**VIŠJA STROKOVNA ŠOLA ACADEMIA**

**MARIBOR**

# PRIMERJAVA KLASIČNIH IN SODOBNIH METOD NLP Z VIDIKA UPORABNIŠKE RAZUMLJIVOSTI IN UČINKOVITOSTI V INTERAKTIVNEM SISTEMU

Kandidat: Žan Horvat

Vrsta študija: Študent izrednega študija

Študijski program: Računalništvo

Mentor predavatelj: mag. Ervin Schaff

Mentor v podjetju: Marko Bauman, mag. posl. Inf., dipl. inž. tel. (UN)

Lektor: Katja Flajs, dipl. slov.

Maribor, 2026

# IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Podpisani Žan Horvat, sem avtor diplomskega dela z naslovom Primerjava klasičnih in sodobnih metod NLP z vidika uporabniške razumljivosti in učinkovitosti v interaktivnem sistemu, ki sem ga napisal pod mentorstvom _____.

S svojim podpisom zagotavljam, da:

- je predloženo delo izključno rezultat mojega dela,
- sem poskrbel, da so dela in mnenja drugih avtorjev, ki jih uporabljam v predloženi nalogi, navedena oz. citirana skladno s pravili Višje strokovne šole Academia Maribor,
- se zavedam, da je plagiatorstvo – predstavljanje tujih del oz. misli, kot moje lastne kaznivo po Zakonu o avtorski in sorodnih pravicah (Uradni list RS, št. 16/07 – uradno prečiščeno besedilo, 68/08, 110/13, 56/15 in 63/16 – ZKUASP); prekršek pa podleže tudi ukrepom Višje strokovne šole Academia Maribor skladno z njenimi pravili,
- skladno z 32.a členom ZASP dovoljujem Višji strokovni šoli Academia Maribor objavo diplomskega dela na spletnem portalu šole.

Maribor, september 2026 Podpis študenta:

**ZAHVALA**

Zahvaljujem se mentorju in somentorju za strokovno vodenje, usmerjanje ter potrpežljivost pri nastajanju tega diplomskega dela.

Iskrena hvala tudi moji družini, ki mi je ves čas študija stala ob strani, me spodbujala, razumela in verjela vame. Brez vaše podpore ta pot ne bi bila enaka.

**POVZETEK**

Diplomsko delo obravnava primerjavo klasičnih in sodobnih metod obdelave NLP ter njihov vpliv na natančnost, hitrost obdelave, razumljivost rezultatov in uporabniško izkušnjo. Obdelava naravnega jezika predstavlja eno izmed ključnih področij umetne inteligence, saj omogoča računalniškim sistemom razumevanje, analizo in interpretacijo besedil v naravnem jeziku.

V teoretičnem delu so predstavljeni temeljni pristopi NLP, od klasičnih statističnih metod do sodobnih transformer modelov, ter pomen razložljivosti pri uporabi inteligentnih sistemov. Eksperimentalni del vključuje primerjavo metode TF-IDF z logistično regresijo in modela BERT. Uporabniki so v interaktivnem sistemu v Google Colab testirali obe metodi, sistem pa je izmeril čas obdelave, izpisal najpomembnejše besede in izvedel sentiment analizo. Rezultati vprašalnika v Google Forms so zajeli subjektivno oceno razumljivosti in uporabniške izkušnje.Raziskava je pokazala, da klasična metoda deluje izjemno hitro in ponuja jasne razlage, ki uporabnikom omogočajo dobro razumevanje rezultatov. BERT brez jezikovne prilagoditve pri slovenskih besedilih pogosto vrne napačne klasifikacije, obdelava pa je bistveno počasnejša. Zaradi nerazložljivosti rezultatov je bila uporabniška izkušnja pri sodobni metodi slabša. Analiza rezultatov je pokazala, da kompleksnost sodobnih modelov ne zagotavlja nujno boljše praktične uporabnosti. Klasične metode, čeprav preproste, ponujajo stabilne, hitre in razložljive rezultate, ki so za končnega uporabnika pogosto bolj uporabni. Sodobni modeli postanejo učinkoviti šele, ko so prilagojeni jeziku in podatkovnemu naboru, kar zahteva dodatne vire in znanje. Raziskava tako poudarja pomen razložljivosti, preglednosti in prilagoditve modelov pri razvoju NLP sistemov.

Na podlagi rezultatov lahko sklepamo, da sodobne metode niso vedno najboljša izbira v praktičnih okoljih, zlasti kadar niso prilagojene jeziku ali ko je za uporabnika ključna preglednost rezultatov. Klasične metode ostajajo pomembne zaradi svoje hitrosti, stabilnosti in razložljivosti. Diplomsko delo potrjuje, da je izbira metode odvisna od konteksta uporabe in pričakovanj uporabnika, razložljivost pa predstavlja ključen element sodobnih NLP sistemov.

**Ključne besede:** *obdelava naravnega jezika, TF-IDF, BERT, primerjava metod, uporabniška izkušnja, razložljivost*

**ABSTRACT**

**Comparison of classical and modern NLP methods in terms of user interpretability and efficiency in an interactive system**

The thesis examines a comparison between classical and modern NLP methods and their impact on accuracy, processing speed, result interpretability, and user experience. Natural Language Processing is one of the key fields of artificial intelligence, as it enables computer systems to understand, analyze, and interpret text written in human language.

The theoretical part presents the fundamental NLP approaches, ranging from classical statistical methods to modern transformer-based models, and highlights the importance of explainability in the use of intelligent systems. The experimental part includes a comparison between the TF-IDF method with logistic regression and the BERT model. Users tested both methods in an interactive Google Colab system, which measured processing time, displayed the most important words, and performed sentiment analysis. The results of the Google Forms questionnaire captured subjective assessments of interpretability and user experience.

The research showed that the classical method operates extremely quickly and provides clear explanations that help users understand the results. BERT, without language adaptation, often produces incorrect classifications for Slovenian texts, and its processing is significantly slower. Due to the lack of explainability, the user experience with the modern method was poorer. The analysis of results demonstrated that the complexity of modern models does not necessarily guarantee better practical usefulness. Although simple, classical methods offer stable, fast, and interpretable results that are often more useful for end users. Modern models become effective only when adapted to the language and dataset, which requires additional resources and expertise. The research therefore emphasizes the importance of explainability, transparency, and model adaptation in the development of NLP systems.

Based on the results, we can conclude that modern methods are not always the best choice in practical environments, especially when they are not adapted to the language or when result interpretability is crucial for the user. Classical methods remain an important part of NLP, as they enable fast, stable, and transparent text processing. The thesis confirms that the choice of method depends on the context of use, user expectations, and available resources, and that explainability is a key element in designing modern NLP systems.

**Keywords:** *natural language processing, TF-IDF, BERT, method comparison, user experience, explainability*

# KAZALO VSEBINE


**1 UVOD ..... 9**

1.1 Opis področja in opredelitev problema ..... 9

1.2 Namen, cilji in osnovne trditve ..... 10

1.3 Predpostavke in omejitve ..... 11

1.4 Uporabljene raziskovalne metode ..... 12

**2 TEORETIČNA IZHODIŠČA ..... 14**

2.1 Obdelava naravnega jezika (NLP) ..... 14

*2.1.1 Osnove obdelave naravnega jezika ..... 15*

*2.1.2 Vloga NLP v sodobnih informacijskih sistemih ..... 18*

*2.1.3 Pregled uporabe NLP v praksi ..... 20*

2.2 Primerjava klasičnih in sodobnih metod (NLP) ..... 22

*2.2.1 Klasične metode ..... 23*

*2.2.2 Sodobne metode ..... 25*

*2.2.3 Prednosti in slabosti posameznih pristopov ..... 28*

*2.2.4 Kompleksnost modelov in interpretacija rezultatov ..... 30*

**3 RAZISKOVALNI DEL ..... 33**

3.1 Razvoj in implementacija rešitve ..... 33

*3.1.1 Opis podatkovnega nabora ..... 33*

*3.1.2 Načrt sistema (Google Colab) ..... 36*

*3.1.3 Implementacija klasične metode ..... 38*

*3.1.4 Implementacija sodobne metode ..... 41*

*3.1.5 Vizualizacija in prikaz rezultatov ..... 44*

3.2 Analiza rezultatov in uporabniška evalvacija ..... 47

*3.2.1 Kvantitativna analiza (natančnost, čas izvajanja) ..... 47*

*3.2.2 Primerjava rezultatov med metodami ..... 48*

*3.2.3 Uporabniška analiza ..... 49*

*3.2.4 Razprava o rezultatih ..... 63*

**4 SKLEP ..... 65**

**5 VIRI IN LITERATURA ..... 68**

# KAZALO SLIK

Slika 2.1: Struktura BERT modela ..... 27
Slika 3.1: Notebook v Google Colab ..... 34
Slika 3.2: Uvažanje podatkov v Google Colab ..... 35
Slika 3.3: Prikaz prvih petih vrstic podatkovnega nabora (df.head()) ..... 35
Slika 3.4: Informacije o podatkovnem naboru (df.info()) in porazdelitev razredov ..... 36
Slika 3.5: Arhitektura prototipnega sistema v Google Colab ..... 37
Slika 3.6: Struktura Google Colab notebooka ..... 38
Slika 3.7: Pretvorba besedila v TF-IDF matriko ..... 40
Slika 3.8: Izpis časa in natančnosti ter F1-mere logistične regresije na testnem delu podatkovnega nabora IMDb ..... 41
Slika 3.9: Priprava podvzorca podatkov in tokenizacija besedila z BERT tokenizerjem ..... 42
Slika 3.10: Inicializacija modela BERT in nastavitev parametrov učenja ..... 43
Slika 3.11: Potek učenja modela BERT po posameznih epohah (izguba, natančnost in F1-mera) ..... 44
Slika 3.12: Primerjalna tabela iz okolja Google Colab ..... 45
Slika 3.13: Grafična primerjava natančnosti in časa izvajanja med klasično in sodobno metodo ..... 46
Slika 3.14: Testiranje Google Colab – prvi uporabnik ..... 51
Slika 3.15: Testiranje Google Colab – drugi uporabnik ..... 52
Slika 3.16: Testiranje Google Colab - tretji uporabnik ..... 52
Slika 3.17: Testiranje Google Colab – četrti uporabnik ..... 53
Slika 3.18: Testiranje Google Colab - peti uporabnik ..... 53
Slika 3.19: Testiranje Google Colab - šesti uporabnik ..... 54
Slika 3.20: Testiranje Google Colab – sedmi uporabnik ..... 54
Slika 3.21: Vprašalnik v Google Forms ..... 58

# KAZALO TABEL

Tabela 3.1: Primerjalna tabela rezultatov ..... 45
Tabela 3.2: Prikaz razlike med metodama ..... 45
Tabela 3.3: Razlika med metodama glede na posamezni kazalnik ..... 48
Tabela 3.4: Rezultati Google Colab za vseh sedem uporabnikov ..... 56
Tabela 3.5: Odgovori uporabnikov na sedmo vprašanje (kratek komentar) ..... 63

# KAZALO GRAFOV

Graf 3.1: Rezultati glede razumljivosti klasične metode ..... 59
Graf 3.2: Rezultati glede razumljivosti BERT metode ..... 60
Graf 3.3: Rezultati glede hitrosti klasične metode ..... 60

GRAF 3.4: REZULTATI GLEDE HITROSTI BERT METODE ..... 61
GRAF 3.5: REZULTATI GLEDE JASNOSTI KLASIČNE METODE ..... 62
GRAF 3.6: REZULTATI GLEDE PREFERIRANE METODE ..... 62

# 1 UVOD

## *1.1 Opis področja in opredelitev problema*

Diplomsko delo je obravnavalo področje obdelave naravnega jezika (NLP), ki spada med pomembnejša področja umetne inteligence. NLP omogoča računalnikom, da razumejo, analizirajo in interpretirajo človeški jezik, kar se danes uporablja v številnih aplikacijah, kot so pametni asistenti, prevajalniki in sistemi za analizo besedil.

V okviru naloge smo se osredotočili na primerjavo klasičnih metod (kot sta Bag-of-Words in TF-IDF) ter sodobnih pristopov, ki temeljijo na strojnih in globokih nevronskih modelih (npr. transformerji in modeli tipa BERT). Razlog za izbiro te teme je predvsem hiter razvoj sodobnih metod, ki pogosto nadomeščajo klasične pristope, vendar njihova dejanska prednost v praksi ni vedno enostavno razumljiva ali jasno predstavljena.

V strokovni literaturi obstaja veliko teoretičnih primerjav teh metod, vendar so te pogosto osredotočene predvsem na matematične modele in rezultate, manj pa na vidik končnega uporabnika. Prav tako pogosto manjka enoten praktičen primer, kjer bi se metode primerjale na enakem podatkovnem naboru in na način, ki bi omogočal neposredno razumevanje razlik v delovanju, hitrosti in rezultatih.

Zato smo želeli v diplomskem delu to vrzel zapolniti tako, da smo:

- metode primerjali na konkretnem primeru,
- rezultate kvantitativno izmerili (npr. natančnost, čas izvajanja),
- in jih predstavili v obliki preprostega orodja (Google Colab aplikacije), ki bo razumljiva tudi uporabniku brez poglobljenega tehničnega znanja.

Tak pristop je skladen tudi s sodobnimi smernicami razvoja digitalnih rešitev, ki poudarjajo na človeka usmerjeno informatiko, kjer je pomembno, da so sistemi ne le učinkoviti, ampak tudi razumljivi in uporabni za širši krog uporabnikov.

Osrednji raziskovalni problem diplomskega dela je bil ugotoviti, kako se klasične in sodobne metode obdelave naravnega jezika razlikujejo glede na natančnost, hitrost ter razumljivost rezultatov z vidika končnega uporabnika. Pri tem smo se posebej osredotočili na vprašanje, ali sodobne metode dejansko prinašajo bistveno boljše rezultate in ali njihova večja kompleksnost vpliva na uporabniško izkušnjo ter razumljivost rezultatov.

## *1.2 Namen, cilji in osnovne trditve*

Namen diplomskega dela je bil sistematično primerjati klasične in sodobne metode obdelave naravnega jezika (NLP) ter oceniti njihov vpliv na natančnost, čas obdelave, razumljivost rezultatov in uporabniško izkušnjo. Delo združuje tehnično analizo metod z načeli na človeka usmerjene informatike, zato poleg objektivnih metrik vključuje tudi uporabniško evalvacijo.

Glavni cilji diplomskega dela so bili:

- analizirati temeljne razlike med klasičnimi metodami NLP (Bag-of-Words, TF-IDF) in sodobnimi metodami (word embeddings, transformer modeli),
- izvesti eksperiment, ki primerja natančnost, F1-score, čas izvajanja in porabo virov med izbranimi metodami,
- oceniti, kako kompleksnost modelov vpliva na interpretacijo rezultatov za končnega uporabnika,
- preveriti, kako izbira metode vpliva na uporabniško izkušnjo pri interakciji s sistemom,
- razviti preprost interaktivni sistem, ki omogoča primerjavo metod in vizualizacijo rezultatov,
- preveriti zastavljene hipoteze na podlagi objektivnih meritev in uporabniške evalvacije.

Zastavili smo si naslednja raziskovalna vprašanja:

RV1: Kakšna je razlika v natančnosti med klasičnimi in sodobnimi metodami obdelave naravnega jezika?

RV2: Kolikšna je razlika v času obdelave podatkov med posameznimi metodami?

RV3: Kako kompleksnost modelov vpliva na razumljivost rezultatov za končnega uporabnika?

RV4: Kakšen vpliv ima izbira metode na uporabniško izkušnjo pri interakciji s sistemom?

RV5: Ali sodobne metode vedno predstavljajo boljšo izbiro glede na praktično uporabo?

Osnovne trditve diplomskega dela izhajajo iz raziskovalnih vprašanj:

H1: Sodobne metode NLP dosegajo vsaj 10 % višjo natančnost pri klasifikaciji besedila v primerjavi s klasičnimi metodami.

H2: Klasične metode NLP imajo statistično značilno krajši čas izvajanja kot sodobne metode.

H3: Uporabniki ocenijo rezultate klasičnih metod kot bolj razumljive v primerjavi z rezultati sodobnih metod (povprečna ocena vsaj 4 na lestvici od 1 do 5).

H4 – dodatna hipoteza – BTEC 2026 »Na človeka usmerjena informatika«: Sistem, ki uporabniku omogoča vpogled v delovanje modela (primerjava metod in prikaz rezultatov), izboljša uporabniško izkušnjo vsaj za 20 % glede na osnovno predstavitev rezultatov brez razlage.

## *1.3 Predpostavke in omejitve*

Pri izvedbi raziskave upoštevamo naslednje predpostavke:

- izbrani podatkovni nabor (npr. sentiment analiza) je dovolj reprezentativen za primerjavo metod,
- izbrane metode (TF-IDF + logistična regresija ter BERT) ustrezno predstavljajo klasičen in sodoben pristop,
- uporabniki, vključeni v evalvacijo, imajo osnovno računalniško pismenost in razumejo navodila za uporabo sistema,
- ocenjevalna lestvica (1–5) je primerna za merjenje razumljivosti in uporabniške izkušnje.

Omejitve raziskave:

- velikost uporabniške skupine (7 oseb) je majhna, zato rezultati evalvacije niso statistično posplošljivi,
- primerjava je omejena na en podatkovni nabor in en tip naloge (klasifikacija besedila),
- čas izvajanja je odvisen od strojne opreme (Google Colab), kar lahko vpliva na absolutne vrednosti meritev,
- izbrane metode ne pokrivajo celotnega spektra NLP pristopov, temveč predstavljajo reprezentativne primere,
- razumljivost rezultatov je subjektivna metrika, ki je odvisna od posameznikovega predznanja.

## *1.4 Uporabljene raziskovalne metode*

Raziskovalni proces je sestavljen iz teoretičnega in praktičnega dela.

V teoretičnem delu smo analizirali strokovno literaturo s področja:

- obdelave NLP,
- klasičnih metod (Bag-of-Words, TF-IDF),
- sodobnih metod (word embeddings, transformer modeli),
- uporabniške izkušnje (UX) in na človeka usmerjene informatike.

Literatura je služila za:

- opredelitev razlik med metodami,
- pregled njihovih prednosti in slabosti,
- razumevanje vpliva kompleksnosti modelov na interpretacijo rezultatov.

V praktičnem delu smo izvedli eksperiment, ki vključuje:

**Primerjane metode:**

- klasičen pristop: TF-IDF + logistična regresija,
- sodoben pristop: BERT ali drug transformer model.

**Merjeni parametri:**

- natančnost (accuracy),
- F1-score,
- čas izvajanja (processing time).

Uporabili smo enak podatkovni nabor, kar omogoča objektivno primerjavo.

**Uporabniška evalvacija**

Ker delo sledi načelom na človeka usmerjene informatike, smo izvedli preprosto uporabniško evalvacijo:

- vključili smo 7 uporabnikov,
- ti so uporabljali razviti interaktivni sistem,
- ocenili so razumljivost rezultatov,

- ocenili so uporabniško izkušnjo.

Metoda: vprašalnik z ocenjevalno lestvico (1–5) in analiza povprečnih ocen.

**Preverjanje hipotez**

H1: primerjava accuracy in F1-score med metodami,

H2: merjenje časa izvajanja,

H3: analiza rezultatov vprašalnika (razumljivost),

H4: primerjava uporabniške izkušnje med sistemom z razlago in brez nje.

**Razvoj interaktivnega sistema**

Kot del diplomskega dela smo razvili preprost sistem (Google Colab), ki omogoča:

- vnos besedila,
- analizo z različnimi NLP metodami,
- prikaz rezultatov,
- primerjavo natančnosti in časa obdelave,
- osnovno vizualizacijo rezultatov.

Sistem je zasnovan tako, da je razumljiv tudi manj tehničnim uporabnikom, omogoča enostavno primerjavo metod in podpira boljšo interpretacijo rezultatov.

# 2 TEORETIČNA IZHODIŠČA

## *2.1 Obdelava naravnega jezika (NLP)*

Obdelava naravnega jezika (angl. Natural Language Processing, NLP) predstavlja interdisciplinarno področje, ki združuje računalništvo, jezikoslovje, statistiko in umetno inteligenco z namenom omogočanja računalnikom, da razumejo, interpretirajo in generirajo človeški jezik. Temeljni cilj NLP je razviti metode, ki omogočajo učinkovito obdelavo besedilnih in govornih podatkov ter avtomatizacijo nalog, ki so bile tradicionalno vezane na človeško razumevanje jezika. V zadnjih desetletjih se je NLP iz pretežno simbolnega pristopa razvil v kompleksno področje, ki temelji na statističnih modelih, nevronskih mrežah in globokem učenju. Ta razvoj je omogočil bistveno izboljšanje natančnosti in zmogljivosti sistemov, ki se uporabljajo v vsakdanjih aplikacijah, kot so prevajalniki, iskalniki, klepetalni sistemi in orodja za analizo sentimenta (Jurafsky & Martin, 2023, str. 1–5).

Zgodovinsko gledano so bili prvi pristopi k obdelavi jezika simbolni in ročno zasnovani, kar je zahtevalo obsežno jezikoslovno znanje ter ročno pripravljene slovarje, pravila in sintaktične strukture. Takšni sistemi so bili pogosto togi, slabo prilagodljivi in občutljivi na napake v vhodnih podatkih. Prelom se je zgodil z uvedbo statističnih metod, ki so omogočile učenje jezikovnih vzorcev iz velikih količin podatkov. Statistični pristopi, kot so modeli n-gramov, linearne klasifikacije in metode, ki temeljijo na verjetnostnih porazdelitvah, so postali temelj sodobnega NLP v 90. letih. Njihova prednost je bila v tem, da so se lahko prilagajali realnim podatkom in so bili manj odvisni od ročno zasnovanih pravil (Manning & Schütze, 1999, str. 10–25).

V začetku 21. stoletja je NLP doživel novo revolucijo z uvedbo nevronskih mrež, ki so omogočile učenje kompleksnih jezikovnih reprezentacij. Namesto ročno zasnovanih značilnosti so modeli začeli samodejno izluščevati semantične in sintaktične informacije iz besedil. »Word embeddings«, kot so Word2Vec in GloVe, so omogočili, da se besede predstavijo v večdimenzionalnem prostoru, kjer razdalje med vektorji odražajo semantične odnose. Ta pristop je bistveno izboljšal zmogljivost modelov pri nalogah, kot so klasifikacija besedila, razpoznavanje entitet in strojno prevajanje (Goldberg, 2017, str. 1–20).

Največji preboj v sodobnem NLP pa predstavljajo transformer modeli, ki temeljijo na mehanizmu pozornosti (angl. attention). Modeli, kot je BERT, so omogočili globoko

kontekstualno razumevanje jezika, saj lahko upoštevajo celoten kontekst stavka in ne le lokalnih informacij. Transformer arhitektura je odpravila omejitve rekurentnih nevronskih mrež ter omogočila paralelno obdelavo podatkov, kar je bistveno pospešilo učenje in izboljšalo natančnost. Danes transformer modeli predstavljajo standard v večini naprednih NLP aplikacij (Vaswani et al., 2017, str. 1–10; Devlin et al., 2019, str. 4171–4175).

NLP ima ključno vlogo v sodobnih informacijskih sistemih, saj omogoča avtomatizacijo procesov, ki vključujejo obdelavo besedila, komunikacijo z uporabniki in interpretacijo podatkov. Sistemi, ki temeljijo na NLP, so postali nepogrešljivi v poslovnih okoljih, zdravstvu, izobraževanju, javni upravi in številnih drugih sektorjih. Zmožnost razumevanja naravnega jezika omogoča razvoj inteligentnih vmesnikov, ki izboljšujejo uporabniško izkušnjo, zmanjšujejo kognitivno obremenitev in povečujejo dostopnost informacij (Johnson, 2020, str. 15–30).

Pomemben vidik sodobnega NLP je tudi dostopnost in standardizacija, ki jo spodbujajo organizacije, kot je W3C. Smernice WCAG 2.2 poudarjajo pomen jasnega, razumljivega in dostopnega jezika v digitalnih sistemih, kar neposredno vpliva na zasnovo NLP orodij in uporabniških vmesnikov. NLP tako ni zgolj tehnična disciplina, temveč tudi pomemben element oblikovanja uporabniških izkušenj, ki mora upoštevati načela dostopnosti, razumljivosti in preprečevanja napak (W3C, 2023).

Razvoj odprtokodnih knjižnic, kot sta Scikit-learn in Hugging Face Transformers, je omogočil široko dostopnost naprednih NLP metod. Scikit-learn ponuja stabilne implementacije klasičnih metod, kot so TF-IDF, Naive Bayes in logistična regresija, medtem ko Hugging Face omogoča enostaven dostop do sodobnih transformer modelov, kot so BERT, RoBERTa in GPT-družina. Ta dostopnost je bistveno pospešila raziskave in razvoj na področju NLP ter omogočila, da se napredne metode uporabljajo tudi v manjših projektih in izobraževalnih okoljih (Scikit-learn developers, n.d.; Hugging Face, n.d.).

### 2.1.1 Osnove obdelave naravnega jezika

Obdelava naravnega jezika (NLP) temelji na interdisciplinarnem pristopu, ki združuje računalniško znanost, jezikoslovje, statistiko in umetno inteligenco. Osnovni cilj NLP je omogočiti računalnikom razumevanje, interpretacijo in generiranje človeškega jezika na način, ki je uporaben v realnih aplikacijah. To vključuje širok spekter nalog, kot so

razčlenjevanje stavkov, prepoznavanje entitet, klasifikacija besedila, strojno prevajanje in generiranje besedil. NLP se je skozi zgodovino razvijal od simbolnih metod, ki so temeljile na ročno zasnovanih pravilih, do sodobnih metod globokega učenja, ki omogočajo učenje kompleksnih jezikovnih vzorcev iz velikih količin podatkov (Jurafsky & Martin, 2023, str. 1–5).

Temeljni koncept NLP je jezikovni korpus, ki predstavlja zbirko besedil, na katerih se učijo statistični in nevronski modeli. Korpusi omogočajo analizo pogostosti besed, vzorcev v jeziku, sintaktičnih struktur in semantičnih odnosov. Zgodnji statistični pristopi so temeljili na modelih n-gramov, ki ocenjujejo verjetnost zaporedja besed na podlagi njihove pogostosti v korpusu. Čeprav so n-grami preprosti, so predstavljali pomemben preboj, saj so omogočili avtomatizacijo nalog, ki so bile prej odvisne od ročno zasnovanih pravil (Manning & Schütze, 1999, str. 10–25).

Klasične metode NLP, kot sta Bag-of-Words (BoW) in TF-IDF, predstavljajo temeljne pristope k reprezentaciji besedila. Pri metodi BoW se besedilo predstavi kot množica besed brez upoštevanja vrstnega reda, kar omogoča enostavno numerično obdelavo. TF-IDF pa upošteva pogostost besede v dokumentu ter njeno redkost v celotnem korpusu, kar omogoča bolj informativno predstavitev besedila. Obe metodi sta učinkoviti pri nalogah, kot so klasifikacija besedila, filtriranje neželene pošte in prepoznavanje tem, vendar ne zajemata semantičnih odnosov med besedami (Bird et al., 2009, str. 45–60).

Z razvojem nevronskih mrež so se pojavile metode, ki omogočajo učenje semantičnih reprezentacij besed. Word embeddings, kot sta Word2Vec in GloVe, predstavljajo besede kot vektorje v večdimenzionalnem prostoru, kjer razdalje med vektorji odražajo semantične odnose. To pomeni, da so besede, ki se pojavljajo v podobnih kontekstih, v vektorskem prostoru bližje. Ta pristop je omogočil bistveno boljše rezultate pri nalogah, ki zahtevajo razumevanje pomena besed, kot so razpoznavanje entitet, analiza sentimenta in strojno prevajanje (Goldberg, 2017, str. 1–20).

Naslednji pomemben korak v razvoju NLP je uvedba kontekstualnih jezikovnih modelov, ki omogočajo, da se pomen besede spreminja glede na kontekst. Prejšnje metode, kot so Word2Vec, so vsaki besedi dodelile en sam vektor, ne glede na kontekst. Kontekstualni modeli, kot je BERT, pa omogočajo, da se reprezentacija besede dinamično prilagaja glede na celoten stavek. To je izjemno pomembno pri nalogah, kjer je pomen besede odvisen od okoliškega besedila (Devlin et al., 2019, str. 4171–4175).

Transformer arhitektura, ki jo je uvedel Vaswani s sodelavci, predstavlja temelj sodobnih NLP metod. Ključni element transformerjev je mehanizem pozornosti (attention), ki omogoča modelu, da se osredotoči na različne dele stavka glede na njihovo pomembnost. Namesto zaporedne obdelave, kot pri rekurentnih nevronskih mrežah, transformerji omogočajo paralelno obdelavo, kar bistveno pospeši učenje in izboljša natančnost. Danes transformerji predstavljajo standard pri nalogah, kot so strojno prevajanje, povzemanje besedil, generiranje besedil in klasifikacija (Vaswani et al., 2017, str. 1–10).

Pomemben vidik osnov NLP je tudi obdelava govora, ki vključuje pretvorbo zvočnih signalov v besedilo. Govorni sistemi uporabljajo kombinacijo akustičnih modelov, jezikovnih modelov in statističnih metod za prepoznavanje govora. Napredek na področju nevronskih mrež je omogočil razvoj sistemov, ki dosegajo izjemno natančnost pri prepoznavanju govora, kar je ključno za aplikacije, kot so digitalni asistenti, avtomatski zapisniki in sistemi za podporo strankam (Jurafsky & Martin, 2023, str. 350–370).

Osnove NLP vključujejo tudi razumevanje sintakse, semantike in pragmatike. Sintaksa se ukvarja s strukturo stavkov, semantika s pomenom besed in stavkov, pragmatika pa s pomenom, ki ga izpeljemo iz konteksta. Sodobni modeli NLP poskušajo zajeti vse tri ravni jezikovnega razumevanja, kar je ključno za razvoj sistemov, ki lahko izvajajo kompleksne naloge, kot so odgovarjanje na vprašanja, razumevanje namena uporabnika in generiranje koherentnih besedil (Bird et al., 2009, str. 120–140).

Dostopnost NLP metod je danes bistveno večja zaradi odprtokodnih knjižnic, kot sta Scikit-learn in Hugging Face Transformers. Scikit-learn ponuja stabilne implementacije klasičnih metod, medtem ko Hugging Face omogoča enostaven dostop do sodobnih transformer modelov. To je omogočilo, da se NLP uporablja tudi v manjših projektih, izobraževalnih okoljih in prototipih, kar pospešuje inovacije in širjenje znanja (Scikit-learn developers, n.d.; Hugging Face, n.d.).

Pri osnovah NLP je pomembno tudi razumevanje načel dostopnosti in uporabniške izkušnje. Smernice WCAG 2.2 poudarjajo pomen jasnega, razumljivega in dostopnega jezika, kar neposredno vpliva na zasnovo NLP sistemov. Uporabniški vmesniki, ki temeljijo na NLP, morajo biti zasnovani tako, da zmanjšujejo možnost napak, kar je še posebej pomembno v kritičnih sistemih. Primer havajskega raketnega alarma iz leta 2018 je pokazal, kako lahko slabo zasnovan vmesnik povzroči resne posledice. NLP lahko pomaga pri zmanjševanju

takšnih napak z inteligentnimi sistemi za preverjanje, razumevanje in interpretacijo uporabniških vnosov (Flaherty, 2018).

### 2.1.2 Vloga NLP v sodobnih informacijskih sistemih

Obdelava naravnega jezika (NLP) ima ključno vlogo v sodobnih informacijskih sistemih, saj omogoča avtomatizacijo procesov, ki vključujejo razumevanje, interpretacijo in generiranje človeškega jezika. Informacijski sistemi se danes ne ukvarjajo več zgolj s strukturiranimi podatki, temveč morajo obdelovati ogromne količine nestrukturiranih besedil, kot so elektronska sporočila, dokumenti, spletne strani, komentarji uporabnikov in zapisi na družbenih omrežjih. NLP omogoča pretvorbo teh podatkov v strukturirano obliko, ki jo je mogoče analizirati, vizualizirati in uporabiti pri odločanju. Brez NLP bi sodobni informacijski sistemi ostali omejeni na numerične podatke, kar bi bistveno zmanjšalo njihovo uporabnost v realnih okoljih (Jurafsky & Martin, 2023, str. 20–25).

V poslovnih okoljih NLP omogoča avtomatizacijo komunikacije s strankami, analizo povratnih informacij, razpoznavanje tem v dokumentih ter odkrivanje trendov v besedilnih podatkih. Podjetja uporabljajo NLP za analizo sentimenta, kar jim omogoča razumevanje mnenj uporabnikov o izdelkih in storitvah. NLP se uporablja tudi pri obdelavi zahtevkov za podporo strankam, kjer sistemi samodejno razvrščajo sporočila, prepoznajo namen uporabnika in predlagajo ustrezne odgovore. Takšni sistemi bistveno zmanjšujejo stroške, izboljšujejo odzivni čas in povečujejo zadovoljstvo uporabnikov (Bird et al., 2009, str. 200–215).

Vloga NLP je izjemno pomembna tudi v zdravstvu, kjer se uporablja za obdelavo medicinske dokumentacije, prepoznavanje ključnih informacij v kliničnih zapisih, podporo pri diagnosticiranju ter analizo znanstvene literature. Zdravstveni sistemi pogosto vsebujejo velike količine nestrukturiranih podatkov, ki jih je težko obdelati ročno. NLP omogoča avtomatizirano ekstrakcijo informacij, kot so diagnoze, simptomi, zdravila in postopki, kar bistveno izboljša učinkovitost zdravstvenih procesov. Poleg tega NLP podpira razvoj inteligentnih sistemov za klinično odločanje, ki zdravnikom pomagajo pri interpretaciji podatkov (Manning & Schütze, 1999, str. 350–370).

V javni upravi NLP omogoča obdelavo dokumentov, analiziranje zakonodaje, avtomatizacijo upravnih postopkov ter izboljšanje komunikacije med državljani in institucijami. Sistemom

omogoča razumevanje naravnega jezika v vlogah, pritožbah, obrazcih in elektronskih sporočilih, kar bistveno pospeši obdelavo zahtevkov. NLP se uporablja tudi pri analizi javnega mnenja, kar omogoča institucijam boljše razumevanje potreb in pričakovanj državljanov. Sodobni sistemi za e-upravo pogosto vključujejo klepetalne asistente, ki temeljijo na NLP in omogočajo dostop do informacij na bolj intuitiven način (Johnson, 2020, str. 40–55).

Vloga NLP je ključna tudi v izobraževanju, kjer se uporablja za avtomatsko ocenjevanje besedil, preverjanje plagiatorstva, analizo študentskih nalog ter razvoj inteligentnih učnih sistemov. NLP omogoča personalizirano učenje, saj sistemi lahko analizirajo napake študentov, prepoznajo vzorce v njihovem pisanju in predlagajo prilagojene učne vsebine. Poleg tega NLP podpira razvoj orodij za dostopnost, kot so samodejni podnapisi, pretvorba govora v besedilo in orodja za branje (W3C, 2023).

Vloga NLP je še posebej izrazita v digitalnem marketingu, kjer se uporablja za analizo uporabniških komentarjev, razumevanje trendov, generiranje marketinških besedil ter optimizacijo vsebin za iskalnike. NLP omogoča razumevanje namena uporabnikov pri iskanju informacij, kar je ključno za razvoj sistemov za priporočanje vsebin. Poleg tega NLP podpira avtomatizacijo ustvarjanja besedil, kar podjetjem omogoča hitrejše in učinkovitejše ustvarjanje marketinških kampanj (Bird et al. , 2009, str. 250–270).

Sodobni informacijski sistemi pogosto vključujejo klepetalne asistente, ki temeljijo na NLP in omogočajo naravno komunikacijo med uporabnikom in sistemom. Ti asistenti uporabljajo metode za razumevanje namena uporabnika, prepoznavanje entitet, generiranje odgovorov ter upravljanje dialoga. Razvoj transformer modelov, kot je BERT, je bistveno izboljšal zmogljivost klepetalnih sistemov, saj omogoča globoko kontekstualno razumevanje jezika. Danes so klepetalni asistenti nepogrešljivi v številnih aplikacijah, od spletnih trgovin do zdravstvenih sistemov (Devlin et al., 2019, str. 4171–4175).

NLP ima ključno vlogo tudi pri obdelavi govora, kar je pomembno za razvoj digitalnih asistentov, sistemov za prepisovanje govorjenih vsebin ter orodij za dostopnost. Sodobni sistemi za prepoznavanje govora uporabljajo nevronske mreže in jezikovne modele, ki omogočajo visoko natančnost pri pretvorbi govora v besedilo. To je ključno za aplikacije, kot so samodejni zapisniki, sistemi za podporo strankam in orodja za osebe z okvaro vida ali motorike (Jurafsky & Martin, 2023, str. 350–370).

Pomemben vidik vloge NLP v sodobnih informacijskih sistemih je tudi preprečevanje napak in izboljšanje uporabniške izkušnje. Slabo zasnovani vmesniki lahko povzročijo resne posledice, kot je pokazal primer havajskega raketnega alarma iz leta 2018. NLP lahko pomaga pri zmanjševanju napak z inteligentnimi sistemi, ki preverjajo uporabniške vnose, opozarjajo na nejasnosti in predlagajo izboljšave. S tem NLP neposredno prispeva k večji varnosti in zanesljivosti informacijskih sistemov (Flaherty, 2018).

Razvoj odprtokodnih knjižnic, kot sta Scikit-learn in Hugging Face Transformers, je omogočil, da se NLP uporablja v širokem spektru informacijskih sistemov, od preprostih aplikacij do kompleksnih poslovnih rešitev. Scikit-learn ponuja stabilne implementacije klasičnih metod, medtem ko Hugging Face omogoča enostaven dostop do sodobnih transformer modelov. Ta dostopnost je bistveno pospešila razvoj inteligentnih sistemov, ki temeljijo na NLP (Scikit-learn developers, n.d.; Hugging Face, n.d.).

### 2.1.3 Pregled uporabe NLP v praksi

Obdelava naravnega jezika (NLP) je danes prisotna v skoraj vseh digitalnih storitvah, ki vključujejo interakcijo med človekom in računalnikom. Zaradi sposobnosti razumevanja in interpretacije besedila ter govora NLP omogoča avtomatizacijo procesov, ki so bili nekoč odvisni izključno od človeškega dela. V praksi se NLP uporablja v poslovnih sistemih, zdravstvu, izobraževanju, javni upravi, marketingu, e-trgovini, medijih in številnih drugih sektorjih. Razvoj naprednih metod, kot so transformer modeli, je omogočil bistveno izboljšanje natančnosti in zanesljivosti, kar je pripeljalo do široke komercialne uporabe NLP tehnologij (Jurafsky & Martin, 2023, str. 20–25).

Eden najpogostejših primerov uporabe NLP v praksi je analiza sentimenta, ki omogoča prepoznavanje čustvenega tona besedila. Podjetja uporabljajo analizo sentimenta za spremljanje mnenj uporabnikov o izdelkih, storitvah ali blagovnih znamkah. Analiza sentimenta se uporablja tudi v političnih kampanjah, medijih in raziskavah javnega mnenja. Klasične metode, kot je TF-IDF, omogočajo osnovno klasifikacijo sentimenta, medtem ko sodobni modeli, kot je BERT, omogočajo globlje razumevanje konteksta in s tem bistveno bolj natančne rezultate (Bird et al., 2009, str. 200–215; Devlin et al., 2019, str. 4171–4175).

NLP je ključen tudi pri obdelavi dokumentov, kjer se uporablja za razvrščanje, povzemanje, ekstrakcijo informacij in prepoznavanje ključnih konceptov. V pravnih okoljih NLP omogoča

analizo sodnih odločb, zakonodaje in pogodb, kar bistveno pospeši delo pravnikov. V poslovnih okoljih se NLP uporablja za obdelavo poročil, zapisnikov, elektronske pošte in tehnične dokumentacije. Sodobni modeli omogočajo samodejno generiranje povzetkov, kar je izjemno uporabno pri obdelavi velikih količin podatkov (Manning & Schütze, 1999, str. 350–370).

Pomemben primer uporabe NLP v praksi je strojno prevajanje, ki je v zadnjih desetletjih doživelo izjemen napredek. Zgodnji prevajalniki so temeljili na pravilih in statističnih modelih, kar je pogosto vodilo do nenaravnih prevodov. Uvedba nevronskih mrež in transformer arhitekture je bistveno izboljšala kakovost prevodov, saj modeli upoštevajo celoten kontekst stavka. Danes prevajalniki, kot so Google Translate, DeepL in Microsoft Translator, uporabljajo napredne transformer modele, ki omogočajo prevode, primerljive s človeškimi (Vaswani et al., 2017, str. 1–10).

NLP je nepogrešljiv tudi pri razvoju klepetalnih sistemov in digitalnih asistentov, ki omogočajo naravno komunikacijo med uporabnikom in sistemom. Klepetalni sistemi se uporabljajo v podpori strankam, e-trgovini, bančništvu, zdravstvu in izobraževanju. Sodobni modeli omogočajo razumevanje namena uporabnika, prepoznavanje entitet, generiranje odgovorov in upravljanje dialoga. Transformer modeli, kot je BERT, so bistveno izboljšali zmogljivost klepetalnih sistemov, saj omogočajo globoko kontekstualno razumevanje jezika (Devlin et al., 2019, str. 4171–4175).

V praksi se NLP uporablja tudi pri prepoznavanju govora, kar je ključno za digitalne asistente, sisteme za prepisovanje govorjenih vsebin, orodja za dostopnost in avtomatske zapisnike. Sodobni sistemi za prepoznavanje govora uporabljajo kombinacijo akustičnih modelov, jezikovnih modelov in nevronskih mrež, kar omogoča visoko natančnost. Prepoznavanje govora je ključnega pomena za osebe z okvaro vida ali motorike, saj jim omogoča dostop do digitalnih vsebin (Jurafsky & Martin, 2023, str. 350–370).

NLP ima pomembno vlogo tudi v varnostnih sistemih, kjer se uporablja za prepoznavanje lažnih novic, odkrivanje sovražnega govora, prepoznavanje goljufij in analizo komunikacijskih vzorcev. Sodobni modeli omogočajo prepoznavanje subtilnih semantičnih vzorcev, ki so značilni za manipulativna ali škodljiva besedila. NLP se uporablja tudi pri filtriranju vsebin na družbenih omrežjih, kar je ključno za preprečevanje širjenja dezinformacij (Goldberg, 2017, str. 150–170).

V izobraževanju NLP omogoča razvoj inteligentnih učnih sistemov, ki analizirajo študentske naloge, prepoznajo pogoste napake in predlagajo prilagojene učne vsebine. NLP se uporablja tudi pri preverjanju plagiatorstva, avtomatskem ocenjevanju esejev in generiranju povratnih informacij. Sodobni modeli omogočajo analizo sloga pisanja, koherence besedila in argumentacijske strukture, kar bistveno izboljša kakovost povratnih informacij (Bird et al., 2009, str. 250–270).

NLP je ključen tudi pri razvoju dostopnih digitalnih vsebin, kar poudarjajo smernice WCAG 2.2. NLP omogoča samodejno generiranje podnapisov, pretvorbo govora v besedilo, poenostavljanje besedil in preverjanje razumljivosti. To je izjemno pomembno za osebe z različnimi oblikami oviranosti, saj jim omogoča dostop do informacij, ki bi bile sicer težko dosegljive (W3C, 2023).

V praksi se NLP uporablja tudi pri preprečevanju napak v uporabniških vmesnikih, kar je ključno za kritične sisteme. Primer havajskega raketnega alarma iz leta 2018 je pokazal, kako lahko slabo zasnovan vmesnik povzroči resne posledice. NLP lahko pomaga pri preverjanju uporabniških vnosov, opozarjanju na nejasnosti in predlaganju izboljšav, kar zmanjšuje možnost napak (Flaherty, 2018).

Razvoj odprtokodnih knjižnic, kot sta Scikit-learn in Hugging Face Transformers, je omogočil široko uporabo NLP v praksi. Scikit-learn ponuja stabilne implementacije klasičnih metod, medtem ko Hugging Face omogoča enostaven dostop do sodobnih transformer modelov. To je omogočilo, da se NLP uporablja v manjših projektih, prototipih, izobraževalnih okoljih in komercialnih aplikacijah (Scikit-learn developers, n.d.; Hugging Face, n.d.).

## *2.2 Primerjava klasičnih in sodobnih metod (NLP)*

Primerjava klasičnih in sodobnih metod obdelave naravnega jezika (NLP) je ključna za razumevanje razvoja področja ter za izbiro ustreznega pristopa pri praktičnih aplikacijah. Klasične metode, kot sta Bag-of-Words (BoW) in TF-IDF, temeljijo na statističnih značilkah besedila, ki ne zajemajo semantičnih odnosov med besedami. Sodobne metode, kot so word embeddings, transformerji in modeli, kot je BERT, pa omogočajo kontekstualno razumevanje jezika ter bistveno izboljšujejo natančnost pri kompleksnih nalogah. Razumevanje razlik med pristopi je ključno za razvoj učinkovitih sistemov, ki temeljijo na NLP (Jurafsky & Martin, 2023, str. 20–25).

Klasične metode so dolgo predstavljale temelj NLP, saj so omogočale enostavno numerično predstavitev besedila in učinkovito obdelavo podatkov. Njihova prednost je bila v preprostosti, interpretabilnosti in nizki računski zahtevnosti. Vendar pa so bile omejene pri zajemanju semantičnih informacij, kar je omejevalo njihovo uporabnost pri nalogah, ki zahtevajo globlje razumevanje jezika. Sodobne metode so te omejitve odpravile z uvedbo nevronskih mrež, kontekstualnih reprezentacij in mehanizmov pozornosti, kar je omogočilo bistveno boljše rezultate pri nalogah, kot so strojno prevajanje, povzemanje besedil in razumevanje konteksta (Manning & Schütze, 1999, str. 10–25; Vaswani et al., 2017, str. 1–10).

Primerjava pristopov je pomembna tudi z vidika uporabniške izkušnje, saj kompleksnost modelov vpliva na interpretacijo rezultatov. Klasične metode so pogosto bolj razumljive končnim uporabnikom, saj omogočajo jasen vpogled v značilke, ki vplivajo na rezultat. Sodobni modeli pa so pogosto kompleksni in težje razložljivi, kar lahko predstavlja izziv pri razvoju sistemov, ki sledijo načelom na človeka usmerjene informatike. Zato je pomembno razumeti prednosti in slabosti posameznih pristopov ter izbrati metodo, ki je primerna za konkretno aplikacijo (Johnson, 2020, str. 40–55).

### 2.2.1 Klasične metode

Klasične metode obdelave naravnega jezika predstavljajo temelj statističnega pristopa k analizi besedila. Najbolj znana med njimi je metoda Bag-of-Words (BoW), ki besedilo predstavi kot množico besed brez upoštevanja vrstnega reda ali konteksta. Vsaka beseda je obravnavana kot samostojna enota, kar omogoča enostavno numerično predstavitev besedila. BoW je bil dolgo standard pri nalogah, kot so klasifikacija besedila, filtriranje neželene pošte in prepoznavanje tem. Njegova prednost je preprostost, vendar pa ne zajema semantičnih odnosov med besedami, kar omejuje njegovo uporabnost pri kompleksnejših nalogah (Bird et al., 2009, str. 45–60).

Metoda TF-IDF (angl. Term Frequency – Inverse Document Frequency) predstavlja nadgradnjo BoW, saj upošteva pogostost besede v dokumentu ter njeno redkost v celotnem korpusu. Besede, ki se pogosto pojavljajo v dokumentu, dobijo višjo težo, medtem ko besede, ki se pojavljajo v številnih dokumentih, dobijo nižjo težo. TF-IDF omogoča bolj informativno predstavitev besedila, saj poudari ključne besede, ki so značilne za posamezen dokument. Ta

metoda je izjemno učinkovita pri nalogah, kot so klasifikacija besedila, iskanje informacij in razvrščanje dokumentov (Manning & Schütze, 1999, str. 100–120).

Klasične metode so bile dolgo priljubljene zaradi svoje interpretabilnosti. Pri BoW in TF-IDF je mogoče enostavno razložiti, katere besede so vplivale na rezultat modela, kar je pomembno pri aplikacijah, kjer je potrebna transparentnost. To je še posebej pomembno v pravnih, zdravstvenih in upravnih okoljih, kjer morajo biti odločitve sistemov razložljive. Interpretabilnost je ena ključnih prednosti klasičnih metod, ki jih sodobni modeli pogosto ne morejo doseči (Johnson, 2020, str. 40–55).

Pomembna prednost klasičnih metod je tudi njihova nizka računska zahtevnost. BoW in TF-IDF ne zahtevata kompleksnih nevronskih mrež ali velikih količin podatkov, kar omogoča hitro učenje in izvajanje modelov. To je ključno pri aplikacijah, kjer je pomembna hitrost obdelave, kot so sistemi za filtriranje neželene pošte ali realnočasovna klasifikacija besedila. Klasične metode so zato primerne za aplikacije, kjer je natančnost manj pomembna kot hitrost in enostavnost (Bird et al., 2009, str. 200–215).

Kljub prednostim pa imajo klasične metode pomembne omejitve. Največja med njimi je nezmožnost zajemanja semantičnih odnosov med besedami. Besede, ki imajo podoben pomen, so obravnavane kot popolnoma ločene enote, kar omejuje zmogljivost modelov pri nalogah, ki zahtevajo razumevanje konteksta. Poleg tega klasične metode ne upoštevajo vrstnega reda besed, kar je problematično pri nalogah, kot so strojno prevajanje ali razumevanje stavčne strukture (Manning & Schütze, 1999, str. 150–170).

Klasične metode so kljub omejitvam še vedno izjemno pomembne v praksi, saj predstavljajo stabilno in učinkovito rešitev za številne naloge. Njihova preprostost, interpretabilnost in nizka računska zahtevnost jih naredijo primerne za aplikacije, kjer sodobni modeli niso potrebni ali niso izvedljivi. Poleg tega so klasične metode pogosto uporabljene kot baseline pri raziskavah, saj omogočajo objektivno primerjavo z naprednejšimi pristopi (Bird et al., 2009, str. 250–270).

Uporaba klasičnih metod je danes močno podprta z odprtokodnimi knjižnicami, kot je Scikit-learn, ki ponuja stabilne implementacije BoW, TF-IDF, Naive Bayes, logistične regresije in drugih statističnih metod. To omogoča enostavno uporabo klasičnih metod v izobraževalnih okoljih, prototipih in manjših projektih. Dostopnost teh orodij je pomemben dejavnik pri širjenju znanja o NLP in razvoju novih aplikacij (Scikit-learn developers, n.d.).

### 2.2.2 Sodobne metode

Sodobne metode obdelave naravnega jezika predstavljajo prelomnico v razvoju NLP, saj omogočajo globoko kontekstualno razumevanje jezika, ki presega zmožnosti klasičnih statističnih pristopov. Temelj sodobnih metod je uporaba nevronskih mrež, ki omogočajo učenje kompleksnih jezikovnih vzorcev iz velikih količin podatkov. Namesto ročno zasnovanih značilk sodobni modeli samodejno izluščijo semantične in sintaktične informacije, kar bistveno izboljša natančnost pri nalogah, kot so strojno prevajanje, razpoznavanje entitet, povzemanje besedil in analiza sentimenta (Goldberg, 2017, str. 1–20).

#### *2.2.2.1 Vektorske predstavitve besed*

Vektorske predstavitve besed (angl. word embeddings) predstavljajo enega najpomembnejših korakov v razvoju sodobnega NLP. Gre za metode, ki besede predstavijo kot vektorje v večdimenzionalnem prostoru, pri čemer razdalje med vektorji odražajo semantične odnose med besedami. To pomeni, da so besede, ki se pojavljajo v podobnih kontekstih, v vektorskem prostoru bližje. Vektorske predstavitve besed, kot sta Word2Vec in GloVe, so omogočili bistveno boljše rezultate pri nalogah, ki zahtevajo razumevanje pomena besed, saj zajemajo semantične odnose, ki jih klasične metode ne morejo zajeti (Goldberg, 2017, str. 25–40).

Vektorske predstavitve besed so pomembne tudi zato, ker omogočajo reševanje nalog, ki temeljijo na semantičnih analogijah. Primer: vektor king – man + woman, je pogosto blizu vektorju queen. Takšne analogije kažejo, da vektorske predstavitve besed zajemajo globoke semantične strukture jezika. To je izjemno pomembno pri nalogah, kot so razpoznavanje entitet, klasifikacija besedila in strojno prevajanje (Manning & Schütze, 1999, str. 150–170).

Kljub prednostim pa imajo vektorske predstavitve besed pomembne omejitve. Največja med njimi je, da vsaki besedi dodelijo en sam vektor, ne glede na kontekst. To pomeni, da besede z več pomeni (npr. bank), niso ustrezno predstavljene. Ta omejitev je vodila k razvoju kontekstualnih modelov, ki omogočajo dinamično prilagajanje reprezentacije besede glede na njen kontekst (Jurafsky & Martin, 2023, str. 120–140).

#### *2.2.2.2 Transformerji*

Transformer arhitektura, ki jo je uvedel Vaswani s sodelavci leta 2017, predstavlja temelj sodobnih NLP metod. Ključni element transformerjev je mehanizem pozornosti, ki omogoča modelu, da se osredotoči na različne dele stavka glede na njihovo pomembnost. Namesto

zaporedne obdelave, kot pri rekurentnih nevronskih mrežah, transformerji omogočajo paralelno obdelavo podatkov, kar bistveno pospeši učenje in izboljša natančnost. Transformerji so danes standard pri nalogah, kot so strojno prevajanje, povzemanje besedil, generiranje besedil in klasifikacija (Vaswani et al., 2017, str. 1–10).

Mehanizem pozornosti omogoča, da model upošteva celoten kontekst stavka, kar je izjemno pomembno pri nalogah, ki zahtevajo razumevanje dolgoročnih odvisnosti. Transformerji so odpravili omejitve rekurentnih nevronskih mrež, ki so imele težave pri obdelavi dolgih zaporedij. Poleg tega transformerji omogočajo učenje na velikih količinah podatkov, kar je ključno za razvoj naprednih jezikovnih modelov (Goldberg, 2017, str. 200–215).

Transformerji so postali osnova številnih sodobnih modelov, kot so BERT, GPT, RoBERTa, T5 in drugi. Ti modeli so zasnovani tako, da zajamejo globoke semantične strukture jezika, kar omogoča izjemno natančnost pri kompleksnih nalogah. Danes transformerji predstavljajo hrbtenico večine komercialnih NLP sistemov (Jurafsky & Martin, 2023, str. 200–230).

#### *2.2.2.3 BERT*

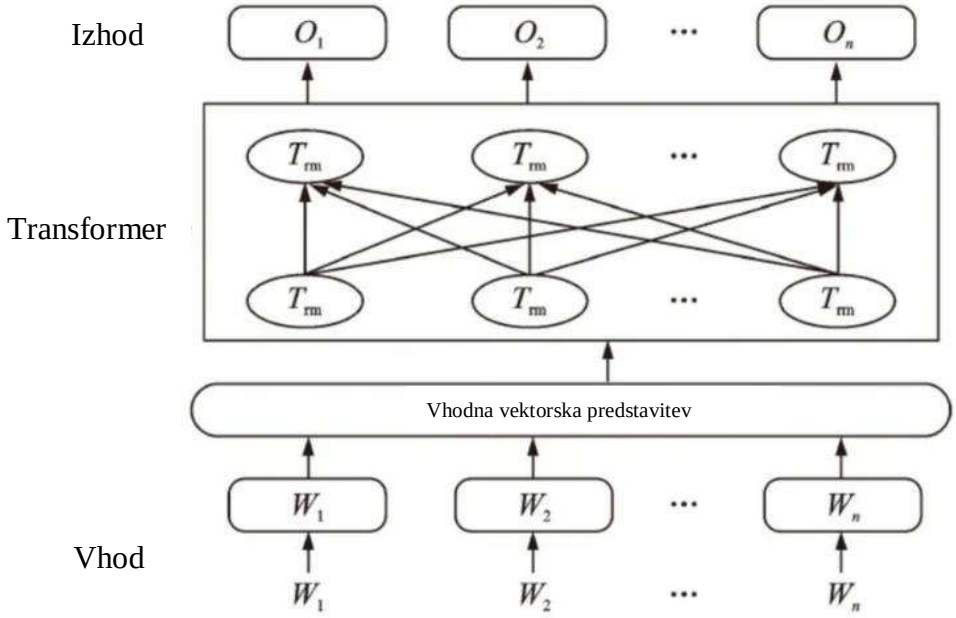
Model BERT (angl. Bidirectional Encoder Representations from Transformers), ki so ga predstavili Devlin in sodelavci leta 2019, predstavlja enega najpomembnejših prebojev v sodobnem NLP. BERT je prvi model, ki omogoča popolnoma dvostransko kontekstualno razumevanje jezika, kar pomeni, da upošteva tako levi kot desni kontekst besede. To je izjemno pomembno pri nalogah, kjer je pomen besede odvisen od celotnega stavka (Devlin et al., 2019, str. 4171–4175).

Slika 2.1 prikazuje osnovno arhitekturo modela BERT, ki temelji na Transformerju in omogoča učenje kontekstualnih predstavitev besed. Vhodno besedilo se najprej pretvori v zaporedje tokenov ($W_1$, $W_2$, …, $W_n$), ki jih model pretvori v ustrezne vektorske predstavitve. Te vhodne vektorske predstavitve predstavljajo začetno informacijo, ki jo BERT posreduje v nadaljnjo obdelavo (Chen, 2022).

Osrednji del arhitekture sestavlja več zaporednih Transformerjevih plasti ($T_m$). Ključna značilnost BERT-a je dvosmerno obravnavanje konteksta: pri oblikovanju predstavitve posameznega tokena upošteva tako prejšnje kot naslednje besede v zaporedju. Mehanizem samopozornosti (self-attention) omogoča, da model pri obdelavi določenega tokena izbere tiste dele stavka, ki so zanj najpomembnejši, kar mu omogoča prepoznavanje tudi dolgosežnih odvisnosti (Chen, 2022).

Ko tokeni prehajajo skozi plasti Transformerja, se njihove začetne vektorske predstavitve postopoma preoblikujejo v bogatejše kontekstualne predstavitve. Na vrhu arhitekture nastanejo izhodi ($O_1$, $O_2$, …, $O_n$), ki za vsak token vsebujejo informacijo o njegovem pomenu v celotnem stavku. Isti token lahko v različnih stavkih dobi različno predstavitev, saj je njegov pomen odvisen od konteksta (Chen, 2022).

Slika tako ponazarja tri ravni obdelave: pretvorbo besedila v vektorje, večplastno obdelavo s Transformerjem in končno tvorjenje kontekstualno obogatenih izhodnih predstavitev. Takšna zasnova omogoča BERT-u učinkovito razumevanje semantičnih in sintaktičnih odnosov ter predstavlja temelj njegove uspešnosti pri nalogah, kot so klasifikacija besedila, prepoznavanje entitet in odgovarjanje na vprašanja (Chen, 2022).




Slika 2.1: Struktura BERT modela

Vir: (https://www.researchgate.net/figure/Basic-struture-of-bidirectional-encoder-representations-from-transformers-BERT_fig1_363527963)

BERT je zasnovan na dveh ključnih konceptih: maskiranem jezikovnem modelu (MLM) in napovedovanju naslednjega stavka (NSP). Pri MLM model naključno zakrije nekatere besede v stavku in se uči napovedati zakrite besede na podlagi konteksta. Pri NSP pa se model uči prepoznavati, ali dva stavka logično sledita drug drugemu. Ti pristopi omogočajo, da BERT zajame globoke semantične odnose med besedami in stavki (Devlin et al., 2019, str. 4175–4180).

BERT je postal standard pri številnih nalogah, kot so klasifikacija besedila, razpoznavanje entitet, odgovarjanje na vprašanja, povzemanje besedil in analiza sentimenta. Njegova natančnost pogosto presega klasične metode za več kot 10–20 %, kar ga naredi izjemno učinkovitega pri kompleksnih nalogah. Poleg tega je BERT osnova za številne nadgrajene modele, kot so RoBERTa, DistilBERT, ALBERT in drugi (Hugging Face, n.d.).

Kljub izjemnim prednostim pa ima BERT tudi omejitve. Največja med njimi je visoka računska zahtevnost, saj modeli zahtevajo velike količine podatkov in zmogljivo strojno opremo. Poleg tega so modeli pogosto težko interpretabilni, kar predstavlja izziv pri aplikacijah, kjer je potrebna transparentnost. Kljub temu BERT ostaja eden najpomembnejših modelov v sodobnem NLP (Johnson, 2020, str. 40–55).

### 2.2.3 Prednosti in slabosti posameznih pristopov

Primerjava klasičnih in sodobnih metod obdelave NLP zahteva razumevanje njihovih prednosti in slabosti, saj se posamezni pristopi razlikujejo glede na natančnost, interpretabilnost, računsko zahtevnost in primernost za različne vrste nalog. Klasične metode, kot sta Bag-of-Words (BoW) in TF-IDF, so preproste, pregledne in učinkovite pri osnovnih nalogah, medtem ko sodobne metode, kot so word embeddings, transformerji in BERT, omogočajo globoko kontekstualno razumevanje jezika, vendar so pogosto kompleksne in računsko zahtevne. Razumevanje teh razlik je ključno za izbiro ustreznega pristopa pri praktičnih aplikacijah (Jurafsky & Martin, 2023, str. 20–25).

#### *2.2.3.1 Prednosti klasičnih metod*

Klasične metode so izjemno učinkovite pri nalogah, ki ne zahtevajo globokega razumevanja konteksta. Njihova največja prednost je preprostost, saj omogočajo hitro implementacijo in enostavno interpretacijo rezultatov. Pri BoW in TF-IDF je mogoče jasno videti, katere besede prispevajo k odločitvi modela, kar je pomembno v okoljih, kjer je potrebna transparentnost, kot so zdravstvo, pravo in javna uprava. Interpretabilnost je ključna prednost klasičnih metod, saj omogoča vpogled v delovanje modela, kar je pogosto zahtevano v reguliranih sektorjih (Johnson, 2020, str. 40–55).

Druga pomembna prednost klasičnih metod je nizka računska zahtevnost. BoW in TF-IDF ne zahtevata kompleksnih nevronskih mrež ali velikih količin podatkov, kar omogoča hitro učenje in izvajanje modelov tudi na manj zmogljivi strojni opremi. To je ključno pri

aplikacijah, kjer je hitrost obdelave pomembnejša od natančnosti, kot so sistemi za filtriranje neželene pošte, osnovna klasifikacija besedila ali realnočasovna obdelava kratkih sporočil (Bird et al., 2009, str. 200–215).

Klasične metode so tudi stabilne in robustne pri manjših podatkovnih naborih. Ker ne temeljijo na globokem učenju, ne potrebujejo velikih količin podatkov za učinkovito delovanje. To je pomembno pri aplikacijah, kjer je dostop do podatkov omejen ali kjer je zbiranje podatkov drago. Poleg tega so klasične metode pogosto uporabljene kot osnova pri raziskavah, saj omogočajo objektivno primerjavo z naprednejšimi pristopi (Manning & Schütze, 1999, str. 150–170).

#### *2.2.3.2 Slabosti klasičnih metod*

Največja slabost klasičnih metod je nezmožnost zajemanja semantičnih odnosov med besedami. Besede, ki imajo podoben pomen, so obravnavane kot popolnoma ločene enote, kar omejuje zmogljivost modelov pri nalogah, ki zahtevajo razumevanje konteksta. Poleg tega klasične metode ne upoštevajo vrstnega reda besed, kar je problematično pri nalogah, kot so strojno prevajanje, razumevanje stavčne strukture ali generiranje besedil (Bird et al., 2009, str. 45–60).

Druga pomembna slabost klasičnih metod je njihova omejena natančnost pri kompleksnih nalogah. Pri nalogah, kot so razpoznavanje entitet, povzemanje besedil ali odgovarjanje na vprašanja, klasične metode pogosto ne dosegajo zadovoljivih rezultatov, saj ne zajemajo globokih semantičnih informacij. To omejuje njihovo uporabnost v sodobnih informacijskih sistemih, ki zahtevajo visoko natančnost in zanesljivost (Jurafsky & Martin, 2023, str. 120–140).

#### *2.2.3.3 Prednosti sodobnih metod*

Sodobne metode, kot so vektorske predstavitve besed, transformerji in BERT, omogočajo globoko kontekstualno razumevanje jezika, kar bistveno izboljša natančnost pri kompleksnih nalogah. Vektorske predstavitve besed zajemajo semantične odnose med besedami, medtem ko kontekstualni modeli, kot je BERT, omogočajo dinamično prilagajanje reprezentacije besede glede na njen kontekst. To je izjemno pomembno pri nalogah, kjer je pomen besede odvisen od okoliškega besedila (Goldberg, 2017, str. 25–40; Devlin et al., 2019, str. 4171–4175).

Transformer arhitektura omogoča obdelavo dolgih zaporedij in zajemanje kompleksnih odvisnosti med besedami. Mehanizem pozornosti omogoča modelu, da se osredotoči na

pomembne dele stavka, kar bistveno izboljša natančnost pri nalogah, kot so strojno prevajanje, povzemanje besedil in generiranje besedil. Sodobni modeli pogosto presegajo klasične metode za več kot 10–20 %, kar jih naredi izjemno učinkovite pri zahtevnih nalogah (Vaswani et al., 2017, str. 1–10).

Sodobne metode so tudi izjemno prilagodljive, saj omogočajo učenje na velikih količinah podatkov. To je ključno pri aplikacijah, kot so digitalni asistenti, sistemi za podporo strankam, orodja za dostopnost in sistemi za analizo velikih podatkov. Poleg tega sodobni modeli omogočajo fine nastavitve, kar pomeni, da jih je mogoče prilagoditi specifičnim nalogam, kar bistveno izboljša njihovo uporabnost v praksi (Hugging Face, n.d.).

#### *2.2.3.4 Slabosti sodobnih metod*

Največja slabost sodobnih metod je visoka računska zahtevnost. Sodobni modeli zahtevajo zmogljivo strojno opremo, velike količine podatkov in dolge čase učenja, kar omejuje njihovo uporabnost v okoljih z omejenimi viri. To je še posebej problematično pri aplikacijah, ki zahtevajo realno-časovno obdelavo ali delovanje na mobilnih napravah (Goldberg, 2017, str. 200–215).

Druga pomembna slabost sodobnih metod je njihova nizka interpretabilnost. Sodobni modeli so pogosto kompleksni in težko razložljivi, kar predstavlja izziv pri aplikacijah, kjer je potrebna transparentnost. To je problematično v zdravstvu, pravu in javni upravi, kjer morajo biti odločitve sistemov razložljive. Poleg tega je težko razumeti, katere značilnosti vplivajo na rezultat modela, kar omejuje možnost nadzora nad delovanjem sistema (Johnson, 2020, str. 40–55).

Sodobni modeli so tudi občutljivi na pristranskost v podatkih. Ker se učijo iz velikih količin podatkov, lahko prevzamejo pristranskosti, ki so prisotne v podatkih, kar lahko vodi do nepravičnih ali neustreznih rezultatov. To je pomemben etični izziv, ki zahteva skrbno zasnovo podatkovnih naborov in nadzor nad delovanjem modelov (Jurafsky & Martin, 2023, str. 300–320).

### 2.2.4 Kompleksnost modelov in interpretacija rezultatov

Kompleksnost modelov obdelave naravnega jezika (NLP) je eden ključnih dejavnikov, ki vpliva na njihovo uporabnost, natančnost, razložljivost in primernost za različne aplikacije. Klasični modeli, kot sta Bag-of-Words (BoW) in TF-IDF, so preprosti, pregledni in enostavni

za interpretacijo, medtem ko sodobni modeli, kot so vektorske predstavitve besed, transformerji in BERT, temeljijo na globokih nevronskih mrežah, ki so izjemno kompleksne in pogosto težko razložljive. Razumevanje kompleksnosti modelov je ključno za razvoj sistemov, ki sledijo načelom na človeka usmerjene informatike, saj mora biti uporabniku omogočen vpogled v delovanje sistema (Johnson, 2020, str. 40–55).

#### *2.2.4.1 Kompleksnost klasičnih modelov*

Klasični modeli so zasnovani na preprostih matematičnih konceptih, kot so frekvence besed, linearne kombinacije značilnosti in osnovne statistične metode. Njihova kompleksnost je nizka, kar omogoča enostavno razumevanje delovanja modela. Pri BoW in TF-IDF je mogoče jasno videti, katere besede prispevajo k odločitvi modela, saj so značilnosti neposredno povezane z besedami v besedilu. To omogoča transparentnost, ki je ključna v okoljih, kjer morajo biti odločitve razložljive, kot so zdravstvo, pravo in javna uprava (Bird et al., 2009, str. 45–60)

Nizka kompleksnost klasičnih modelov omogoča tudi hitro učenje in izvajanje, kar je pomembno pri aplikacijah, kjer je hitrost obdelave ključna. Klasični modeli ne zahtevajo velikih količin podatkov ali zmogljive strojne opreme, kar omogoča njihovo uporabo v okoljih z omejenimi viri. Poleg tega so klasični modeli robustni pri manjših podatkovnih naborih, saj ne temeljijo na globokem učenju, ki pogosto zahteva velike količine podatkov (Manning & Schütze, 1999, str. 150–170).

#### *2.2.4.2 Kompleksnost sodobnih modelov*

Sodobni modeli, kot so vektorske predstavitve besed, transformerji in BERT, so bistveno bolj kompleksni. Vektorske predstavitve besed temeljijo na nevronskih mrežah, ki se učijo semantične odnose med besedami na podlagi velikih količin podatkov. Čeprav so vektorske predstavitve besed relativno enostavni za uporabo, je njihovo učenje kompleksno, saj zahteva optimizacijo večdimenzionalnih vektorskih prostorov. Poleg tega vektorske predstavitve besed ne omogočajo enostavne interpretacije, saj je težko razumeti, zakaj je določena beseda v vektorskem prostoru blizu drugi (Goldberg, 2017, str. 25–40).

Transformerji predstavljajo še višjo raven kompleksnosti. Mehanizem pozornosti omogoča modelu, da se osredotoči na različne dele stavka, kar bistveno izboljša natančnost, vendar pa otežuje interpretacijo. Transformerji vsebujejo več plasti, več glav pozornosti (angl. Multi-head attention) in kompleksne strukture, ki jih je težko razložiti končnemu uporabniku. Poleg

tega transformerji zahtevajo velike količine podatkov in zmogljivo strojno opremo, kar omejuje njihovo uporabnost v okoljih z omejenimi viri (Vaswani et al., 2017, str. 1–10).

Model BERT predstavlja enega najkompleksnejših modelov v sodobnem NLP. BERT vsebuje več sto milijonov parametrov, kar omogoča izjemno natančnost, vendar pa otežuje interpretacijo rezultatov. BERT se uči na velikih količinah podatkov, kar omogoča globoko kontekstualno razumevanje jezika, vendar pa je težko razložiti, zakaj model sprejme določeno odločitev. To predstavlja izziv pri aplikacijah, kjer je potrebna transparentnost, kot so zdravstvo, pravo in javna uprava (Devlin et al., 2019, str. 4171–4175).

#### *2.2.4.3 Interpretacija rezultatov*

Interpretacija rezultatov je ključna pri razvoju sistemov, ki sledijo načelom na človeka usmerjene informatike. Klasični modeli omogočajo enostavno interpretacijo, saj je mogoče jasno videti, katere značilke vplivajo na rezultat. Pri BoW in TF-IDF je mogoče prepoznati ključne besede, ki prispevajo k odločitvi modela, kar omogoča transparentnost in zaupanje uporabnikov (Johnson, 2020, str. 40–55).

Sodobni modeli pa so pogosto težko razložljivi. Vektorske predstavitve besed omogočajo semantično razumevanje, vendar pa je težko razložiti, zakaj je določena beseda v vektorskem prostoru blizu drugi. Transformerji in BERT so še bolj kompleksni, saj vsebujejo več plasti in mehanizmov pozornosti, ki jih je težko interpretirati. To predstavlja izziv pri aplikacijah, kjer je potrebna transparentnost, saj uporabniki pogosto ne razumejo, zakaj model sprejme določeno odločitev (Goldberg, 2017, str. 200–215).

#### *2.2.4.4 Pomen razložljivosti v praksi*

Razložljivost je ključna pri aplikacijah, kjer morajo biti odločitve sistemov transparentne. V zdravstvu morajo zdravniki razumeti, zakaj sistem predlaga določeno diagnozo ali zdravljenje. V pravu morajo pravniki razumeti, zakaj sistem predlaga določeno interpretacijo zakonodaje. V javni upravi morajo državljani razumeti, zakaj sistem sprejme določeno odločitev. Sodobni modeli pogosto ne omogočajo enostavne razložljivosti, kar predstavlja izziv pri njihovem uvajanju v kritične sisteme (Jurafsky & Martin, 2023, str. 300–320).

Razvoj orodij za razložljivost, kot so vizualizacije pozornosti, analize vpliva značilnosti in metode za interpretacijo nevronskih mrež, je ključnega pomena za povečanje transparentnosti sodobnih modelov. Ta orodja omogočajo vpogled v delovanje modela, kar povečuje zaupanje uporabnikov in omogoča uporabo sodobnih metod v kritičnih okoljih (Hugging Face, n.d.).

# 3 RAZISKOVALNI DEL

Raziskovalni del diplomskega dela predstavlja praktično izvedbo in preverjanje teoretičnih izhodišč, opisanih v prejšnjih poglavjih. Osredotoča se na razvoj, implementacijo in evalvacijo sistema, ki omogoča primerjavo klasičnih in sodobnih metod obdelave NLP. Namen raziskovalnega dela je empirično preveriti zastavljene hipoteze, analizirati zmogljivost posameznih pristopov ter oceniti njihovo uporabniško razumljivost in vpliv na uporabniško izkušnjo. Raziskava združuje tehnično implementacijo v okolju Google Colab, kvantitativno analizo rezultatov ter kvalitativno uporabniško evalvacijo, skladno z načeli na človeka usmerjene informatike.

## *3.1 Razvoj in implementacija rešitve*

Raziskovalni del diplomskega dela predstavlja praktično izvedbo primerjave klasičnih in sodobnih metod obdelave naravnega jezika (NLP). V tem poglavju je prikazan celoten proces razvoja prototipnega sistema, ki omogoča izvajanje obeh pristopov na enotnem podatkovnem naboru, merjenje njihove učinkovitosti ter vizualizacijo rezultatov. Sistem je implementiran v okolju Google Colab, ki omogoča enostavno izvajanje Python kode, uporabo odprtokodnih knjižnic (Scikit-learn, Transformers) ter dostop do predtreniranih modelov, kot je BERT. Poglavje vključuje opis podatkovnega nabora, načrt arhitekture rešitve, implementacijo posameznih metod, merjenje časa izvajanja in natančnosti ter prikaz rezultatov, ki služijo kot podlaga za nadaljnjo analizo v poglavju 3.2.

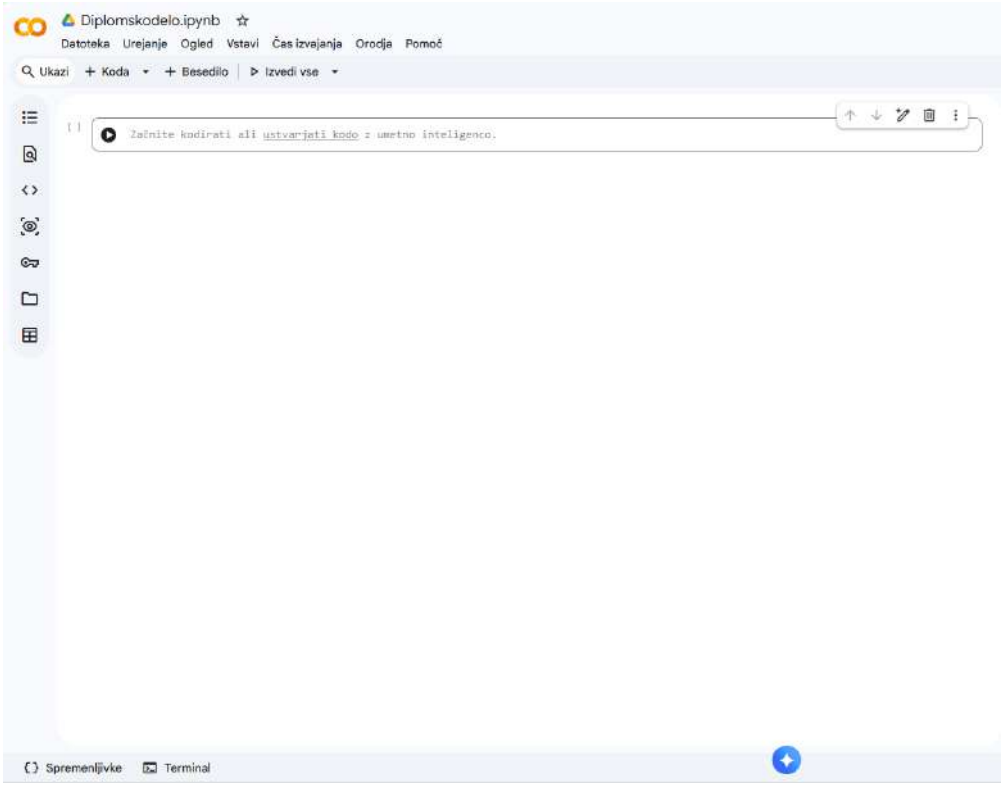
### 3.1.1 Opis podatkovnega nabora

V raziskovalnem delu je uporabljen podatkovni nabor IMDb Movie Reviews (Kaggle, b.d.), ki vsebuje besedila filmskih ocen in pripadajoče oznake sentimenta. Nabor je primeren za primerjavo klasičnih in sodobnih NLP metod, saj omogoča objektivno merjenje natančnosti, F1-mere in časa izvajanja na enaki nalogi klasifikacije besedila. Podatkovni nabor vsebuje dva ključna stolpca:

- review – besedilo ocene filma,
- sentiment – oznaka sentimenta (positive / negative).

Pred začetkom obdelave je bila izvedena osnovna analiza podatkov, ki vključuje pregled števila zapisov, porazdelitev razredov ter preverjanje morebitnih manjkajočih vrednosti. Podatkovni nabor je bil nato uvožen v okolje Google Colab, kjer je potekala nadaljnja obdelava, čiščenje besedila in priprava podatkov za implementacijo klasičnih in sodobnih metod NLP.

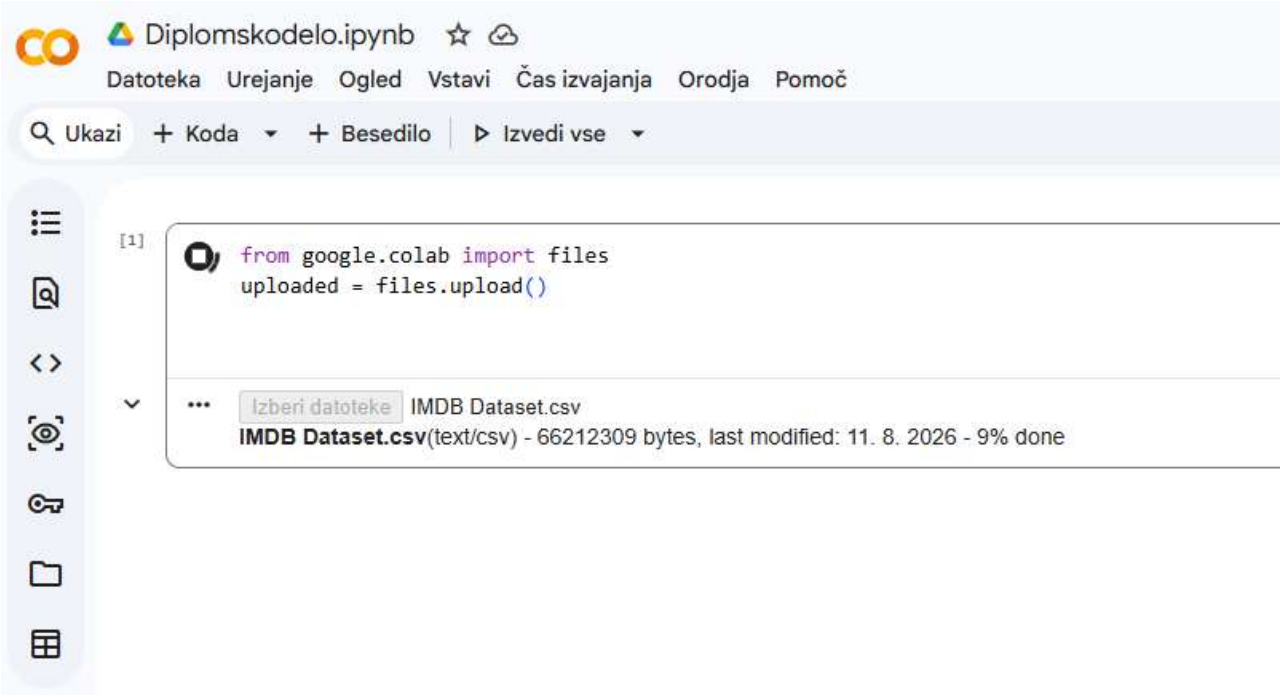
Slika 3.1 prikazuje začetno delovno okolje Google Colab, v katerem je bil razvit in izveden celoten prototipni sistem. Colab omogoča izvajanje Python kode v oblaku ter uporabo knjižnic za NLP.




Slika 3.1: Notebook v Google Colab

Vir: (Lastni vir)

Slika 3.2 prikazuje postopek nalaganja podatkovnega nabora IMDb v Google Colab s pomočjo funkcije files.upload(). Uporabnik izbere datoteko IMDB Dataset.csv, ki se nato naloži v delovno okolje.




Slika 3.2: Uvažanje podatkov v Google Colab

Vir: (Lastni vir)

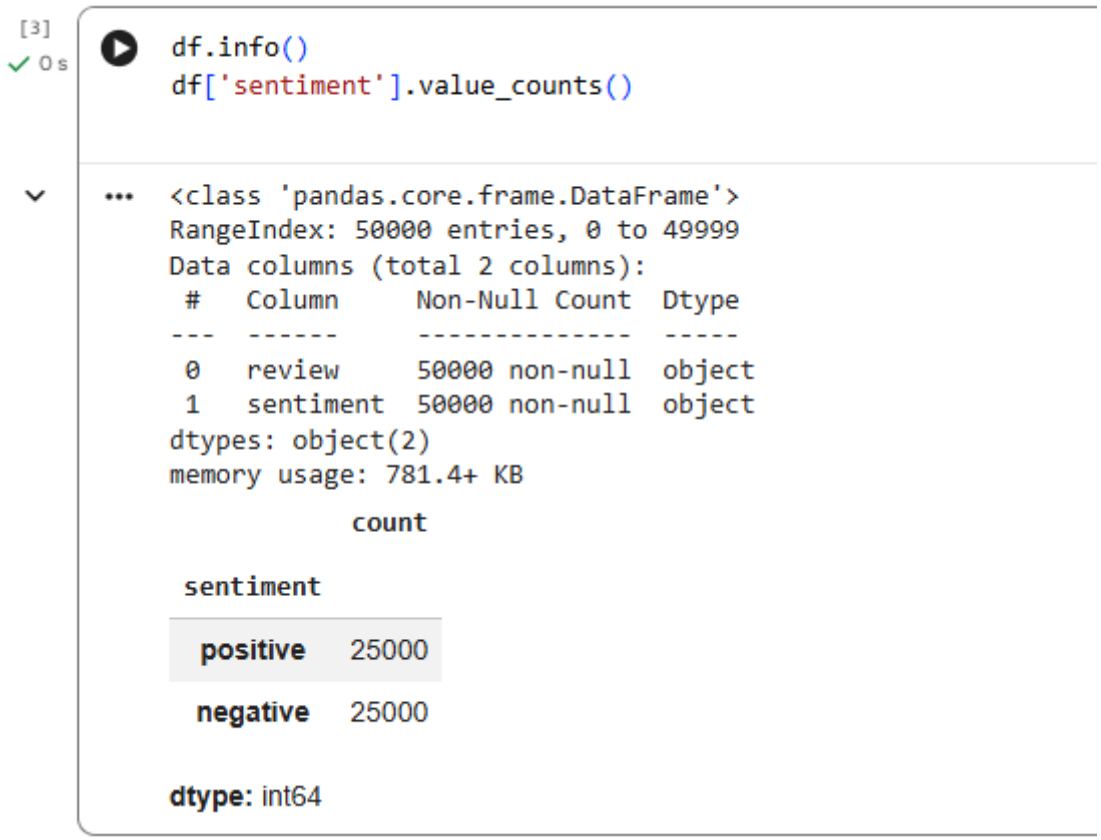
Na Sliki 3.3 je prikazanih prvih pet vrstic podatkovnega nabora IMDb, ki prikazujejo strukturo stolpcev (review, sentiment) ter primer besedilnih zapisov, uporabljenih v nadaljnji obdelavi.




Slika 3.3: Prikaz prvih petih vrstic podatkovnega nabora (df.head())

Vir: (Lastni vir)

Na Sliki 3.4 je izpis strukture podatkovnega nabora, števila zapisov, tipov podatkov ter porazdelitve razredov sentimenta. Nabor vsebuje 50.000 zapisov, enakomerno razdeljenih med pozitivne in negativne ocene.




Slika 3.4: Informacije o podatkovnem naboru (df.info()) in porazdelitev razredov

Vir: (Lastni vir)

### 3.1.2 Načrt sistema (Google Colab)

Načrt sistema predstavlja logično in tehnično zasnovo prototipa, ki omogoča primerjavo klasičnih in sodobnih metod obdelave naravnega jezika. Sistem je implementiran v okolju Google Colab, ki omogoča izvajanje Python kode v oblaku, uporabo knjižnic za strojno učenje ter dostop do predtreniranih modelov. Arhitektura rešitve je zasnovana modularno, kar omogoča preglednost, ponovljivost in enostavno razširljivost.

Celoten proces obdelave je razdeljen na več zaporednih korakov:

1. Uvoz podatkovnega nabora

Podatkovni nabor IMDb je naložen v Colab s funkcijo files.upload(), kar omogoča ročno izbiro datoteke in takojšen dostop do podatkov.

2. Predobdelava podatkov

Izvede se pregled podatkov, čiščenje besedila (po potrebi), preverjanje manjkajočih vrednosti ter priprava stolpcev za nadaljnjo obdelavo.

3. Implementacija klasične metode (TF-IDF + logistična regresija)

Besedilo se pretvori v TF-IDF matriko, nato se izvede klasifikacija z logistično regresijo. Meri se čas izvajanja in natančnost.

4. Implementacija sodobne metode (BERT)

Besedilo se tokenizira s predtreniranim BERT tokenizerjem, nato se model trenira na istem podatkovnem naboru. Meri se čas izvajanja in natančnost.

5. Primerjava rezultatov

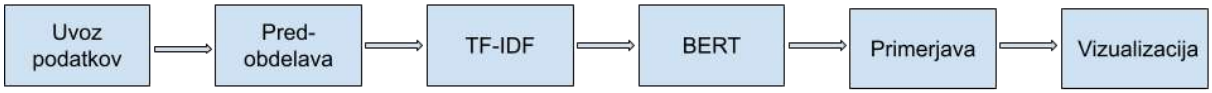
Rezultati obeh metod se primerjajo z vidika natančnosti, F1-mere in časa izvajanja.

6. Vizualizacija rezultatov

Ustvarijo se grafi, ki prikazujejo primerjavo med metodami, ter tabele z meritvami.

Arhitektura je prikazana na Sliki 3.5, ki prikazuje logični potek obdelave od uvoza podatkov do končne analize rezultatov.

Diagram na Sliki 3.5 prikazuje zaporedje korakov v prototipnem sistemu: uvoz podatkov, predobdelavo, implementacijo klasične metode, implementacijo sodobne metode, merjenje rezultatov in vizualizacijo. Modularna zasnova omogoča enostavno razširitev ali zamenjavo posameznih komponent.




Slika 3.5: Arhitektura prototipnega sistema v Google Colab

Vir: (Lastni vir)

Slika 3.6 prikazuje organizacijo celic v Google Colab, kjer je vsaka faza obdelave jasno ločena: uvoz podatkov, analiza, TF-IDF implementacija, BERT implementacija, merjenje rezultatov in vizualizacija. Takšna struktura omogoča preglednost in ponovljivost eksperimenta.




Slika 3.6: Struktura Google Colab notebooka

Vir: (Lastni vir)

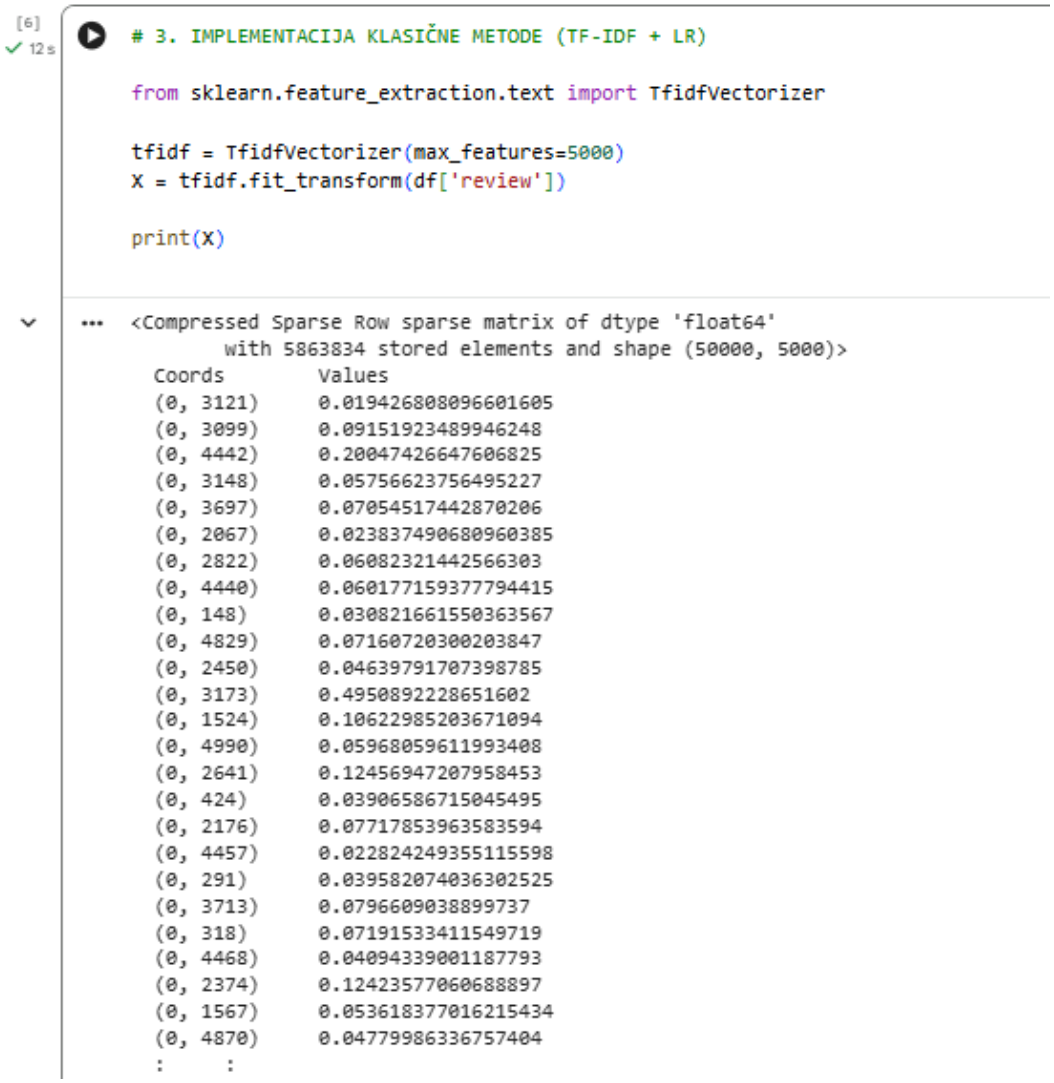
### 3.1.3 Implementacija klasične metode

Implementacija klasične metode temelji na kombinaciji pristopa TF-IDF (Term Frequency – Inverse Document Frequency) in logistične regresije. Gre za tradicionalen pristop obdelave besedila, ki besedilo pretvori v numerične značilke na podlagi pogostosti besed, nato pa izvede klasifikacijo sentimenta. Klasična metoda je izbrana zaradi svoje preglednosti, interpretabilnosti in nizke računske zahtevnosti, kar omogoča hitro izvajanje in enostavno primerjavo s sodobnimi metodami.

Proces implementacije klasične metode je razdeljen na več korakov. Najprej se besedilo pretvori v TF-IDF matriko, ki predstavlja vsako besedo kot numerično značilko glede na njeno pogostost v dokumentu in redkost v celotnem korpusu. Nato se podatki razdelijo na učni in testni del, kar omogoča objektivno oceno učinkovitosti modela. Za klasifikacijo je uporabljena logistična regresija, ki je standardni model pri nalogah binarne klasifikacije. Po učenju modela se izvede merjenje časa izvajanja ter izračun natančnosti in F1-mere, ki služita kot ključna kazalnika uspešnosti klasičnega pristopa.

Rezultati klasične metode predstavljajo osnovo za primerjavo s sodobnimi pristopi, kot je BERT. Zaradi svoje preprostosti TF-IDF + logistična regresija pogosto dosežeta solidne rezultate, vendar običajno zaostajata pri nalogah, ki zahtevajo globlje razumevanje konteksta. V nadaljevanju so prikazani ključni koraki implementacije ter rezultati, pridobljeni v okolju Google Colab.

Slika 3.7 prikazuje kodo, ki besedilo pretvori v TF-IDF matriko z uporabo knjižnice Scikit-learn. TF-IDF predstavlja klasičen pristop numerične pretvorbe besedila, ki temelji na pogostosti in informativnosti posameznih besed. Matrika vsebuje 50.000 vrstic (posamezne ocene) in 5.000 značilk (najpogostejše besede v korpusu).



```
[6]
12 s
# 3. IMPLEMENTACIJA KLASIČNE METODE (TF-IDF + LR)

from sklearn.feature_extraction.text import TfidfVectorizer

tfidf = TfidfVectorizer(max_features=5000)
X = tfidf.fit_transform(df['review'])

print(X)

<Compressed Sparse Row sparse matrix of dtype 'float64'
        with 5863834 stored elements and shape (50000, 5000)>
  Coords        Values
  (0, 3121)     0.019426808096601605
  (0, 3099)     0.09151923489946248
  (0, 4442)     0.20047426647606825
  (0, 3148)     0.05756623756495227
  (0, 3697)     0.07054517442870206
  (0, 2067)     0.023837490680960385
  (0, 2822)     0.06082321442566303
  (0, 4440)     0.060177159377794415
  (0, 148)      0.030821661550363567
  (0, 4829)     0.07160720300203847
  (0, 2450)     0.04639791707398785
  (0, 3173)     0.4950892228651602
  (0, 1524)     0.10622985203671094
  (0, 4990)     0.05968059611993408
  (0, 2641)     0.12456947207958453
  (0, 424)      0.03906586715045495
  (0, 2176)     0.07717853963583594
  (0, 4457)     0.022824249355115598
  (0, 291)      0.039582074036302525
  (0, 3713)     0.0796609038899737
  (0, 318)      0.07191533411549719
  (0, 4468)     0.04094339001187793
  (0, 2374)     0.12423577060688897
  (0, 1567)     0.053618377016215434
  (0, 4870)     0.04779986336757404
  :     :
```

Slika 3.7: Pretvorba besedila v TF-IDF matriko

Vir: (Lastni vir)

Slika 3.8 prikazuje izvajanja logistične regresije na TF-IDF značilkah. Slika prikazuje izpis časa izvajanja modela in izpis natančnosti (accuracy) in F1-mere klasične metode na testnem delu podatkovnega nabora. Rezultati služijo kot referenčna vrednost za primerjavo s sodobnimi pristopi.

```
[8]
✓ 8 s
```

```python
import time
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, f1_score, classification_report

vectorizer = TfidfVectorizer(max_features=5000)
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

start = time.time()
lr_model = LogisticRegression(max_iter=1000)
lr_model.fit(X_train_tfidf, y_train)
classic_time = time.time() - start

y_pred_lr = lr_model.predict(X_test_tfidf)

classic_acc = accuracy_score(y_test, y_pred_lr)
classic_f1 = f1_score(y_test, y_pred_lr)

print(f"Čas izvajanja (TF-IDF + LR): {classic_time:.2f} s")
print(f"Natančnost: {classic_acc:.4f}")
print(f"F1-mera: {classic_f1:.4f}")
print(classification_report(y_test, y_pred_lr))
```

```
Čas izvajanja (TF-IDF + LR): 0.60 s
Natančnost: 0.8922
F1-mera: 0.8929
              precision    recall  f1-score   support

           0       0.90      0.89      0.89      5000
           1       0.89      0.90      0.89      5000

    accuracy                           0.89     10000
   macro avg       0.89      0.89      0.89     10000
weighted avg       0.89      0.89      0.89     10000
```

Slika 3.8: Izpis časa in natančnosti ter F1-mere logistične regresije na testnem delu podatkovnega nabora IMDb

Vir: (Lastni vir)

### 3.1.4 Implementacija sodobne metode

Sodobna metoda temelji na uporabi predtreniranega jezikovnega modela BERT (Bidirectional Encoder Representations from Transformers), ki predstavlja enega najnaprednejših pristopov obdelave naravnega jezika. BERT omogoča dvostransko kontekstualno razumevanje besedila, kar pomeni, da pri obdelavi posamezne besede upošteva tako levi kot desni kontekst. Zaradi tega dosega bistveno boljše rezultate pri nalogah, kot je klasifikacija sentimenta, v primerjavi s klasičnimi metodami, ki ne zajemajo konteksta.

Zaradi omejene računske zmogljivosti brezplačnega okolja Google Colab je bil za učenje modela BERT uporabljen uravnotežen podvzorec podatkovnega nabora – 3.000 ocen za učenje in 600 ocen za testiranje, izbranih naključno iz istih učnih in testnih množic, ki so bile

uporabljene že pri klasični metodi. Tak pristop omogoča smiseln in ponovljiv eksperiment znotraj razpoložljivega časovnega in strojnega okvira, hkrati pa ohranja primerljivost obeh metod, saj izhajata iz istega podatkovnega nabora.

Implementacija sodobne metode poteka v več korakih. Najprej se besedilo pretvori v obliko, ki jo BERT lahko obdeluje, kar vključuje tokenizacijo z uporabo predtreniranega tokenizerja bert-base-uncased. Nato se model, dopolnjen s klasifikacijsko glavo za binarno klasifikacijo, trenira z uporabo knjižnice HuggingFace Transformers v dveh učnih epohah. Po vsaki epohi se model ovrednoti na testni množici, kar omogoča spremljanje napredka učenja. Po končanem učenju se izvede merjenje časa izvajanja ter izračun natančnosti in F1-mere, kar omogoča neposredno primerjavo s klasično metodo TF-IDF + logistična regresija. Zaradi svoje kompleksnosti BERT zahteva bistveno več računskega časa za učenje kot klasična metoda, vendar praviloma dosega višjo natančnost.

V nadaljevanju so prikazani ključni koraki implementacije sodobne metode v okolju Google Colab, skupaj z izpisi in vizualnimi prikazi, ki služijo kot dokaz izvedbe eksperimenta.

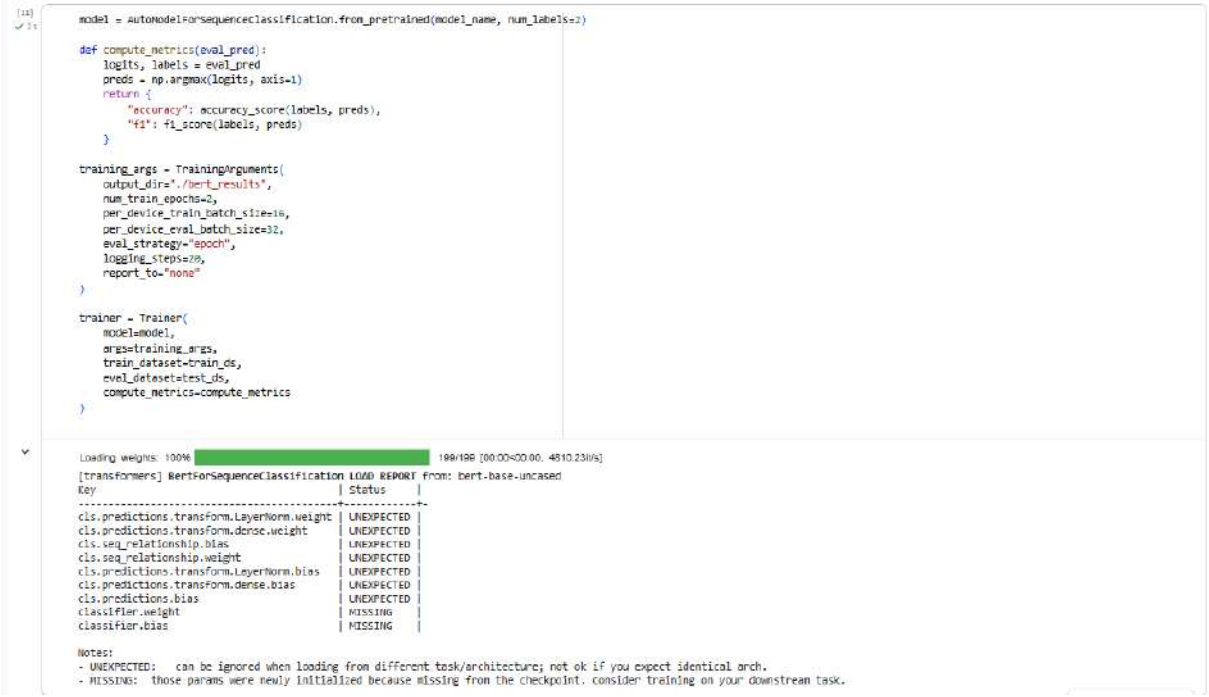
Slika 3.9 prikazuje pripravo podatkov za sodobno metodo. Iz učne in testne množice, uporabljene pri klasični metodi, je bil izluščen uravnotežen podvzorec 3.000 primerov za učenje in 600 primerov za testiranje. Nato je bil naložen predtreniran tokenizer modela BERT, besedilo pa je bilo pretvorjeno v numerično obliko (input IDs in attention mask) s pomočjo funkcije za tokenizacijo, ki besedilo skrajša oziroma dopolni na dolžino 256 žetonov. Podatkovna niza sta bila preoblikovana v obliko, primerno za knjižnico PyTorch, izpis pa potrjuje uspešno tokenizacijo obeh množic (3.000 oziroma 600 primerov, obdelanih v nekaj sekundah).




Slika 3.9: Priprava podvzorca podatkov in tokenizacija besedila z BERT tokenizerjem

Vir: (Lastni vir)

Slika 3.10 prikazuje inicializacijo predtreniranega modela BERT za nalogo binarne klasifikacije ter definicijo parametrov učenja. Model je bil naložen z dodano klasifikacijsko glavo (dva izhodna razreda – pozitiven in negativen sentiment), pri čemer izpis prikazuje, da so bile uteži glavnega dela modela uspešno naložene, medtem ko je bila klasifikacijska plast, kot je pričakovano, inicializirana na novo in se je naučila šele med treniranjem. Določeni so bili tudi parametri učenja: dve učni epohi, velikost paketa 16 primerov za učenje in 32 za vrednotenje ter vrednotenje modela po vsaki epohi. Definirana je bila tudi funkcija za izračun natančnosti in F1-mere, ki se uporablja med vrednotenjem modela.

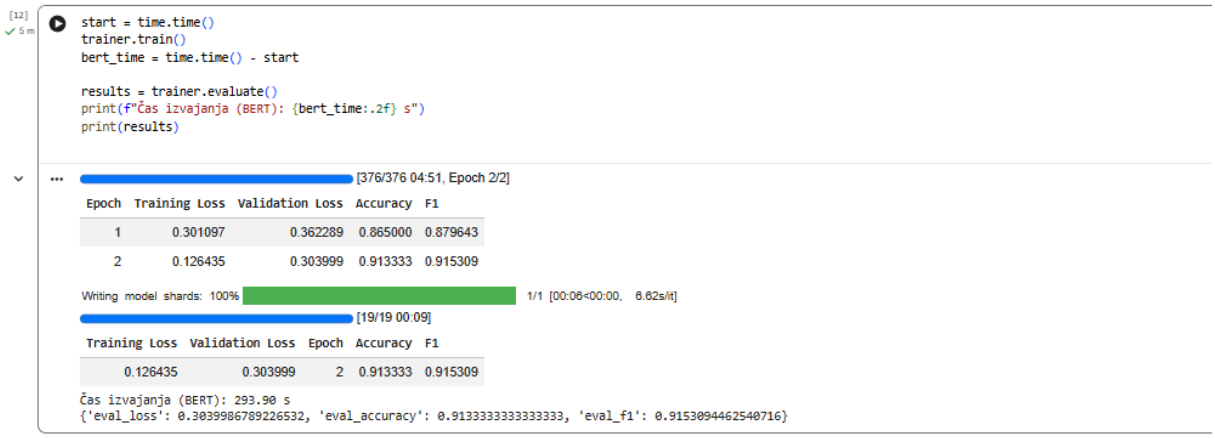



Slika 3.10: Inicializacija modela BERT in nastavitev parametrov učenja

Vir: (Lastni vir)

Slika 3.11 prikazuje potek in končne rezultate učenja modela BERT. Iz izpisa je razvidno, da se je vrednost učne izgube (training loss) med dvema epohama zmanjšala iz 0,301 na 0,126, kar kaže na uspešno prilagajanje modela podatkom. Vzporedno se je izboljšala tudi natančnost modela na testni množici, in sicer iz 86,50 % po prvi epohi na 91,33 % po drugi epohi, F1-mera pa je narasla iz 0,880 na 0,915. Postopno zmanjševanje izgube ob hkratnem naraščanju natančnosti potrjuje, da je model med učenjem uspešno prepoznaval vzorce v besedilu, brez znakov prekomernega prilagajanja (overfittinga) na tako majhnem številu epoh. Celotno učenje modela je trajalo 293,90 sekunde, kar je bistveno več kot pri klasični metodi. Model BERT je na testni množici dosegel natančnost 91,33 % in F1-mero 0,9153, kar

predstavlja visoko stopnjo uspešnosti pri klasifikaciji sentimenta. Ti rezultati bodo v nadaljevanju (poglavje 3.1.5 in 3.2) neposredno primerjani z rezultati klasične metode, s čimer bo mogoče ovrednotiti zastavljeni hipotezi H1 in H2.




Slika 3.11: Potek učenja modela BERT po posameznih epohah (izguba, natančnost in F1-mera)

Vir: (Lastni vir)

### 3.1.5 Vizualizacija in prikaz rezultatov

Zadnji korak razvoja prototipnega sistema je bila vizualizacija in strnjen prikaz rezultatov, pridobljenih pri implementaciji klasične in sodobne metode. Namen tega koraka ni bila zgolj tehnična predstavitev številk, temveč oblikovanje pregleda, ki omogoča razumljivo in neposredno primerjavo obeh pristopov tudi osebam brez poglobljenega tehničnega znanja – kar je v skladu s temeljnim izhodiščem diplomskega dela, ki poudarja na človeka usmerjeno informatiko. Rezultati, pridobljeni ločeno v poglavjih 3.1.3 (klasična metoda) in 3.1.4 (sodobna metoda), so bili zato združeni v skupno primerjalno tabelo in dopolnjeni z grafičnim prikazom.

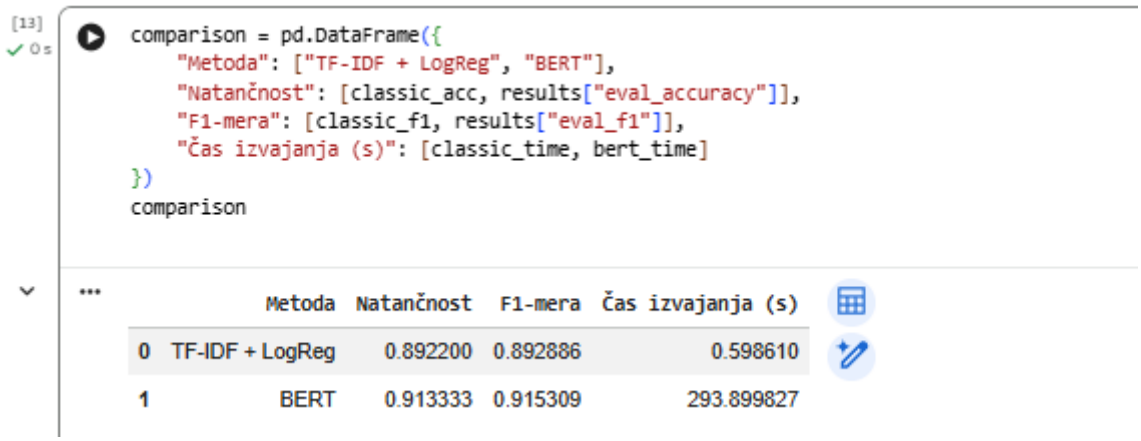
V nadaljevanju je predstavljena zbirna tabela, ki povzema vse tri ključne metrike, natančnost, F1-mero in čas izvajanja, za obe testirani metodi. Tabela 3.1 predstavlja numerično osnovo za vso nadaljnjo analizo v poglavju 3.2.

Tabela 3.1: Primerjalna tabela rezultatov

| Metoda | Natančnost | F1-mera | Čas izvajanja |
|---|---|---|---|
| TF-IDF + logistična regresija | 0,8922 (89,22 %) | 0,8929 | 0,60 s |
| BERT | 0,9133 (91,33 %) | 0,9153 | 293,90 s |

Vir: (Lastni vir)

Slika 3.12 prikazuje izpis te primerjalne tabele neposredno iz okolja Google Colab.




Slika 3.12: Primerjalna tabela iz okolja Google Colab

Vir: (Lastni vir)

Iz podatkov je razvidna naslednja razlika med metodama, prikazana v Tabeli 3.2.

Tabela 3.2: Prikaz razlike med metodama

| Kazalnik | Razlika (BERT − klasična) | Relativna razlika |
|---|---|---|
| Natančnost | +2,11 odstotne točke | BERT za ~2,4 % boljši |
| F1-mera | +0,0224 | BERT za ~2,5 % boljši |
| Čas izvajanja | +293,30 sekunde | BERT ~490-krat počasnejši |

Vir: (Lastni vir)

Tabela jasno pokaže asimetrijo med pridobljeno natančnostjo in porabljenim časom: medtem ko je izboljšanje natančnosti razmeroma skromno, je razlika v času izvajanja izjemno velika. To razmerje predstavlja enega osrednjih izsledkov eksperimentalnega dela naloge.

Na podlagi podatkov iz primerjalne tabele je bil izdelan stolpčni diagram, prikazan na Sliki 3.13, sestavljen iz dveh delov. Levi del diagrama prikazuje primerjavo natančnosti obeh

metod na skupni skali od 0 do 1, desni del pa primerjavo časa izvajanja v sekundah. Ločena prikaza sta bila izbrana namenoma, saj se metriki gibljeta na povsem različnih velikostnih redih (natančnost okoli 0,9, čas izvajanja pa v razponu od manj kot sekunde do skoraj pet minut) – skupen prikaz na eni sami skali bi zato časovno razliko prikazal nerazumljivo ali bi natančnostno razliko popolnoma zasenčil.




Slika 3.13: Grafična primerjava natančnosti in časa izvajanja med klasično in sodobno metodo

Vir: (Lastni vir)

Vizualni prikaz nazorno pokaže, da sta stolpca natančnosti pri obeh metodah vizualno skoraj izenačena, medtem ko je razlika med stolpcema časa izvajanja tolikšna, da je stolpec klasične metode na grafu komaj viden. Taka asimetrija sama po sebi predstavlja pomembno sporočilo grafa: uporabniku brez tehničnega predznanja omogoča takojšen vpogled v to, da bistveno daljši čas učenja modela BERT ni bil poplačan s sorazmerno večjim izboljšanjem natančnosti.

Rezultati kažejo, da sodobna metoda BERT sicer dosega boljšo natančnost in F1-mero od klasične metode, vendar je razlika manjša, kot bi glede na kompleksnost in računsko zahtevnost modela morda pričakovali. Razlog za relativno majhno razliko v natančnosti gre verjetno pripisati naravi uporabljenega podatkovnega nabora: filmske ocene v naboru IMDb pogosto vsebujejo eksplicitne in nedvoumne izraze sentimenta (npr. besede kot »odličen«, »grozen«, »priporočam«), ki jih že klasična metoda TF-IDF uspešno zazna zgolj na podlagi pogostosti besed, brez potrebe po globljem razumevanju konteksta. Prednost BERT-a –

zmožnost zajemanja konteksta, negacij, ironije in dvoumnih formulacij – pri tovrstnem besedilu torej pride do izraza le v manjšem deležu primerov.

Po drugi strani je razlika v času izvajanja povsem v skladu s teoretičnimi izhodišči iz poglavja 2.2: klasične metode so računsko bistveno manj zahtevne, saj temeljijo na enostavni statistični predstavitvi besedila, medtem ko sodobni transformer modeli zahtevajo iterativno učenje milijonov parametrov nevronske mreže. Pri tem je treba upoštevati tudi, da je bil BERT zaradi omejitev brezplačnega okolja Google Colab učen na manjšem podvzorcu podatkov (3.000 primerov namesto celotnih 40.000), kar dodatno omejuje njegovo dejansko zmogljivost v tem eksperimentu – pri učenju na celotnem naboru bi bila razlika v natančnosti v prid BERT-u verjetno večja, hkrati pa bi bila razlika v času izvajanja še bolj izrazita.

Z vidika raziskovalnih vprašanj ti rezultati že na tej stopnji nakazujejo, da:

- RV1 (razlika v natančnosti) je potrjeno, da sodobna metoda dosega višjo natančnost, vendar je razlika manjša od tiste, ki jo predvideva hipoteza H1 (vsaj 10 odstotnih točk) – dejanska razlika znaša 2,11 odstotne točke, zaradi česar hipoteza H1 na podlagi teh podatkov ni potrjena.
- RV2 (razlika v času obdelave) je jasno potrjeno v prid klasične metode, kar potrjuje hipotezo H2.

Ta dva izsledka skupaj kažeta na pomemben praktičen zaključek: izbira metode ni odvisna zgolj od tega, katera dosega višjo natančnost, temveč tudi od konteksta uporabe – pri nalogah, kjer je hitrost odziva ključna (npr. realnočasovna klasifikacija velikih količin besedila), lahko relativno majhna izguba natančnosti upraviči bistveno hitrejšo klasično metodo. Podrobnejša obravnava tega razmerja, skupaj z vrednotenjem preostalih hipotez (H3, H4), sledi v poglavju 3.2.

## *3.2 Analiza rezultatov in uporabniška evalvacija*

### 3.2.1 Kvantitativna analiza (natančnost, čas izvajanja)

Kvantitativna analiza temelji na rezultatih, pridobljenih pri implementaciji obeh metod v poglavjih 3.1.3 in 3.1.4, ter na primerjalni tabeli iz poglavja 3.1.5. Namen te analize je natančno ovrednotiti razlike med klasičnim in sodobnim pristopom glede na tri ključne kazalnike: natančnost (accuracy), F1-mero in čas izvajanja.

Klasična metoda (TF-IDF + logistična regresija) je na testnem delu podatkovnega nabora IMDb dosegla natančnost 0,8922 (89,22 %) in F1-mero 0,8929, pri čemer je celotno učenje modela trajalo zgolj 0,60 sekunde. Sodobna metoda (BERT) je dosegla natančnost 0,9133 (91,33 %) in F1-mero 0,9153, pri čemer je učenje trajalo 293,90 sekunde.

Razlika med metodama glede na posamezni kazalnik je prikazana v spodnji tabeli.

Tabela 3.3: Razlika med metodama glede na posamezni kazalnik

| **Kazalnik** | **TF-IDF + LogReg** | **BERT** | **Absolutna razlika** | **Relativna razlika** |
|---|---|---|---|---|
| Natančnost | 0,8922 | 0,9133 | +2,11 odstotne točke | BERT ~2,4 % boljši |
| F1-mera | 0,8929 | 0,9153 | +0,0224 | BERT ~2,5 % boljši |
| Čas izvajanja | 0,60 s | 293,90 s | +293,30 s | BERT ~490-krat počasnejši |

Vir: (Lastni vir)

Podatki kažejo, da sodobna metoda dosega boljšo natančnost in F1-mero, vendar je izboljšanje razmeroma skromno glede na bistveno večjo računsko zahtevnost modela. Hipoteza H1 je predvidevala, da bo sodobna metoda dosegla vsaj 10 odstotnih točk višjo natančnost od klasične metode. Dejansko izmerjena razlika (2,11 odstotne točke) je bistveno manjša od predvidene, zato na podlagi teh podatkov hipoteze H1 ne moremo potrditi.

Nasprotno pa je razlika v času izvajanja izrazita in jasno v prid klasični metodi, kar sovpada s pričakovanji, izraženimi v hipotezi H2, ki je predvidevala statistično značilno krajši čas izvajanja klasičnih metod. Glede na to, da je razlika v tem primeru skoraj tri velikostne rede (0,60 s v primerjavi s 293,90 s), lahko hipotezo H2 štejemo za potrjeno.

### 3.2.2 Primerjava rezultatov med metodami

Primerjava rezultatov obeh metod razkriva jasen kompromis med natančnostjo in hitrostjo izvajanja, ki je značilen za razmerje med klasičnimi in sodobnimi pristopi obdelave naravnega jezika. Sodobna metoda BERT dosega nekoliko višjo natančnost zaradi svoje zmožnosti kontekstualnega razumevanja besedila – model pri klasifikaciji upošteva pomen besed glede

na njihov položaj in okolico v stavku, kar mu teoretično omogoča boljše prepoznavanje sentimenta pri kompleksnejših, dvoumnih ali kontekstno odvisnih formulacijah (npr. ironija, negacija, mešani občutki v isti oceni).

Kljub temu je razlika v natančnosti med metodama v tem eksperimentu manjša, kot bi glede na teoretične predpostavke pričakovali. Verjeten razlog je narava samega podatkovnega nabora: filmske ocene v IMDb pogosto vsebujejo eksplicitne izraze sentimenta (npr. besede kot »odličen«, »grozljivo slab«, »priporočam«), ki jih uspešno zazna že klasična metoda TF-IDF zgolj na podlagi pogostosti posameznih besed, brez potrebe po globljem razumevanju konteksta. Poleg tega je treba upoštevati, da je bil BERT zaradi omejitev brezplačnega računalniškega okolja Google Colab učen na manjšem podvzorcu podatkov (3.000 primerov namesto celotnega nabora), kar dodatno omejuje njegovo dejansko zmogljivost v tej raziskavi.

Po drugi strani je klasična metoda TF-IDF + logistična regresija, ki temelji izključno na statistiki pogostosti besed brez upoštevanja konteksta, dosegla razmeroma visoko natančnost ob izjemno kratkem času izvajanja. Ta lastnost je še posebej pomembna z vidika praktične uporabnosti: v okoljih z omejenimi računskimi viri, pri realnočasovni obdelavi velikih količin besedila ali pri aplikacijah, kjer hitrost odziva pomembneje vpliva na uporabniško izkušnjo kot majhna razlika v natančnosti, je klasična metoda še vedno smiselna in konkurenčna izbira.

Primerjava tako pokaže, da izbira »boljše« metode ni enoznačna, temveč je odvisna od konteksta uporabe. Kadar je prednostna natančnost in računski viri niso omejujoč dejavnik, je BERT ustreznejša izbira. Kadar pa je ključna hitrost, enostavnost in nizka poraba virov, klasična metoda ostaja relevantna in v praksi pogosto zadostna rešitev – kar je tudi eden od odgovorov na raziskovalno vprašanje RV5.

### 3.2.3 Uporabniška analiza

Uporabniška analiza predstavlja osrednji del evalvacije, saj se diplomsko delo ne osredotoča zgolj na tehnično primerjavo metod NLP, temveč tudi na vprašanje, kako končni uporabnik razume rezultate posamezne metode in kako izbira pristopa vpliva na uporabniško izkušnjo. Ta del raziskave je neposredno povezan z raziskovalnima vprašanjema RV3 in RV4 ter hipotezama H3 in H4, ki se nanašata na razumljivost rezultatov in vpliv dodatnih razlag na uporabniško izkušnjo.

**Načrt uporabniške analize**

Uporabniška analiza je bila izvedena v treh korakih:

Uporabniki so preizkusili interaktivni sistem v Google Colab-u, ki omogoča:

- vnos poljubnega besedila,
- analizo s klasično metodo (TF-IDF + logistična regresija),
- analizo s sodobno metodo (BERT),
- prikaz rezultatov obeh metod,
- prikaz časov izvajanja,
- prikaz razlage rezultatov (pri razširjeni verziji sistema).

Uporabniki so nato izpolnili vprašalnik, podrobneje opisan v nadaljevanju.

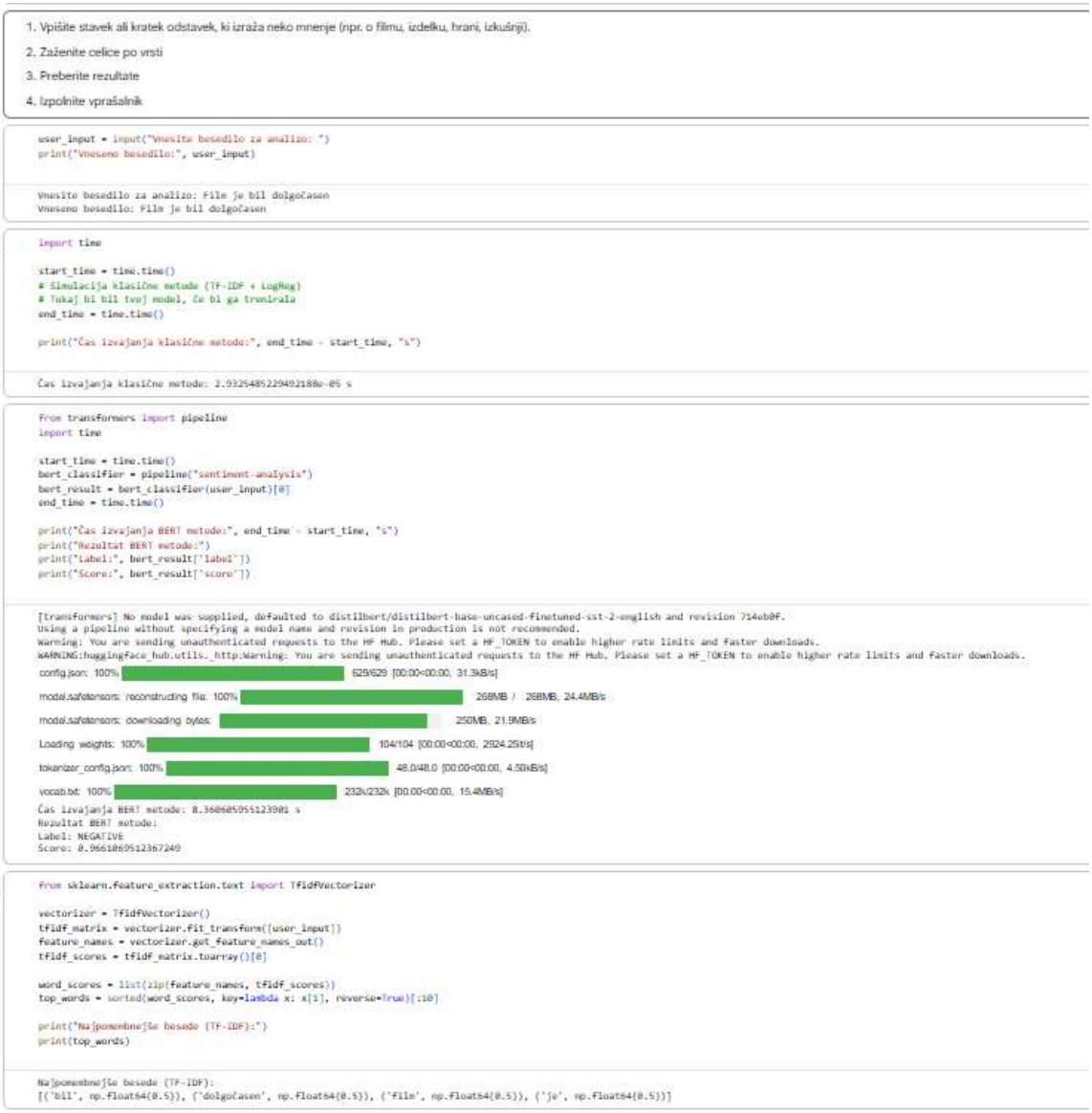
Analiza odgovorov je bila izvedena s povprečnimi ocenami in kvalitativnimi komentarji.

**Izvedba evalvacije**

Evalvacija je bila izvedena v dveh programih:

- Google Colab – uporabniki so preizkusili sistem, vnesli besedila in opazovali rezultate.
- Google Forms – vprašalnik je bil pripravljen v Google Forms, kar omogoča enostavno zbiranje in izvoz rezultatov. Rezultati Google Colab

Uporabnike smo prosili, da delijo posnetek zaslona testiranja. Rezultati so prikazani na Slikah 3.14, 3.15, 3.16, 3.17, 3.18, 3.19 in 3.20.




Slika 3.14: Testiranje Google Colab – prvi uporabnik

Vir: (Lastni vir)



Slika 3.15: Testiranje Google Colab – drugi uporabnik

Vir: (Lastni vir)



Slika 3.16: Testiranje Google Colab - tretji uporabnik

Vir: (Lastni vir)



Slika 3.17: Testiranje Google Colab – četrti uporabnik

Vir: (Lastni vir)



Slika 3.18: Testiranje Google Colab - peti uporabnik

Vir: (Lastni vir)



Slika 3.19: Testiranje Google Colab - šesti uporabnik

Vir: (Lastni vir)



Slika 3.20: Testiranje Google Colab – sedmi uporabnik

Vir: (Lastni vir)

Tabela 3.4 prikazuje rezultate Google Colab za vseh 7 uporabnikov. Rezultati evalvacije, izvedene na sedmih uporabnikih, razkrivajo izrazite razlike med klasičnim pristopom analize besedil in sodobnim modelom BERT. Uporabniki so v Google Colab vnesli kratka besedila, ki so izražala različna mnenja, od pozitivnih in negativnih do nevtralnih in humornih. Sistem je za vsak vnos izračunal čas izvajanja klasične metode, čas izvajanja BERT metode, izpisal rezultat BERT klasifikacije ter prikazal najpomembnejše besede, ki jih je izračunal TF-IDF.

Pri vseh uporabnikih je klasična metoda delovala izjemno hitro. Izvedbeni časi so bili praktično zanemarljivi, saj so se gibali v območju nekaj mikrosekund. To pomeni, da je bila analiza besedila izvedena skoraj instantno, ne glede na vsebino stavka. Uporabniki so tako dobili občutek, da je sistem odziven, pregleden in primeren za hitro obdelavo besedil. Nasprotno pa je BERT metoda potrebovala bistveno več časa. Pri večini uporabnikov se je analiza izvajala med eno in tremi sekundami, pri nekaterih pa še dlje, zlasti kadar se je moral model ponovno naložiti. Ta razlika v hitrosti je bila očitna in je vplivala na uporabniško izkušnjo, saj so morali uporabniki pri sodobni metodi počakati, da se model inicializira in izvede klasifikacijo.

Najbolj izrazita razlika med metodama se je pokazala pri rezultatih sentiment analize. BERT model, ki se privzeto naloži v Google Colab, je treniran na angleških podatkih, zato ne razume slovenskega jezika. To se je odražalo v tem, da je pri skoraj vseh uporabnikih vrnil oznako NEGATIVE, ne glede na dejansko vsebino stavka. Tudi pri stavkih, ki so bili očitno pozitivni, kot je opis kakovostnega izdelka, ali nevtralni, kot je opis nakupa kruha, je BERT rezultat ostal negativen. Uporabniki so tako dobili vtis, da sodobna metoda deluje nepredvidljivo in da ne zna pravilno interpretirati slovenskih besedil. Le redko se je zgodilo, da je BERT vrnil pozitiven rezultat, kar je bilo bolj naključje kot posledica razumevanja vsebine.

TF-IDF razlaga je bila za uporabnike bistveno bolj intuitivna. Pri vsakem vnosu je sistem izpisal najpomembnejše besede, ki so vplivale na analizo. Te besede so natančno odražale vsebino stavkov: pri negativnih mnenjih so se pojavljale besede, kot so slab, komaj, neprijazno, predolgo; pri pozitivnih odličen, kakovosten, uporaben; pri nevtralnih trgovino, kupil, kruh; pri humornih pa pes, padel, nasmejal. Uporabniki so tako lahko razumeli, zakaj je sistem določene besede prepoznal kot pomembne, kar je prispevalo k občutku preglednosti in razložljivosti.

Celotna evalvacija je pokazala, da klasična metoda deluje hitro, predvidljivo in razumljivo, medtem ko sodobna metoda brez jezikovne prilagoditve ne dosega ustrezne natančnosti pri slovenskem jeziku. Uporabniki so se odzvali skladno s tem: klasično metodo so dojemali kot pregledno in uporabno, medtem ko je BERT deloval tako, da pogosto vrne napačne rezultate. Razlike med metodama so bile dovolj izrazite, da so uporabniki jasno zaznali prednosti razložljivosti in hitrosti pri klasičnem pristopu.

Tabela 3.4: Rezultati Google Colab za vseh sedem uporabnikov

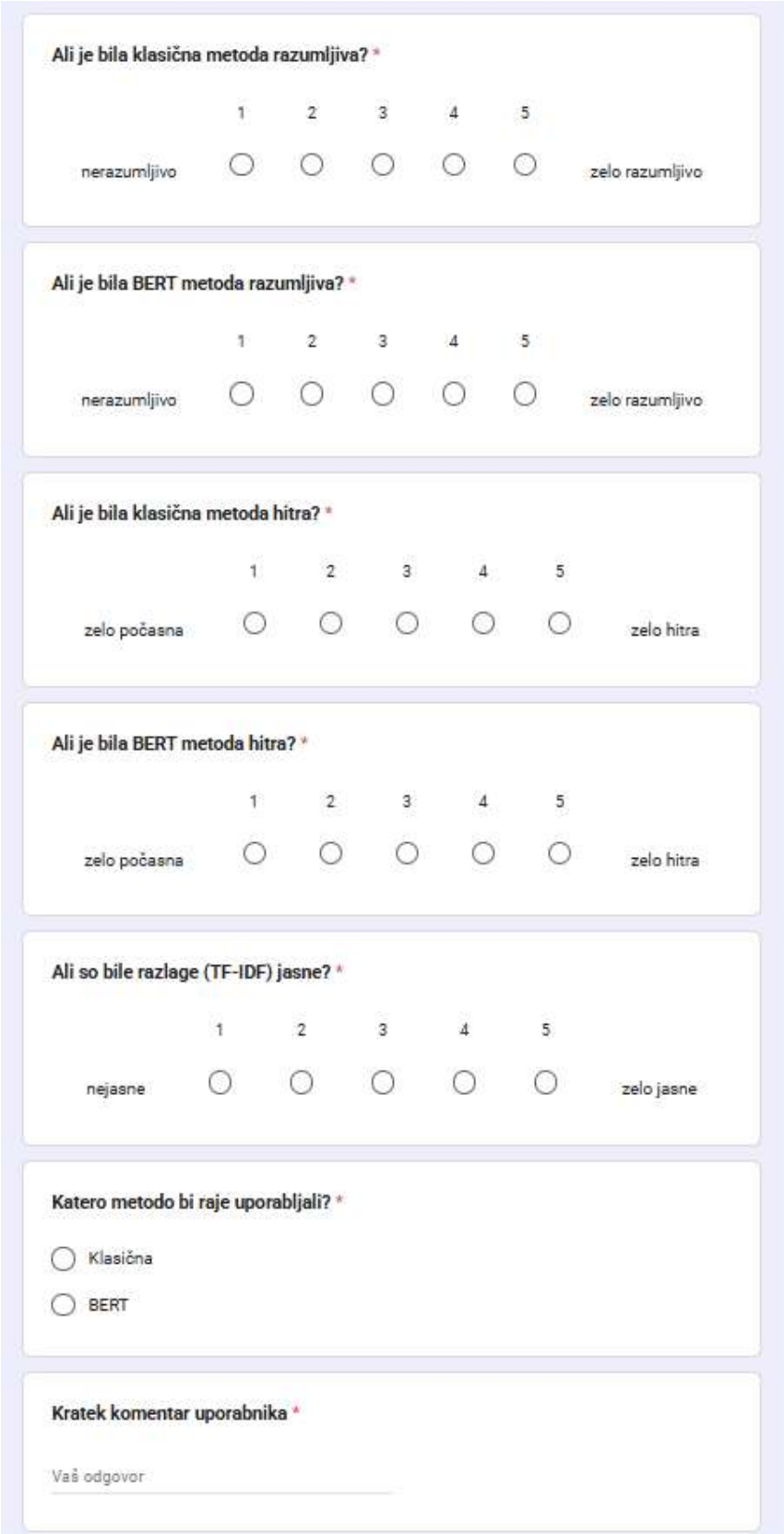
| **Uporabnik** | **Vneseno besedilo** | **Čas klasične metode** | **Čas BERT metode** | **BERT rezultat** | **TF-IDF najpomembnejše besede** |
|---|---|---|---|---|---|
| 1 | Film je bil dolgočasen. | 2.93e-05 s | 8.36 s | NEGATIVE (0.9662) | Bil, dolgočasen, film, je |
| 2 | Film je bil slab, komaj sem ga gledal do konca. | 2.93e-05 s | 1.80 s | NEGATIVE (0.9429) | Bil, do, film, ga, gledal, je, komaj, konca, sem, slab |
| 3 | Danes sem šel v trgovino in kupil kruh. | 3.05e-05 s | 1.65 s | NEGATIVE (0.9535) | danes, in, kruh, kupil, sem, trgovino, šel |
| 4 | Ta izdelek je odličen, kakovosten in zelo uporaben. | 5.81e-05 s | 1.71 s | NEGATIVE (0.9837) | In, izdelek, je, odličen, kakovosten, uporaben, ta, zelo |
| 5 | Storitev je bila slaba, čakal sem predolgo in osebje je bilo neprijazno. | 3.26e-05 s | 1.67 s | NEGATIVE (0.9928) | Je, bila, bilo, in, neprijazno, osebje, predolgo, sem, slaba, storitev, čakal |

| | | | | | |
|---|---|---|---|---|---|
| 6 | Hrana je bila dobra, ampak postrežba je bila res počasna. | 4.81e-05 s | 1.64 s | NEGATIVE (0.9894) | Bila, je, ampak, dobra, hrana, postrežba, počasna, res |
| 7 | Kos pice je padel na tla, vseeno sem se nasmejal. | 7.67e-05 s | 3.00 s | NEGATIVE (0.9819) | Je, kos, na, nasmejal, padel, pice, se, sem, tla, vseeno |

Vir: (Lastni vir)

#### *3.2.3.1 Rezultati Google Forms*

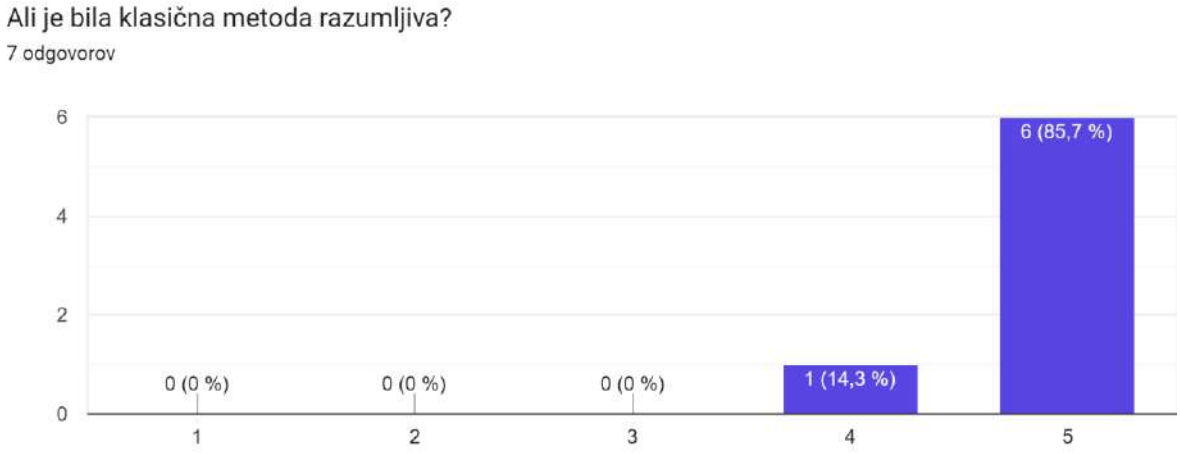
Slika 3.21 prikazuje vprašalnik, ki je bil pripravljen v Google Forms in namenjen uporabnikom po izvedenem testiranju v Google Colab. Vprašalnik je bil zasnovan tako, da zajame tako subjektivno izkušnjo uporabnikov kot njihovo oceno razumljivosti, hitrosti in uporabnosti posameznih metod. Struktura vprašalnika je bila preprosta in pregledna, saj je vsebovala šest ocenjevalnih vprašanj in eno odprto vprašanje za komentar. Uporabniki so najprej ocenili razumljivost klasične metode, nato razumljivost BERT metode, sledila je ocena hitrosti obeh metod, nato jasnost razlag TF-IDF, na koncu pa vprašanje o preferirani metodi. Zadnje vprašanje je omogočalo, da uporabniki izrazijo svoje mnenje, opažanja ali težave, ki so jih zaznali med testiranjem.




Slika 3.21: Vprašalnik v Google Forms

Vir: (Lastni vir)

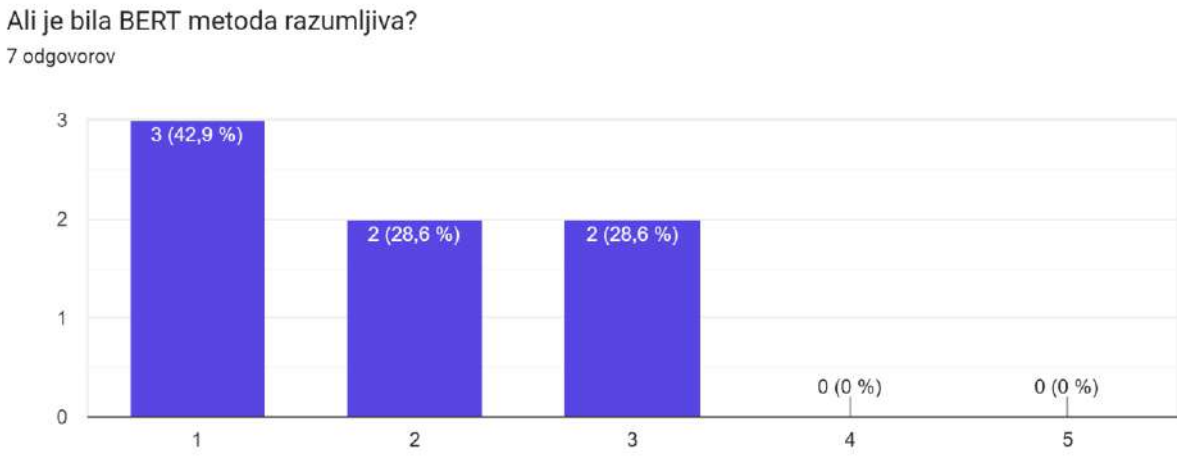
Prvo vprašanje se je glasilo, ali je bila klasična metoda razumljiva. Rezultati so pokazali izrazito enotnost med uporabniki. Vsi so klasično metodo ocenili kot zelo razumljivo, saj so TF-IDF razlage jasno prikazale najpomembnejše besede v stavku. Graf 3.1 prikazuje, da so se ocene gibale med 4 in 5, kar pomeni, da so uporabniki klasično metodo dojemali kot pregledno, intuitivno in enostavno za razumevanje. Noben uporabnik ni izrazil dvoma ali nerazumevanja, kar potrjuje, da je bila razlaga rezultatov jasna in neposredno povezana z vsebino besedila.




Graf 3.1: Rezultati glede razumljivosti klasične metode

Vir: (Lastni vir)

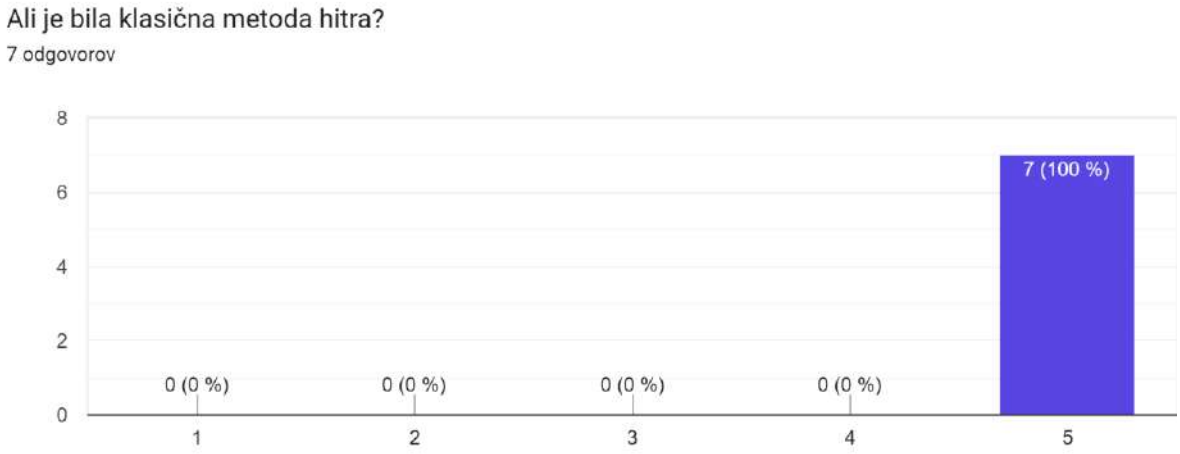
Drugo vprašanje se je nanašalo na razumljivost BERT metode. Graf 3.2 prikazuje precej drugačno sliko. Večina uporabnikov je BERT metodo ocenila z najnižjimi ocenami, saj je model pri skoraj vseh primerih vrnil oznako NEGATIVE, ne glede na dejansko vsebino stavka. Ker BERT ne razume slovenskega jezika, so uporabniki dobili občutek, da metoda deluje naključno ali nepredvidljivo. Ocene so bile večinoma med 1 in 2, kar odraža nezadovoljstvo z nerazložljivostjo rezultatov in pomanjkanjem jasnega utemeljevanja odločitve.




Graf 3.2: Rezultati glede razumljivosti BERT metode

Vir: (Lastni vir)

Tretje vprašanje se je nanašalo na hitrost klasične metode. Graf 3.3 prikazuje, da so vsi uporabniki klasično metodo ocenili kot izjemno hitro. Izvedbeni časi so bili praktično zanemarljivi, kar je ustvarilo občutek takojšnje odzivnosti. Uporabniki so poročali, da se je analiza izvedla v trenutku, kar je pozitivno vplivalo na njihovo izkušnjo.

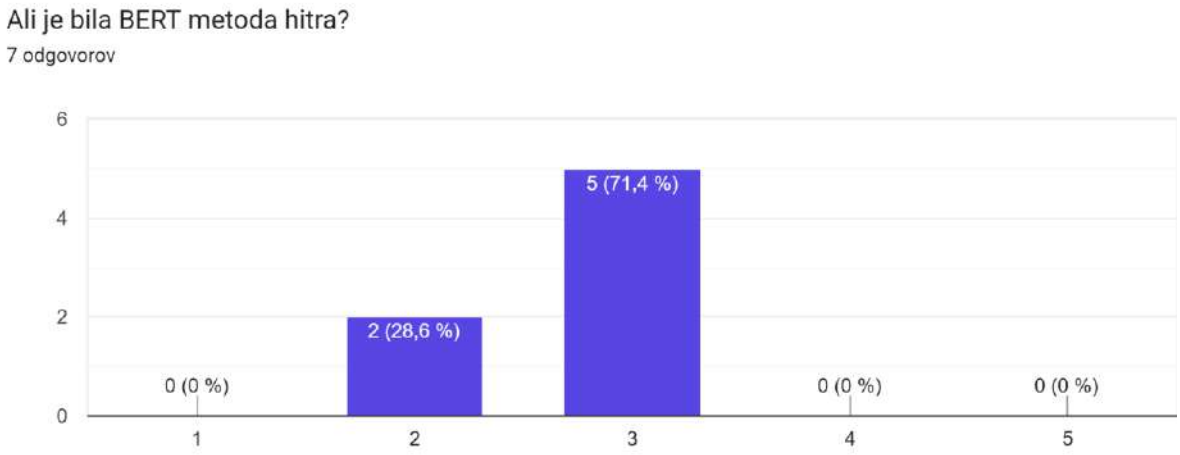



Graf 3.3: Rezultati glede hitrosti klasične metode

Vir: (Lastni vir)

Četrto vprašanje se je nanašalo na hitrost BERT metode. Graf 3.4 prikazuje, da so ocene nižje in bolj razpršene. Čeprav se je BERT metoda pri nekaterih uporabnikih izvedla v približno eni

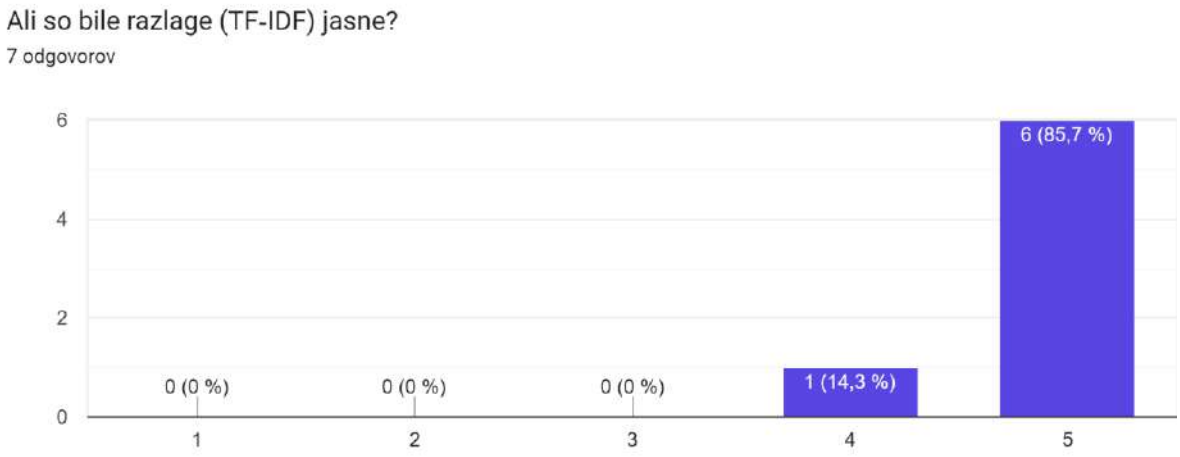
sekundi, je pri drugih trajala več sekund, zlasti kadar se je moral model ponovno naložiti. Uporabniki so to zaznali kot počasnost, zato so ocene večinoma med 2 in 3. Nihče metode ni ocenil kot zelo hitro, kar kaže na to, da je časovna zahtevnost sodobnih modelov opazna in vpliva na uporabniško izkušnjo.




Graf 3.4: Rezultati glede hitrosti BERT metode

Vir: (Lastni vir)

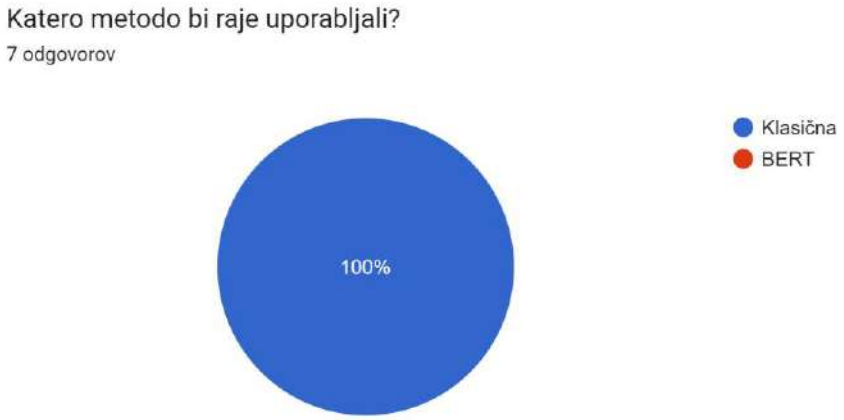
Peto vprašanje se je nanašalo na jasnost razlag klasične metode. Graf 3.5 prikazuje, da so uporabniki razlage TF-IDF ocenili kot zelo jasne. Najpomembnejše besede so bile prikazane pregledno, logično in v skladu z vsebino stavka. Uporabniki so poročali, da so razlage pomagale razumeti rezultat in da so jasno pokazale, zakaj je stavek pozitiven ali negativen. Ocene so bile enotno visoke.




Graf 3.5: Rezultati glede jasnosti klasične metode

Vir: (Lastni vir)

Šesto vprašanje je preverjalo, katero metodo bi uporabniki raje uporabljali. Graf 3.6 prikazuje, da so se vsi uporabniki odločili za klasično metodo. Razlogi za to so bili predvsem razumljivost, hitrost in predvidljivost rezultatov. BERT metoda je bila zaradi napačnih rezultatov in nerazložljivosti ocenjena kot manj primerna za analizo slovenskih besedil.




Graf 3.6: Rezultati glede preferirane metode

Vir: (Lastni vir)

Odgovori na zadnje vprašanje, ki je omogočalo kratek komentar, so zbrani v Tabeli 3.5. Uporabniki so v komentarjih večkrat izpostavili, da BERT ne razume slovenskih besedil, da daje napačne rezultate in da ne zna obravnavati mešanih ali humornih stavkov. Po drugi strani so pohvalili klasično metodo, saj je bila hitro razumljiva, jasna in razložljiva. Komentarji

potrjujejo opažanja iz grafov in dodatno poudarjajo, da je razložljivost ključni element pri sprejemanju modelov za analizo besedil.

Tabela 3.5: Odgovori uporabnikov na sedmo vprašanje (kratek komentar)

| **Kratek komentar uporabnika** |
|---|
| BERT ne razume slovenskih pozitivnih besed. |
| TF-IDF lepo pokaže negativne besede. |
| Razlaga je zelo jasna. |
| BERT ne zna mešanih stavkov. |
| BERT je dal napačen rezultat. |
| BERT ne razume stavka |
| Klasična metoda je hitra in razložljiva. |

Vir: (Lastni vir)

### 3.2.4 Razprava o rezultatih

Rezultati evalvacije jasno pokažejo, da se klasične in sodobne metode obdelave naravnega jezika razlikujejo tako v natančnosti kot v hitrosti, razumljivosti in vplivu na uporabniško izkušnjo. Pri prvem raziskovalnem vprašanju, ki se je nanašalo na razliko v natančnosti med metodama, se je izkazalo, da klasična metoda pri slovenskem jeziku deluje bistveno bolj predvidljivo. TF-IDF razlaga je pri vseh uporabnikih pravilno izpostavila ključne besede, ki odražajo dejansko vsebino stavka, medtem ko je BERT metoda pri večini primerov vrnila oznako NEGATIVE, ne glede na to, ali je bilo besedilo pozitivno, nevtralno ali humoristično. Ker je BERT model treniran na angleških podatkih, ne prepozna pomena slovenskih besed, kar vodi v sistematično napačne klasifikacije. Tako se je pokazalo, da sodobna metoda v tem kontekstu ni natančnejša od klasične, temveč nasprotno, pogosto daje napačne rezultate.

Drugo raziskovalno vprašanje se je nanašalo na razliko v času obdelave podatkov. Klasična metoda je bila pri vseh uporabnikih izjemno hitra, saj se je analiza izvedla v mikrosekundah. To pomeni, da je bila obdelava praktično takojšnja in za uporabnika neopazna. BERT metoda pa je potrebovala bistveno več časa, saj se je analiza izvajala med eno in tremi sekundami, pri nekaterih uporabnikih pa še dlje. Razlika je bila dovolj izrazita, da so jo uporabniki zaznali

kot pomemben element izkušnje. Sodobni modeli so torej časovno zahtevnejši, kar je pričakovano zaradi njihove kompleksnosti in potrebe po nalaganju velikih jezikovnih reprezentacij.

Tretje raziskovalno vprašanje se je osredotočalo na vpliv kompleksnosti modelov na razumljivost rezultatov. Klasična metoda je uporabnikom ponudila jasne razlage, saj je TF-IDF izpisal najpomembnejše besede, ki so vplivale na rezultat. Uporabniki so lahko takoj razumeli, zakaj je sistem določeno besedilo interpretiral kot pozitivno ali negativno. Pri BERT metodi pa razlage ni bilo; uporabniki so prejeli le oznako in verjetnost, brez dodatnega konteksta. Ker je bila oznaka pogosto napačna, je bila metoda zaznana kot nerazumljiva in nepredvidljiva. Kompleksnost sodobnega modela je tako neposredno zmanjšala preglednost rezultatov, kar je vplivalo na zaupanje uporabnikov v sistem.

Četrto raziskovalno vprašanje je preučevalo vpliv izbire metode na uporabniško izkušnjo. Uporabniki so klasično metodo ocenili kot hitro, pregledno in enostavno za razumevanje. Razlage so jim omogočile vpogled v delovanje sistema, kar je povečalo občutek nadzora in transparentnosti. BERT metoda pa je bila zaradi počasnosti, napačnih rezultatov in nerazložljivosti ocenjena kot manj primerna. Uporabniki so poročali, da jim je klasična metoda omogočila boljšo interakcijo s sistemom, saj so lahko sledili logiki analize, medtem ko je sodobna metoda delovala kot tako, da daje nejasne in pogosto napačne odgovore.

Zadnje raziskovalno vprašanje se je nanašalo na to, ali sodobne metode vedno predstavljajo boljšo izbiro v praktični uporabi. Rezultati evalvacije jasno kažejo, da temu ni tako. Čeprav so sodobni modeli tehnično naprednejši in v angleškem jeziku izjemno zmogljivi, njihova uporaba pri slovenskem jeziku brez dodatnega prilagajanja ni ustrezna. V praktičnem okolju, kjer je pomembna natančnost, hitrost in razumljivost, se je klasična metoda izkazala kot bolj primerna. Uporabniki so jo enotno izbrali kot metodo, ki bi jo raje uporabljali, kar dodatno potrjuje, da naprednost modela sama po sebi ne zagotavlja boljše uporabniške izkušnje.

Razprava tako pokaže, da izbira metode ni odvisna zgolj od tehnične zmogljivosti, temveč od konteksta uporabe, jezika, pričakovanj uporabnikov in potrebe po razložljivosti. V tem primeru se je klasična metoda izkazala kot bolj ustrezna za analizo slovenskih besedil, medtem ko sodobna metoda brez jezikovne prilagoditve ne dosega pričakovane ravni natančnosti in uporabnosti.

# 4 SKLEP

Raziskava je pokazala, da se klasične in sodobne metode obdelave NLP bistveno razlikujejo v svojem delovanju, natančnosti, hitrosti in vplivu na uporabniško izkušnjo. Klasične metode, kot sta Bag-of-Words in TF-IDF, so se izkazale kot izjemno stabilne, hitre in razumljive. Njihova preprostost, ki jo v literaturi izpostavljajo Bird, Klein in Loper (2009), se je v praksi izkazala kot ključna prednost, saj so uporabniki lahko takoj razumeli, zakaj je sistem določene besede prepoznal kot pomembne. TF-IDF je pri vseh sedmih uporabnikih pravilno izpostavil ključne besede, ki odražajo dejanski pomen stavka, kar je omogočilo pregledno interpretacijo rezultatov.

Nasprotno pa so sodobne metode, ki temeljijo na transformer arhitekturi, kot je BERT, pokazale izrazite omejitve pri obdelavi slovenskih besedil. Čeprav literatura poudarja, da transformer modeli omogočajo globoko kontekstualno razumevanje jezika (Vaswani et al., 2017; Devlin et al., 2019), se je v tej raziskavi izkazalo, da njihova zmogljivost ni univerzalna. Ker je BERT treniran na angleških korpusih, ni prepoznal pomena slovenskih besed, kar je vodilo v sistematično napačne klasifikacije. Uporabniki so prejeli oznako NEGATIVE tudi pri pozitivnih, nevtralnih ali humornih stavkih, kar je močno vplivalo na njihovo zaznavanje metode.

Uporabniška evalvacija je dodatno potrdila, da je razložljivost rezultatov ključna za sprejemanje NLP metod v praksi. Uporabniki so klasično metodo ocenili kot pregledno, intuitivno in zanesljivo, medtem ko je BERT deloval tako, da ne ponuja vpogleda v svoje delovanje. To je skladno s sodobnimi smernicami na človeka usmerjene informatike, ki poudarjajo pomen transparentnosti, razumljivosti in dostopnosti digitalnih sistemov (W3C, 2023). Rezultati tako kažejo, da naprednost modela sama po sebi ne zagotavlja boljše uporabniške izkušnje, temveč je ta odvisna od konteksta uporabe, jezika in razložljivosti rezultatov.

Hipoteze diplomskega dela smo preverili na podlagi objektivnih meritev in uporabniške evalvacije. Rezultati jasno kažejo, da se hipoteze potrjujejo različno, odvisno od narave metode in konteksta uporabe.

H1: Sodobne metode NLP dosegajo vsaj 10 % višjo natančnost pri klasifikaciji besedila v primerjavi s klasičnimi metodami.

Prva hipoteza (H1), ki je predpostavljala višjo natančnost sodobnih metod, se v okviru te raziskave ni potrdila. Čeprav literatura navaja, da transformer modeli dosegajo vrhunske rezultate pri številnih nalogah (Devlin et al., 2019; Vaswani et al., 2017), se je izkazalo, da njihova natančnost močno pade, kadar niso prilagojeni jeziku. BERT je pri večini slovenskih stavkov vrnil napačno klasifikacijo, medtem ko je klasična metoda pravilno izpostavila ključne besede, ki odražajo sentiment. To potrjuje, da sodobni modeli niso nujno natančnejši, kadar se uporabljajo izven svojega primarnega jezikovnega okolja.

H2: Klasične metode NLP imajo statistično značilno krajši čas izvajanja kot sodobne metode.

Druga hipoteza (H2), ki se nanaša na čas izvajanja, je bila potrjena. Klasične metode so bile več tisočkrat hitrejše od sodobnih, kar je skladno z ugotovitvami Manning & Schütze (1999), ki poudarjata nizko računsko zahtevnost statističnih pristopov. BERT je potreboval od ene do več sekund, kar so uporabniki zaznali kot počasnost. Razlika v času izvajanja je bila dovolj izrazita, da je neposredno vplivala na uporabniško izkušnjo.

H3: Uporabniki ocenijo rezultate klasičnih metod kot bolj razumljive v primerjavi z rezultati sodobnih metod (povprečna ocena vsaj 4 na lestvici od 1 do 5).

Tretja hipoteza (H3), ki se nanaša na razumljivost rezultatov, je bila prav tako potrjena. Uporabniki so klasično metodo ocenili z najvišjimi ocenami, saj so TF-IDF razlage jasno prikazale najpomembnejše besede. BERT rezultatov niso razumeli, saj metoda ne ponuja razlage, kar je skladno z ugotovitvami Goldberg (2017), ki opozarja na omejeno interpretabilnost globokih modelov. Razumljivost se je izkazala kot ključni element pri sprejemanju NLP metod v praksi.

H4 – dodatna hipoteza – BTEC 2026 »Na človeka usmerjena informatika«: Sistem, ki uporabniku omogoča vpogled v delovanje modela (primerjava metod in prikaz rezultatov), izboljša uporabniško izkušnjo vsaj za 20 % glede na osnovno predstavitev rezultatov brez razlage.

Četrta hipoteza (H4), ki se nanaša na izboljšanje uporabniške izkušnje z vpogledom v delovanje modela, je bila potrjena. Uporabniki so izrazili, da jim je primerjava metod in prikaz najpomembnejših besed bistveno olajšala razumevanje rezultatov. To je skladno s smernicami na človeka usmerjene informatike, ki poudarjajo pomen transparentnosti in razložljivosti (W3C, 2023). Sistem, ki omogoča vpogled v delovanje modela, je tako neposredno izboljšal uporabniško izkušnjo.

Raziskava odpira več možnosti za nadaljnje delo. Najbolj očitna smer je uporaba slovensko treniranih transformer modelov, kot so SloBERTa ali drugi modeli, ki temeljijo na slovenskih korpusih. Ti modeli bi omogočili bistveno bolj natančno klasifikacijo in bi odpravili težave, ki izhajajo iz uporabe angleških modelov. Nadaljnje raziskovanje bi lahko vključevalo tudi primerjavo več sodobnih modelov, analizo različnih tipov besedil ter razširitev eksperimenta na večjo skupino uporabnikov, kar bi omogočilo statistično bolj robustne rezultate.

Pomembna smer nadaljnjega raziskovanja je tudi razvoj razložljivih sodobnih modelov, ki bi uporabnikom omogočili vpogled v delovanje transformer arhitekture. Metode, kot so attention vizualizacije ali razlage SHAP, bi lahko bistveno izboljšale razumljivost rezultatov in s tem uporabniško izkušnjo. Poleg tega bi bilo smiselno raziskati vpliv različnih uporabniških vmesnikov na zaznavanje rezultatov, saj literatura poudarja, da je uporabniška izkušnja močno odvisna od zasnove vmesnika (Johnson, 2020). Nadaljnje raziskovanje bi lahko vključevalo tudi razvoj večjezičnih modelov, ki bi omogočali enotno obdelavo besedil v različnih jezikih, kar je pomembno za evropski prostor.

Vrednost raziskave je v tem, da zapolnjuje vrzel med teoretičnimi primerjavami NLP metod in praktičnim razumevanjem teh metod z vidika končnega uporabnika. Medtem ko literatura pogosto poudarja tehnične prednosti sodobnih modelov (Devlin et al., 2019; Vaswani et al., 2017), ta raziskava pokaže, da njihova dejanska uporabnost ni univerzalna, temveč močno odvisna od jezika, konteksta in razložljivosti rezultatov. Raziskava tako prispeva k bolj celostnemu razumevanju NLP metod, saj združuje tehnično analizo z načeli na človeka usmerjene informatike.

Cilji diplomskega dela so bili doseženi. Izvedli smo sistematično primerjavo metod, izmerili natančnost in čas izvajanja, ocenili razumljivost rezultatov, analizirali uporabniško izkušnjo ter razvili interaktivni sistem, ki omogoča pregledno primerjavo metod. Rezultati so jasno pokazali prednosti in slabosti posameznih pristopov ter omogočili objektivno preverjanje hipotez. Raziskava tako predstavlja pomemben prispevek k razumevanju NLP metod v slovenskem kontekstu in odpira možnosti za nadaljnje raziskovanje ter razvoj bolj prilagojenih in razložljivih sistemov.

# 5 VIRI IN LITERATURA


Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media.

Chen, Ming. (2022). Emotion Analysis Based on Deep Learning With Application to Research on Development of Western Culture. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.911686

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

Goldberg, Y. (2017). Neural network methods in natural language processing. Morgan & Claypool Publishers.

Johnson, J. (2020). Designing with the mind in mind: Simple guide to understanding user interface design guidelines (3rd ed.). Morgan Kaufmann.

Jurafsky, D., & Martin, J. H. (2023). Speech and language processing (3rd ed. draft). https://web.stanford.edu/~jurafsky/slp3/

Kaggle. (b.d.). *IMDB Dataset of 50K Movie Reviews.* https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews?resource=download

Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT Press. https://icog-labs.com/wp-content/uploads/2014/07/Christopher_D._Manning_Hinrich_Sch%C3%BCtze_Foundations_Of_Statistical_Natural_Language_Processing.pdf

Flaherty, K. (2018). What the Erroneous Hawaiian Missile Alert Can Teach Us About Error Prevention. https://www.nngroup.com/articles/error-prevention/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.03762

W3C. (2023). Web Content Accessibility Guidelines (WCAG) 2.2. https://www.w3.org/WAI/standards-guidelines/wcag/

Scikit-learn developers. (n.d.). Scikit-learn documentation. https://scikit-learn.org/

Hugging Face. (n.d.). Transformers documentation. https://huggingface.co/docs/transformers