v3, which employs feature fusion through a bidirectional LSTM based on an attention mechanism. The research findings demonstrate that custom-implemented models significantly improve sentiment analysis performance compared to baseline models. The thesis also highlights the challenges of working with imbalanced datasets and offers guidelines for future research in this field.

Keywords: multimodal sentiment analysis; memes; emotions; BERT; ResNet; EfficientNet; GMU; LSTM; feature fusion; GPT-4; GPT-4o; sentiment analysis