Razvoj sustava za prepoznavanje i identifikaciju sintetički generiranih glasova korištenjem tehnika strojnog učenja

Vučinić, Mirta

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 723

RAZVOJ SUSTAVA ZA PREPOZNAVANJE I IDENTIFIKACIJU SINTETIČKI GENERIRANIH GLASOVA KORIŠTENJEM TEHNIKA STROJNOG UČENJA

Mirta Vučinić

Zagreb, veljača 2025.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 723

RAZVOJ SUSTAVA ZA PREPOZNAVANJE I IDENTIFIKACIJU SINTETIČKI GENERIRANIH GLASOVA KORIŠTENJEM TEHNIKA STROJNOG UČENJA

Mirta Vučinić

Zagreb, veljača 2025.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 30. rujna 2024.

DIPLOMSKI ZADATAK br. 723

Pristupnica:	**Mirta Vučinić (0036524440)**
Studij:	Računarstvo
Profil:	Znanost o podacima
Mentor:	prof. dr. sc. Davor Petrinović

Zadatak: **Razvoj sustava za prepoznavanje i identifikaciju sintetički generiranih glasova korištenjem tehnika strojnog učenja**

Opis zadatka:

U okviru diplomskog rada potrebno je razviti sustav za detekciju deepfake glasova korištenjem skupa podataka DEEP-VOICE s Kagglea. Skup podataka sadrži primjere stvarnih ljudskih govora i njihove deepfake verzije, generirane metodom "Retrieval-based Voice Conversion". Zadatak uključuje ekstrakciju značajki audio signala poput Mel-frekvencijskih cepstralnih koeficijenata (MFCC), te primjenu klasifikacijskih algoritama kao što su SVM, KNN ili neuronskih mreža poput LSTM-a. Potrebno je implementirati, obučiti i evaluirati model koristeći standardne metrike (točnost, preciznost, odziv). Model treba omogućiti identifikaciju manipulacija u govoru, s potencijalnom primjenom u zaštiti od lažnih audio zapisa i zlonamjernih manipulacija.

Rok za predaju rada: 14. veljače 2025.

Zahvala

Zahvaljujem svom mentoru, prof. dr. sc. Davoru Petrinoviću, na izuzetnom vodstvu tijekom dvije godine diplomskog studija, kao i na stručnoj podršci tijekom izrade ovog diplomskog rada.

Sadržaj

Uvod

U današnje doba digitalizacije sintetički generirani glasovi, poznati i kao deepfake audiozapisi, postaju sve veći izazov za sigurnost, autentičnost i povjerenje u komunikaciju. Napretkom tehnologije umjetne inteligencije i strojnog učenja danas je moguće generiranje audiozapisa koji gotovo u potpunosti imitiraju ljudske glasove. To može prouzrokovati razne zlouporabe kao što su političke manipulacije, financijske prijevare, ugrožavanje privatnosti i sigurnosti.

Iz tih razloga, razvoj učinkovitih sustava za prepoznavanje i identifikaciju sintetički generiranih glasova postao je nužan. Ovaj rad predstavlja uporabu naprednih tehnika strojnog učenja za detekciju sintetički generiranih glasova, posebno usmjeren prema stvaranju sustava koji može precizno razlikovati stvarne ljudske glasove od onih stvorenih korištenjem tehnologija za generiranje i manipulaciju glasa.

Ključni faktor u prepoznavanju sintetički generiranih glasova je otkrivanje neprirodnih elemenata, poput robotskih tonova, nepravilnih pauza i neprirodnog ritma govora.

Metodologija ovog istraživanja uključuje klasifikatore kao što su logistička regresija (Logistic Regression), stroj potpornih vektora (Support Vector Machine, SVM), K-najbližih susjeda (K-Nearest Neighbors, KNN), slučajne šume (Random Forest), stablo odluke (Decision Tree), naivni Bayes (Naive Bayes), CatBoost klasifikator, ekstremno povećanje gradijenta (eXtreme Gradient Boosting, XGBoost), ali i duboke neuronske mreže poput dugoročno-kratkoročne memorije (Long Short Term Memory, LSTM) i konvolucijske neuronske mreže (Convolutional Neural Network, CNN). Ovim metodama je postignuta visoka točnost u razdvajanju stvarnih od sintetičkih glasova.

Slična istraživanja

Anagha, Arya, Hari Narayan, Abhishek i Anjali (2023) [5] istražuju detekciju audio s deepfakeova s pomoću konvolucijskih neuronskih mreža (CNN). Modeli su trenirani na ASVspoof 2019 skupu podataka uz primjenu augmentacije i Mel spektrogramske reprezentacije podataka. Koriste Adam optimizator i evaluiraju performanse pomoću F1-mjere, ROC krivulje i prosječne preciznosti. Njihov najbolji model postiže točnost od 85 %, AUC vrijednost od 0.87 i prosječnu preciznost od 0.90, što potvrđuje njegovu visoku učinkovitost u razlikovanju stvarnih i lažnih audiozapisa.

Dua, Meena, Neelam, Amisha i Chakravarty (2023) [4] istražuju detekciju deepfake audiozapisa korištenjem Gammatone kepstralnih koeficijenata (GTCC) i spektrograma koeficijenata grafičkog frekvencijskog kepstra (GFCC), s različitim modelima strojnog učenja i dubokog učenja. Eksperimenti su provedeni na ASVspoof 2019 i ASVspoof 2021 skupovima podataka. Rezultati pokazuju da kombinacija GFCC spektrograma i unaprijed istreniranog ResNet50 modela postiže najnižu stopu jednakih grešaka (EER) od 1,78 % i najmanju funkciju troškova detekcije (t-DCF) od 0,0458, nadmašujući sve druge testirane metode.

Altalahin, AlZu'bi, Alqudah i Mughaid (2023) [2] koriste CNN-LSTM arhitekturu s Mel-frekvencijski kepstralni koeficijentima (MFCC) za poboljšanje detekcije deepfake audiozapisa. Model je treniran na ASVspoof 2019 skupu podataka i postiže 88 % točnosti, nadmašujući tradicionalne metode strojnog učenja. Njihovi rezultati potvrđuju da kombinacija konvolucijskih mreža i rekurentnih mreža poboljšava sposobnost modela u otkrivanju manipuliranog govora.

Geerthik, Senthil, Jayashree i Abinaya (2024) [3] predlažu multimodalni pristup detekciji deepfake sadržaja, kombinirajući analizu zvuka i videa. Koriste MFCC značajke za analizu audiozapisa i Attention-based CNN za vizualnu analizu, trenirajući model na FakeAVCeleb skupu podataka. Njihova metoda postiže 94 % točnosti, čime pokazuju da integracija audiozapisa i vizualnih podataka može poboljšati pouzdanost detekcije deepfake sadržaja.

Mutica, Mihalache i Burileanu (2024) [1] fokusiraju se na detekciju sintetičkog govora s pomoću dubokih neuronskih mreža (DNN). Koriste tri pristupa: višeslojne perceptrone (MLP), CNN te EfficientNetV2 s transfer learningom. Modeli su trenirani na Fake-or-Real skupu podataka, pri čemu CNN pokazuje najbolje generalizacijske sposobnosti na

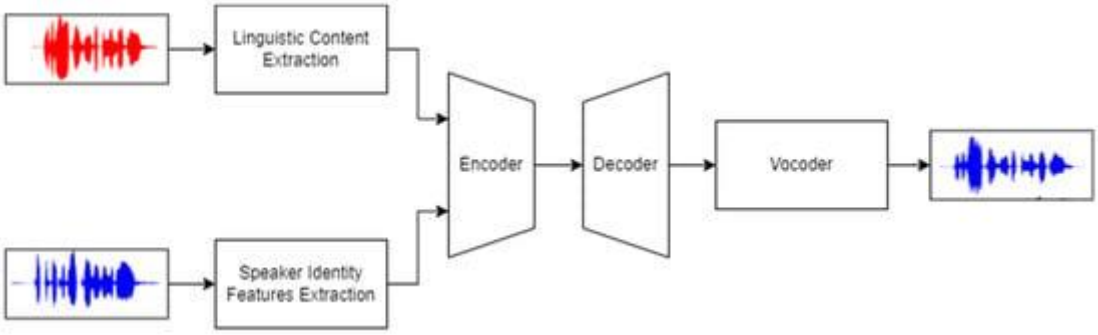
nepoznatim uzorcima postižući točnost na validacijskom skupu od 98.9 % i točnost na test skupu od 83.9 %. Rad naglašava potrebu za razvojem robusnih modela sposobnih za detekciju nepoznatih sintetičkih glasova.

Metode za generiranje sintetičkih glasova

Metode za generiranje sintetičkih glasova koriste napredne tehnologije, poput neuronskih mreža i umjetne inteligencije, kako bi stvorile glasove koji gotovo u potpunosti imitiraju ljudske glasove. Generiranje sintetičkog govora ima dvije glavne vrste generiranja, a to su pretvaranje teksta u govor (Text-to-Speech, TTS) i pretvorba glasa (Voice Conversion, VC).

Pretvorba glasa

Pretvorba glasa (VC) je proces transformacije glasa jednog govornika u glas drugog, bez promjene jezičnog sadržaja. Ova tehnologija manipulira govornim karakteristikama, kao što su identitet glasa, emocije i naglasci. Tipični VC sustav uključuje tri faze: analizu govora, mapiranje i rekonstrukciju. U fazi analize, glas iz izvornog govora se dijeli na značajke koje predstavljaju različite osobitosti govora, poput ritma i intonacije. U fazi mapiranja, izdvojene značajke prilagođavaju se tako da odgovaraju glasovnim karakteristikama ciljnog govornika s pomoću encodera. Na kraju, u fazi rekonstrukcije, decoder/vocoder odgovorni su za obradu podataka dobivenih od encodera kako bi proizveli odgovarajući manipulirani audiozapis (Sl. 0.1).



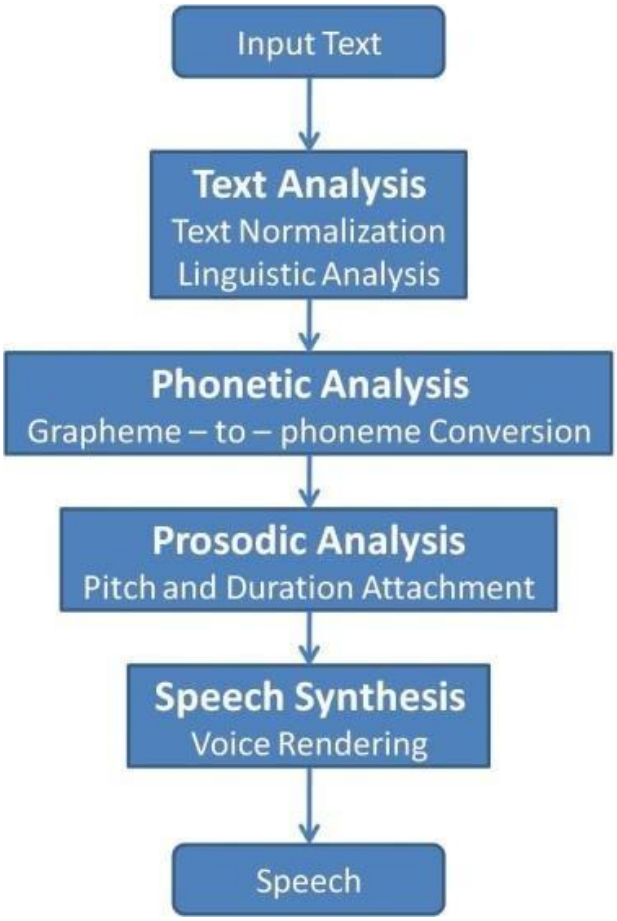
Sl. 0.1 Sustav pretvorbe glasa [1]

Sustavi za pretvaranje teksta u govor

Pretvorba teksta u govor (TTS) odvija se kroz nekoliko faza (Sl. 0.2). Prva faza je analiza teksta, gdje sustav razbija tekst na rečenice i riječi kako bi razumio njihovu strukturu i značenje. Nakon toga, slijedi jezična obrada u kojoj se tekst pretvara u foneme.

Nakon što su fonemi definirani, sustav dodaje odgovarajuće naglaske, ritam i intonaciju kako bi govor zvučao prirodno.

Posljednji korak uključuje sintezu glasa, proizvodeći govor koji blisko oponaša ljudski.



Sl. 0.2 Proces pretvorbe teksta u govor [8]

Potreba za razvojem učinkovitih sustava za detekciju manipuliranih audiozapisa

Razvoj tehnologija za generiranje sintetičkih audiozapisa, donio je brojne korisne primjene poput stvaranja prirodnijih glasovnih asistenata, podršku osobama s govornim oštećenjima, razne primjene za obrazovne, kreativne i mnoge druge svrhe. Međutim, uz ove pozitivne primjene dolazi i niz ozbiljnih prijetnji koje mogu imati značajne posljedice.

Neke od ključnih prijetnji uključuju:

- Širenje dezinformacija
  - Manipulirani audiozapisi omogućuju stvaranje lažnih izjava koje zvuče uvjerljivo iako ih nije izgovorila stvarna osoba i mogu biti iskorišteni za širenje dezinformacija.
- Narušavanje privatnosti i identiteta
  - Tehnologije za generiranje manipuliranih audiozapisa omogućuju imitaciju glasova bez pristanka pojedinaca, čime se narušava njihova privatnost i sigurnost. Takvi audiozapisi mogu biti zloupotrijebljeni u neetičke ili nezakonite svrhe, poput prijevara, iznude ili manipulacija, što ugrožava kako osobu čiji je glas neovlašteno korišten, tako i one koji su cilj takvih radnji.
- Manipulacija dokazima u pravnim postupcima
  - Sintetički generirani audiozapisi mogu biti korišteni za lažno prikazivanje izjava u pravnim slučajevima. Ako se manipulacija dokazima ne prepozna na vrijeme, takvi dokazi mogu dovesti do pogrešnih presuda.

Zbog tih prijetnji, nužno je razviti učinkovite i pouzdane sustave za detekciju manipuliranih audiozapisa koji pravovremeno prepoznaju lažne sadržaje. Njihova glavna svrha je zaštita privatnosti i identiteta pojedinaca. Kako tehnologija za generiranje sintetičkih audiozapisa napreduje, raste i potreba za alatima za njihovo prepoznavanje.

Opis skupa podataka

„DEEP-VOICE“ skup podataka, preuzet s platforme Kaggle, sadrži primjere stvarnog ljudskog govora osam poznatih osoba i sintetički generirane verzije tih govora korištenjem Retrieval-based Voice Conversion (RVC) tehnologije.

RVC je algoritam za konverziju glasa temeljen na umjetnoj inteligenciji koji omogućuje realistične transformacije govora, pri čemu precizno očuva intonaciju i audio karakteristike izvornog govornika. Za razliku od sustava za pretvorbu teksta u govor (Text-to-Speech), RVC pruža pretvorbu govora u govor (Speech-to-Speech), zadržavajući boju glasa i emocionalni ton izvornog govora.

Uz odgovarajuće resurse, poput snažne GPU jedinice i kvalitetnog modela glasa, RVC generira glasove koji su gotovo neprimjetno različiti od stvarnih.

Skup podataka je dostupan u dva formata:

- Audiozapisi: Smješteni u direktorij "AUDIO" s poddirektorijima "REAL" i "FAKE". Nazivi datoteka označavaju koji govornici su pružili stvarni govor i u čije glasove su konvertirani. Prije konverzije, pozadinska buka je uklonjena, a zatim ponovno dodana nakon konverzije.
- Ekstrahirane značajke: Nalaze se u datoteci "DATASET-balanced.csv". Značajke su ekstrahirane iz audiozapisa pomoću prozora duljine jedne sekunde i uravnotežene slučajnim uzorkovanjem.

Fokus ovog istraživanja bio je na analizi i detekciji sintetički generiranog govora putem audiozapisa, pri čemu su iz tih zapisa ekstrahirani Mel-frekvencijski kepstralni koeficijenti (Mel-Frequency Cepstral Coefficients, MFCC) i spektralni prikazi.

MFCC značajke su jedna od najkorištenijih metoda za predstavljanje zvučnih signala u domeni obrade govora. Ove značajke temelje se na ljudskoj percepciji zvuka i koriste Mel ljestvicu kako bi obradile frekvencije na način koji je bliži načinu na koji ih uho interpretira. Proces uključuje segmentaciju zvučnog signala u vremenske prozore, analizu frekvencijskog spektra putem Fourierove transformacije i primjenu Mel filtara koji naglašavaju relevantne frekvencije. Dobiveni koeficijenti sažimaju ključne karakteristike zvuka, omogućujući učinkovitu analizu govora i detekciju razlika između stvarnih i sintetičkih zapisa. Korištenjem tih ekstrahiranih značajki, istraživanje se usmjerilo na izazove real-time

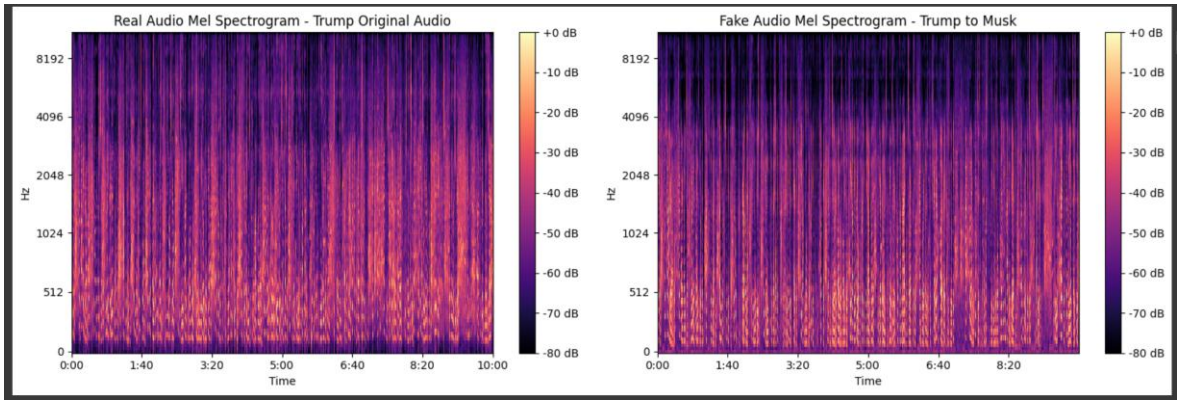
detekcije sintetički generiranih audio zapisa, omogućujući brzu i preciznu identifikaciju umjetno stvorenog govora.

Metodologija

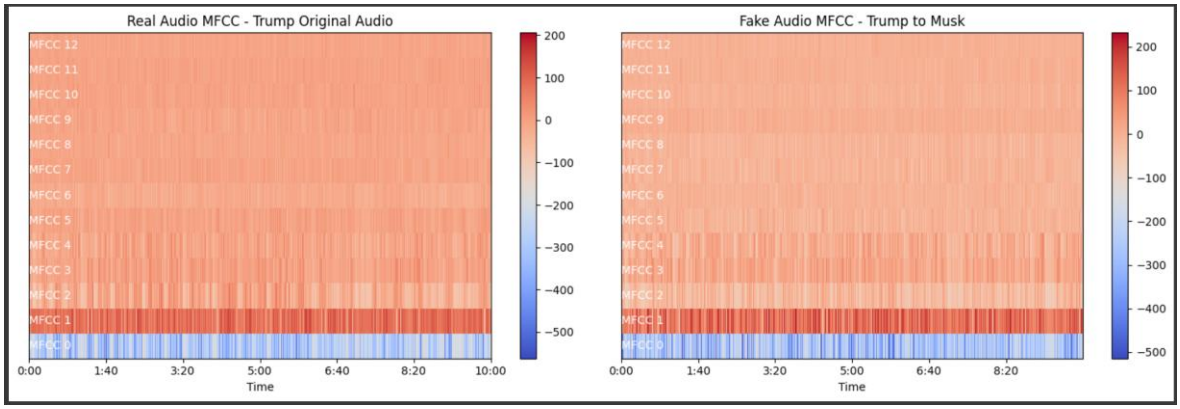
Obrada podataka

Iako nisu dio glavne analize, Mel spektrogrami (Sl. 0.1) i MFCC značajke (Sl. 0.2) prikazani su radi vizualizacije razlika između stvarnog i konvertiranog govora. Mel spektrogrami predstavljaju vizualni prikaz snage signala u frekvencijskom području kroz vrijeme. Ovi prikazi omogućuju usporedbu frekvencijskog sadržaja između originalnog i sintetičkog govora te mogu ukazati na potencijalne promjene u spektralnim karakteristikama prilikom konverzije.

MFCC značajke modeliraju način na koji ljudsko uho percipira zvuk, sažimajući relevantne informacije o govoru. Ove značajke često se koriste u zadacima prepoznavanja govora jer naglašavaju fonetske karakteristike, a prikaz omogućuje usporedbu razlika između prirodnog i sintetičkog govora na temelju njihovih spektralnih svojstava.



Sl. 0.1 Mel spektrogrami stvarnog (lijevo) i sintetičkog (desno) govora

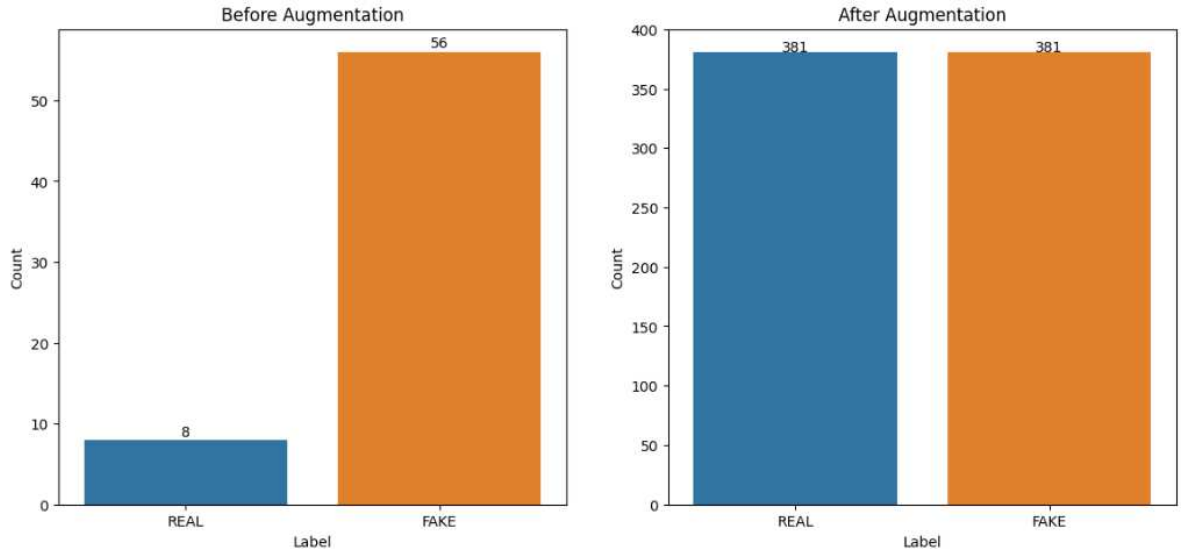


Sl. 0.2 Prikaz MFCC značajki stvarnog (lijevo) i sintetičkog (desno) govora

Kako bi se audiozapisi pripremili za različite tipove modela strojnog učenja, provedeno je nekoliko koraka obrade podataka, uključujući segmentaciju, augmentaciju, ekstrakciju značajki i organizaciju podataka za treniranje.

Prvi korak u obradi podataka bio je razdvajanje audiozapisa u manje segmente. Originalni skup podataka sadržavao je 8 stvarnih (REAL) i 56 generiranih (FAKE) audiozapisa. Kako bi se povećao broj uzoraka i omogućilo bolje učenje modela, svaki audiozapis segmentiran je u kraće dijelove od 10 sekundi s pomoću funkcije split_audio_file. Dobiveni segmenti spremljeni su u odgovarajuće direktorije prema pripadnosti klasi (REAL ili FAKE).

Nakon segmentacije, broj segmenata u klasi FAKE bio je značajno veći nego u klasi REAL, što je stvorilo neravnotežu u podacima. Kako bi se riješio ovaj problem, iz skupa FAKE nasumično je odabrana podskupina segmenata koja je odgovarala broju segmenata iz klase REAL. Odabrani segmenti premješteni su u poseban direktorij, čime je broj segmenata u obje klase izjednačen na 381 (Sl. 0.3). Ovaj postupak osigurao je ravnotežu između klasa, što je ključno za kvalitetnu obuku modela.

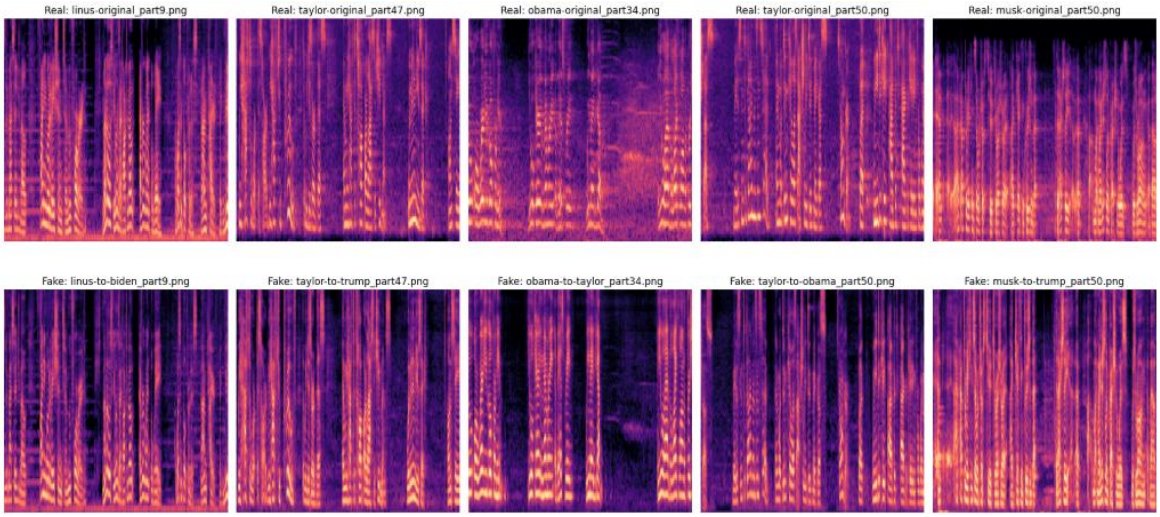


Sl. 0.3 Količina podataka prije i poslije augmentacije

Za različite vrste modela primijenjeni su prilagođeni postupci ekstrakcije značajki. Klasični klasifikatori, kao što su SVM, logistička regresija, Random Forest, naivni Bayes i drugi, koristili su MFCC značajke, koje su ekstrahirane iz audiozapisa koristeći funkciju extract_features. Ove značajke predstavljale su prosječne vrijednosti svakog koeficijenta, što ih čini pogodnima za takve modele. Za LSTM modele, MFCC značajke pripremljene su kao vremenske sekvence s pomoću funkcije extract_mfcc_sequence, čime je očuvan vremenski

kontekst podataka. Sekvence su dodatno normalizirane radi konzistentnosti i učinkovitosti tijekom treniranja.

CNN modeli koristili su spektrograme kao ulazne podatke. Audiozapisi su pretvoreni u spektrograme (Sl. 0.4) s pomoću funkcije save_spectrogram, a rezultirajuće slike pohranjene su u PNG formatu.



Sl. 0.4 Prikaz spektrograma stvarnih i manipuliranih audiozapisa

Podaci su zatim pripremljeni za treniranje različitih modela. Za klasične klasifikatore i LSTM modele, značajke su podijeljene na trening i testni skup pomoću funkcije train_test_split. Pri podjeli podataka korišteno je stratificirano uzorkovanje, čime je osigurano da omjer klasa ostane isti u trening i testnom skupu. Podaci su podijeljeni u omjeru 80:20, pri čemu je 80 % uzoraka korišteno za treniranje modela, dok je preostalih 20 % korišteno za testiranje. Za LSTM modele, uz stratificiranu podjelu, dodatno je provedeno one-hot enkodiranje klasa, čime su oznake (0 za Fake i 1 za Real) pretvorene u binarne vektore kompatibilne s neuronskim mrežama.

Za CNN modele, spektrogrami su organizirani u direktorije prema klasama kako bi se omogućilo učenje s pomoću ImageDataGenerator funkcije. Umjesto ručne podjele podataka, korišten je ugrađeni parametar validation_split = 0.2, koji odvaja 20 % podataka za validaciju, dok preostalih 80 % ostaje za treniranje. Osim podjele, ImageDataGenerator je izvršio i normalizaciju piksela skaliranjem vrijednosti piksela na raspon [0,1], čime se poboljšava stabilnost treniranja i smanjuje osjetljivost modela na intenzitete piksela.

Ovakva priprema podataka omogućila je optimalno treniranje i evaluaciju modela, osiguravajući da su podaci ravnomjerno raspoređeni u svim skupovima, čime se sprječava potencijalna pristranost modela prema određenoj klasi.

Opis modela

Naivni Bayesov klasifikator

Naivni Bayes je klasifikacijski model koji se temelji na Bayesovom teoremu (1) i pretpostavlja uvjetnu nezavisnost značajki unutar svake klase.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Gdje:

- P(A|B) predstavlja uvjetnu vjerojatnost događaja A, s obzirom na događaj B.
- P(B|A) predstavlja vjerojatnost događaja B, s obzirom na događaj A.
- P(A) predstavlja vjerojatnost događaja A.
- P(B) predstavlja vjerojatnost događaja B.

Gaussov Bayesov klasifikator (GNB), korišten u ovom istraživanju, prilagođen je za kontinuirane podatke s pretpostavkom da svaka značajka slijedi Gaussovu (normalnu) raspodjelu. Ova pretpostavka omogućuje pojednostavljenje izračuna vjerojatnosti, jer je dovoljno procijeniti samo parametre srednje vrijednosti i standardne devijacije za svaku značajku unutar klase, umjesto procjene pune distribucije. GNB se pokazuje vrlo učinkovitim kada podaci zaista prate Gaussovu distribuciju, a njegova jednostavnost osigurava brzo treniranje i precizne predikcije.

K-najbližih susjeda

K-najbližih susjeda (KNN) je algoritam za klasifikaciju koji se temelji na pretpostavci da slični podaci u prostoru značajki pripadaju istoj klasi. Kod klasifikacije novog uzorka, KNN identificira K najbližih uzoraka iz trening skupa koristeći odabranu mjeru udaljenosti, poput Euklidske, Manhattan ili Minkowski udaljenosti. Klasa novog uzorka određuje se na temelju klase kojoj pripada većina njegovih najbližih susjeda.

Broj susjeda, K, značajno utječe na točnost KNN algoritma. Mali broj K može dovesti do odziva na šum i odstupajućih vrijednosti, dok veliki K može zamagliti razliku među klasama.

Glavni nedostatak KNN-a je njegova računalna zahtjevnost kod velikih skupova podataka, jer za svaku klasifikaciju mora izračunati udaljenosti između novog podatka i svih točaka u trening skupu. Također, KNN je osjetljiv na dimenzionalnost podataka, što znači da s porastom broja značajki učinkovitost algoritma može opadati.

U ovom istraživanju, kako bi se poboljšale performanse KNN modela, priprema podataka provedena je kroz nekoliko koraka. Prvo je primijenjena metoda RobustScaler za skaliranje značajki. Ova metoda smanjuje utjecaj odstupajućih vrijednosti, čineći model otpornijim na anomalije u podacima. Nakon skaliranja, korištena je SelectKBest metoda za odabir značajki na temelju njihove važnosti. Ovaj postupak koristi ANOVA F-statistiku kako bi procijenio doprinos svake značajke razlici među klasama, osiguravajući da model koristi samo najrelevantnije informacije.

F-statistika ocjenjuje omjer između varijance između različitih klasa i varijance unutar pojedinačnih klasa. Visoke vrijednosti F-statistike ukazuju na značajke koje bolje razdvajaju klase, čime se omogućava smanjenje dimenzionalnosti podataka zadržavajući samo najinformativnije značajke.

Za optimizaciju KNN modela korišten je GridSearchCV, koji omogućava sustavno pretraživanje najboljih kombinacija hiperparametara uz unakrsnu validaciju. Ovom metodom ispitani su različiti brojevi susjeda, metode prilagodbe težina i metričke funkcije udaljenosti, čime se osigurava odabir parametara koji daju najbolje performanse na validacijskom skupu podataka.

Stroj potpornih vektora

Stroj potpornih vektora (SVM) vrsta je algoritma nadziranog učenja koji se koristi u strojnom učenju za rješavanje zadataka klasifikacije i regresije. Poznat je po svojoj sposobnosti učinkovitog rješavanja linearnih i nelinearnih problema. Osnovna ideja SVM-a temelji se na pronalaženju optimalne granice odluke, odnosno hiperravnine, koja razdvaja dvije klase u prostoru značajki s maksimalnim razmakom između najbližih uzoraka iz svake klase. Najbliži uzorci, koji leže na margini hiperravnine, nazivaju se potporni vektori i ključni su za određivanje granice odluke.

Ključna prednost SVM-a je njegova sposobnost da maksimizira razdvajanje među klasama, čime se smanjuje rizik od pogrešne klasifikacije. Ovaj princip maksimizacije razdvajanja čini SVM vrlo učinkovitom metodom, posebno u situacijama kada podaci sadrže šum ili odstupajuće vrijednosti. Dodatno, ako podaci nisu linearno odvojivi u svojoj izvornoj dimenzionalnosti, SVM omogućuje mapiranje podataka u viši dimenzionalni prostor. U tom prostoru podaci postaju linearno odvojivi, čime SVM može pronaći optimalnu granicu. Najčešće korištene jezgrene funkcije uključuju linearnu jezgrenu funkciju, koja je prikladna za podatke koji su prirodno linearno odvojivi, te Radial Basis Function (RBF), koji omogućuje rješavanje složenih, nelinearnih problema klasifikacije.

U ovom istraživanju, SVM model je integriran u tok obrade podataka, odnosno pipeline, kako bi se osigurala njegova optimalna primjena na podacima. Kao prvi korak, korišten je RobustScaler za skaliranje značajki, čime se smanjuje utjecaj odstupajućih vrijednosti.

Nakon skaliranja, primijenjena je metoda SelectKBest za odabir najvažnijih značajki.

I na kraju, korišten je SVC (Support Vector Classifier), implementacija SVM algoritma za klasifikaciju, kako bi se izgradio prediktivni model.

Kako bi se postigla optimalna konfiguracija modela, korišten je GridSearchCV. GridSearchCV je ispitivao različite kombinacije hiperparametara, uključujući vrijednosti regularizacijskog parametra C, različite vrijednosti gama parametra i vrste jezgrenih funkcija (linearna ili RBF).

Logistička regresija

Logistička regresija jedan je od najpoznatijih i najjednostavnijih statističkih modela koji se koristi za binarne klasifikacijske zadatke. Osnovni princip ovog modela je predviđanje vjerojatnosti pripadnosti uzorka jednoj od dviju klasa na temelju ulaznih značajki. Model koristi sigmoidnu funkciju, koja transformira linearnu kombinaciju ulaznih značajki u vrijednost u rasponu između 0 i 1. Ove vrijednosti se interpretiraju kao vjerojatnosti, a konačna klasifikacija se temelji na graničnoj vrijednosti (najčešće 0.5).

Jedna od ključnih pretpostavki logističke regresije je linearna razdvojivost podataka u prostoru značajki. To znači da se podaci mogu odvojiti s pomoću ravnine, tj. linearne granice odluke. Ako podaci nisu linearno odvojivi, performanse modela mogu biti ograničene, no u takvim slučajevima mogu se koristiti metode za proširenje značajki, poput dodavanja nelinearnih kombinacija značajki ili korištenja regularizacije kako bi se smanjio problem pretreniranja.

Prednost logističke regresije je njena sposobnost da se koristi s različitim regularizacijskim tehnikama, poput L1 (Lasso) i L2 (Ridge) regularizacije. Ove tehnike smanjuju složenost modela i sprječavaju pretreniranje, što je posebno korisno kod visoke dimenzionalnosti podataka ili kada podaci sadrže mnogo irelevantnih značajki. Regularizacija uvodi kaznu na veličinu koeficijenata modela, čime se naglašavaju najvažnije značajke.

Za optimizaciju i unapređenje performansi modela, korišten je tok obrade podataka (pipeline), koji je omogućio automatsko primjenjivanje različitih transformacija na podatke i integraciju koraka pretprocesiranja s treniranjem modela. Pipeline se sastoji od tri glavna koraka. Prvi korak uključuje skaliranje značajki korištenjem RobustScaler metode. Drugi korak uključuje SelectKBest metodu i posljednji korak u pipelineu uključuje LogisticRegression model za klasifikaciju.

Za optimalnu konfiguraciju modela, korišten je GridSearchCV koji je ispitivao različite kombinacije hiperparametara kao što su regularizacijski parametar C, vrste regularizacije (L1, L2) i odabir solvera (liblinear, saga).

Slučajne šume

Random Forest je metoda koja koristi skup stabala odluke za rješavanje klasifikacijskih zadataka. Osnovna ideja ove metode je da se kroz pridruživanje predikcija više stabala odluke dobije stabilniji i robusniji rezultat nego što bi to bio slučaj sa samo jednim stablom. Svako stablo u šumi donosi svoju predikciju na temelju vlastitih odluka, a konačna klasifikacija temelji se na većinskom glasanju svih stabala. Ovaj pristup značajno smanjuje rizik od pretreniranosti (overfittinga), koji je često prisutan kod pojedinačnih stabala odluke, jer Random Forest koristi nasumično odabrane podskupove značajki i uzoraka za treniranje svakog stabla, čime se povećava generalizacija modela.

U ovom istraživanju, Random Forest model integriran je u pipeline, koji je omogućio integraciju koraka pretprocesiranja podataka s treniranjem modela. Pipeline započinje s RobustScaler metodom, zatim sa SelectKBest metodom, te posljednji korak koristi RandomForestClassifier za klasifikaciju.

Za optimizaciju hiperparametara korišten je GridSearchCV gdje su testirani parametri poput broja stabala, maksimalne dubine stabala, te parametra koji kontrolira uporabu bootstrappinga.

Stablo odluke

Stablo odlučivanja je nadzirani algoritam učenja koji se koristi za zadatke klasifikacije i regresije. Osnovna ideja stabla odluke je podijeliti podatke u hijerarhijsku strukturu, gdje svaki čvor predstavlja uvjet koji omogućava razdvajanje podataka na temelju određenih kriterija. Stablo se gradi tako da se na svakom čvoru donosi odluka o tome kako podijeliti podatke, koristeći kriterij razdvajanja kao što su gini ili entropija.

Gini mjeri vjerojatnost da će dva nasumično odabrana elementa iz skupa pripadati različitim klasama, pri čemu niži gini označava veći nivo homogenosti podataka. S druge strane, entropija mjeri nesigurnost ili neuređenost skupa podataka, gdje veća entropija ukazuje na veći stupanj nesigurnosti, odnosno veću pomiješanost klasa.

Svaka podjela podataka u čvorovima nastoji optimizirati kriterij razdvajanja, odabirući vrijednost koja najbolje razdvaja podatke u cilju postizanja homogenih podskupina. Taj proces se ponavlja rekurzivno dok se ne zadovolje određeni uvjeti, kao što su minimalni broj uzoraka u čvorovima ili minimalna dubina stabla.

Iako je model stabla odluke jednostavan za implementaciju i vrlo intuitivan, može biti sklon overfittingu, osobito kada su podaci vrlo kompleksni ili imaju mnogo značajki.

U ovom istraživanju, stablo odluke integrirano je u pipeline koji obuhvaća RobustScaler metode za skaliranje značajki, SelectKBest tehniku koja omogućava odabir najvažnijih značajki na temelju statističkog testa i DecisionTreeClassifier za treniranje modela.

Za optimizaciju hiperparametara korišten je GridSearchCV gdje su testirane kombinacije parametara poput maksimalne dubina stabla, broja minimalnih uzoraka potrebnih za podjelu čvora, broja minimalnih uzoraka potrebnih za listove, te kriterija razdvajanja (gini ili entropy).

CatBoost klasifikator

CatBoost klasifikator je napredni model temeljen na algoritmu povećanja gradijenta (gradient boosting) koji je posebno optimiziran za rad s kategorijskim podacima.

Povećanje gradijenta je tehnika u kojoj se slabiji modeli, obično stabla odluka, treniraju u sekvencama, pri čemu svaki sljedeći model ispravlja pogreške prethodnog. Ovaj pristup optimizira više slabih modela u jedan snažan model koji pokazuje bolje performanse u predviđanju i klasifikaciji.

Jedna od značajki koja čini CatBoost posebnim je način na koji obrađuje kategorijske značajke. Umjesto da zahtijeva ručno kodiranje kategorijskih varijabli, kao što je česta praksa u mnogim modelima, CatBoost automatski prepoznaje ove varijable i primjenjuje specifične tehnike kodiranja koje minimiziraju rizik od overfittinga i smanjuju potrebu za dodatnim pretprocesiranjem podataka.

Iako je CatBoost poznat po svojoj sposobnosti obrade kategorijskih podataka, u ovom istraživanju je korišten za rad s numeričkim značajkama, specifično MFCC značajkama. Ovaj pristup omogućava CatBoostu da efikasno procesuira numeričke podatke bez potrebe za dodatnim kodiranjem kategorijskih značajki, a model može koristiti svoje prednosti i za ove vrste podataka.

U ovom istraživanju, CatBoost model je inicijaliziran s parametrima verbose = 0, koji onemogućuje ispis informacija o tijeku učenja i time smanjuje količinu izlaznih podataka te ubrzava treniranje, i random_state = 42, koji osigurava reproducibilnost rezultata. Time se pri svakom pokretanju koda s istim podacima i parametrima postižu identični ishodi.

Ekstremno povećanje gradijenta

Ekstremno povećanje gradijenta (XGBoost) je jedan od najpoznatijih i najmoćnijih modela temeljenih na algoritmu povećanja gradijenta. Ovaj model optimizira preciznost i brzinu, istovremeno primjenjujući napredne tehnike regularizacije za smanjenje problema prenaučenosti. XGBoost je izuzetno učinkovit u radu s velikim skupovima podataka, jer koristi različite tehnike za ubrzanje procesa treniranja, poput paralelizacije i efikasnog rukovanja s nedostajućim podacima.

U ovom istraživanju, XGBoost model je korišten za klasifikaciju, s postavljenim parametrima use_label_encoder = False i eval_metric = 'logloss'. Parametar use_label_encoder = False isključuje automatsko kodiranje etiketa koje bi inače bilo primijenjeno na kategorijske varijable, što može biti korisno u situacijama kada je kodiranje već obavljeno prije obuke. Parametar eval_metric = 'logloss' korišten je za procjenu učinkovitosti modela, pri čemu se mjeri kvalitetu predikcija modela u odnosu na stvarne klase. Ova metrika omogućuje modelu da se fokusira na minimiziranje nesigurnosti u predikcijama, što rezultira većom točnošću, osobito u zadacima s neuravnoteženim klasama ili gdje je preciznost predikcija ključna.

Model dugoročno – kratkoročne memorije

Model dugoročno – kratkoročne memorije (LSTM) je specifičan tip povratne neuronske mreže (Recurrent Neural Network, RNN) koja je dizajnirana za rad s vremenskim serijama i sekvencijalnim podacima. Ključna prednost LSTM-a u odnosu na klasične RNN-ove je njegova sposobnost pamćenja dugoročnih ovisnosti unutar podataka, što omogućava modelu da učinkovito analizira i prepoznaje obrasce u podacima koji ovise o vremenskim sekvencama. Ova sposobnost pamćenja dugoročnih veza čini LSTM izuzetno prikladnim za analize podataka poput audiozapisa, tekstualnih podataka, govora, glazbe i drugih vrsta vremenski ovisnih informacija.

LSTM model koristi specifičnu arhitekturu koja uključuje LSTM jedinice, dizajnirane za učenje dugoročnih veza i procesiranje sekvencijalnih informacija. Ove jedinice sadrže unutarnje komponente: vrata za zaborav, vrata za ulaz i vrata za izlaz, koje omogućuju modelu da selektivno pamti ili zaboravi informacije iz prethodnih vremenskih koraka. Vrata za zaborav određuju koje informacije iz prethodnog vremenskog koraka trebaju biti zaboravljene. Vrata za ulaz omogućuju modelu da odluči koje nove informacije treba pohraniti u trenutačnom vremenskom koraku, a vrata za izlaz odlučuju koje informacije iz trenutačnog vremenskog koraka će biti proslijeđene prema sljedećem sloju mreže.

U ovom istraživanju, LSTM model je konstruiran s tri LSTM sloja, pri čemu su između tih slojeva uključeni Dropout slojevi. Dropout je tehnika regularizacije koja nasumično isključuje, odnosno postavlja na nulu, određene veze između neurona tijekom treniranja, čime se smanjuje rizik od overfittinga. Korištenje Dropouta pomaže modelu da bolje generalizira na neviđenim podacima, odnosno doprinosi jačanju stabilnosti i preciznosti modela, osobito kada se radi s velikim i kompleksnim skupovima podataka.

Za klasifikaciju, na kraju modela koristi se softmax aktivacijska funkcija, koja omogućava modelu izračunavanje vjerojatnosti pripadnosti podataka jednoj od mogućih klasa. U kontekstu ovog istraživanja, softmax se koristi za klasifikaciju u dvije kategorije: "Fake" i "Real".

U modelu je primijenjena L2 regularizacija, koja dodatno pomaže u smanjenju prekomjernog učenja, jer dodaje penalizaciju na velike težine u mreži. Također, korišten je Adam optimizator, koji omogućuje učinkovitu minimizaciju funkcije gubitka i brzo učenje modela, a gubitak se računa s pomoću unakrsne entropije.

Konvolucijska neuronska mreža

Konvolucijska neuronska mreža (CNN) je specifičan tip neuronske mreže koji je tradicionalno razvijen za analizu i obradu slika, ali u ovom istraživanju koristi se za analizu spektrograma. Spektrogrami su vizualni prikazi audiozapisa, koji prikazuju raspodjelu frekvencija u vremenskom intervalu. Korištenje CNN-a za analizu spektrograma omogućuje modelu da prepozna ključne vremenske i frekvencijske obrasce u audiozapisima.

Arhitektura modela započinje s konvolucijskim slojem, koji koristi 32 filtra veličine (3, 3) i ReLU aktivaciju. Ovaj sloj izvodi konvolucijske operacije na ulaznom spektrogramu, automatski učeći osnovne značajke poput rubova i tekstura. Sloj je postavljen na input_shape = (img_height, img_width, 3), što znači da ulazni podaci predstavljaju slike s 3 kanala, odnosno svaki piksel slike ima tri vrijednosti koje odgovaraju intenzitetima crvene, zelene i plave boje.

Nakon prvog konvolucijskog sloja, slijedi MaxPooling2D sloj, koji smanjuje prostornu dimenziju slike za pola, čime model postaje učinkovitiji u prepoznavanju većih obrazaca. Dropout sloj također je dodan kako bi se spriječilo prenaučenost modela, slučajnim isključivanjem 20 % veza između neurona tijekom treniranja.

Zatim slijedi još jedan konvolucijski sloj sa 64 filtra i veličinom jezgre (3, 3), koji dodatno uči složenije značajke u spektrogramu. Ponovno se koristi MaxPooling2D za smanjenje dimenzionalnosti, uz Dropout sloj koji isključuje 25 % veza. Sljedeći konvolucijski sloj ima 128 filtara, a zatim slijedi još jedan MaxPooling2D i Dropout sloj koji također isključuje 25 % veza.

Nakon što konvolucijski slojevi obrade podatke, izlaz se prenosi u Flatten sloj, koji izravnava višedimenzionalne podatke u jednodimenzionalni vektor, što omogućuje prijenos podataka u Dense slojeve. Prvi Dense sloj sadrži 128 neurona i koristi ReLU aktivaciju za dodavanje nelinearnosti. Dropout sloj koristi se ponovno i isključuje 30 % veza.

Na kraju, model ima output sloj s 2 neurona, koji koristi softmax aktivaciju kojom se izračunava vjerojatnost pripadnosti svakoj od dviju klasa, "Fake" i "Real".

Za optimizaciju modela koristi se Adam optimizator, koji je popularan zbog svoje efikasnosti u treniranju dubokih mreža, te gubitak se računa s pomoću unakrsne entropije.

Trening i evaluacija modela

Naivni Bayesov klasifikator

Za bolje razumijevanje rezultata, sljedeće metrike su korištene za ocjenu performansi modela (Sl. 0.5).:

- Točnost (Accuracy): Metrika koja mjeri udio ispravno klasificiranih uzoraka u odnosu na ukupan broj uzoraka. Točnost naivnog Bayesa bila je 79.74 %, što ukazuje na opću sposobnost modela da pravilno klasificira uzorke.
- Preciznost (Precision): Metrika koja mjeri udio točno klasificiranih pozitivnih uzoraka u odnosu na sve uzorke koje je model klasificirao kao pozitivne. Za Naivni Bayesov model, preciznost je iznosila 79.22 %, što znači da je od svih uzoraka koje je model označio kao pozitivne, većina bila točno klasificirana.
- Odziv (Recall): Metrika koja mjeri udio točno klasificiranih pozitivnih uzoraka u odnosu na sve stvarne pozitivne uzorke. Model je postigao odziv od 80.26 %, što pokazuje njegovu sposobnost da identificira većinu stvarnih pozitivnih uzoraka.
- F1-mjera (F1 Score): Harmonijska sredina preciznosti i odziva, koja daje uravnotežen rezultat u slučajevima kada su ove dvije metrike u neskladu. F1-mjera je bila 79.74 %, što ukazuje na uravnoteženost između preciznosti i odziva modela.

```
Accuracy: 0.7973856209150327
Precision: 0.7922077922077922
Recall: 0.8026315789473685
F1-score: 0.7973856209150327
```

Sl. 0.5 Evaluacijske mjere za naivni Bayesov klasifikator

Analiza performansi po klasama (Sl. 0.6) pokazuje da je za klasu "Fake" postignuta preciznost od 80 %, odziv od 79 % i F1-mjera od 80 %, dok su za klasu "Real" rezultati preciznost od 79 %, odziv od 80 % i F1-mjera od 80 %.

	precision	recall	f1-score	support
0	0.80	0.79	0.80	77
1	0.79	0.80	0.80	76
accuracy			0.80	153
macro avg	0.80	0.80	0.80	153
weighted avg	0.80	0.80	0.80	153

Sl. 0.6 Klasifikacijski izvještaj za naivni Bayesov klasifikator

Matrica zabune (Sl. 0.7) pokazuje 61 pravilno klasificiranih "Fake" (True Negatives) i 61 pravilno klasificiranih "Real" uzoraka (True Positives), uz 16 pogrešno klasificiranih "Real" (False Positives) i 15 pogrešno klasificiranih "Fake" uzoraka (False Negatives).

Sl. 0.7 Matrica zabune za naivni Bayesov klasifikator

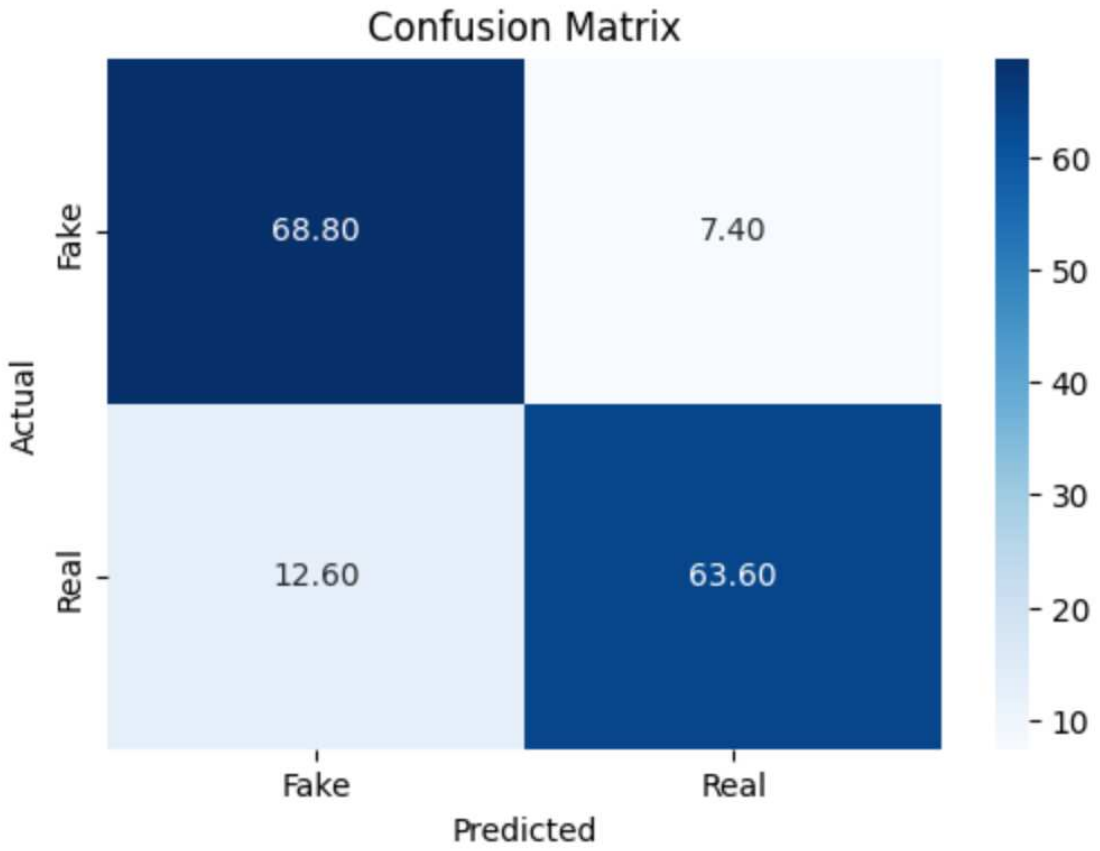
Važno je uzeti u obzir ne samo konačne rezultate, već i prosječne metrike dobivene korištenjem kros-validacije (cv = 5). Ove prosječne metrike omogućuju bolje razumijevanje koliko odstupa trenutačna izvedba modela od njegovih uobičajenih vrijednosti na temelju prethodnih evaluacija. Prosječne metrike iz kros-validacije uključuju prosječnu točnost od

86.87 %, prosječnu preciznost od 89.60 %, prosječni odziv od 83.48 % i prosječnu F1-mjeru od 86.36 % (Sl. 0.8).

```
Average Accuracy from Cross-Validation: 0.8687
Average Precision from Cross-Validation: 0.8960
Average Recall from Cross-Validation: 0.8348
Average F1-Score from Cross-Validation: 0.8636
```

Sl. 0.8 Prosječne evaluacijske mjere za naivni Bayesov klasifikator

Na temelju razlike između prosječnih metrika i rezultata dobivenih pri trenutačnom pokretanju Naivnog Bayesovog klasifikatora, može se uočiti da su svi rezultati (točnost, preciznost, odziv i F1-mjera) u trenutačnom pokretanju bili nešto niži od prosječnih vrijednosti iz kros-validacije. Ove razlike također su vidljive u prikazu prosječnih vrijednosti matrice zabune (Sl. 0.9), gdje su brojevi ispravnih klasifikacija u trenutačnom modelu manji. Ovo može ukazivati na to da trenutačni skup podataka sadrži specifične izazove koji otežavaju modelu da postigne iste visoke performanse kao pri kros-validaciji. Moguće je da su podaci iz trenutačnog pokretanja teže klasificirani zbog buke, neravnoteže između klasa ili drugih faktora koji utječu na model.



Sl. 0.9 Matrica zabune Naivnog Bayesa dobivena kros-validacijom

Iako su rezultati naivnog Bayesovog klasifikatora solidni, njegovo oslanjanje na pretpostavke o normalnoj distribuciji i nezavisnosti značajki može biti ograničavajuće za složenije podatke, posebno kada su podaci ne prate normalnu distribuciju ili kada postoji visoka korelacija između značajki.

K-najbližih susjeda

Model K-najbližih susjeda (KNN) treniran je korištenjem GridSearchCV metode s peterostrukom unakrsnom validacijom, čime su određeni optimalni parametri: broj značajki bio je 13, broj susjeda 3, metoda prilagodbe težina „uniform“, a metrika udaljenosti Euklidska (p = 2). Nakon treninga, model je evaluiran na testnom skupu, gdje je postigao izuzetno visoku točnost od 99.35 %, preciznost od 100 %, odziv 98.68 % i F1-mjeru od 99.34 % (Sl. 0.10).

```
Accuracy: 0.9934640522875817
Precision: 1.0
Recall: 0.9868421052631579
F1-score: 0.9933774834437086
```

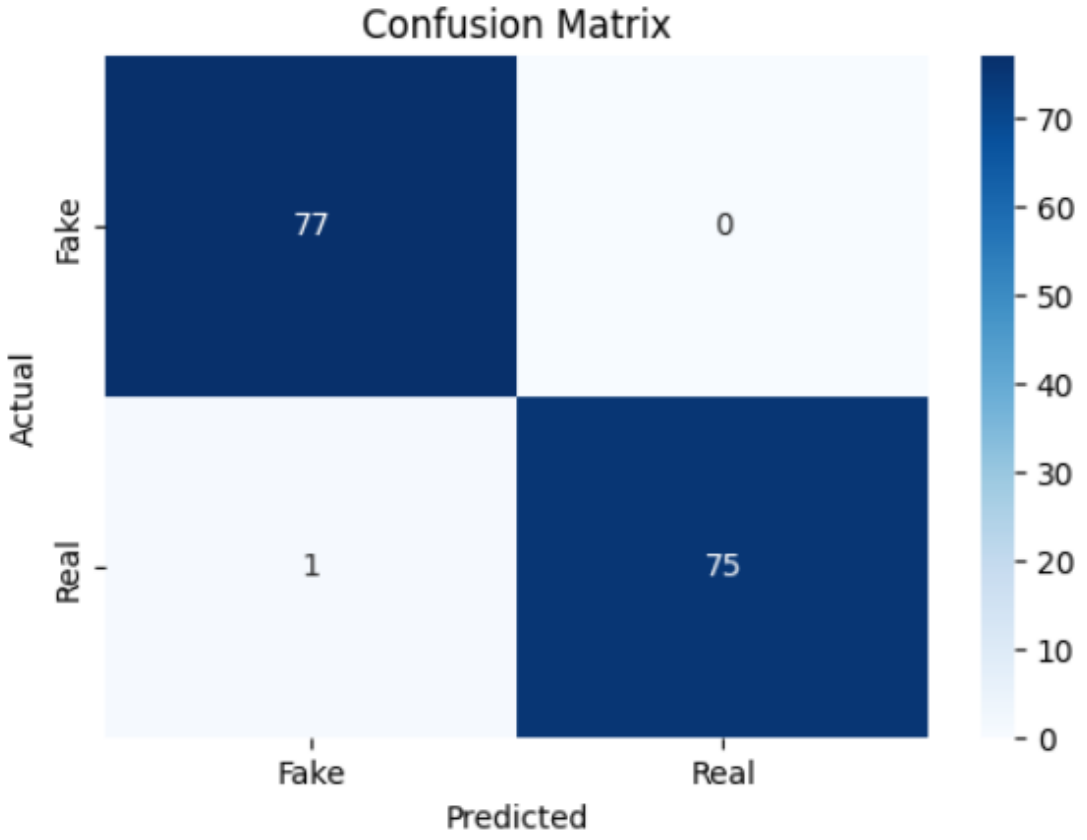
Sl. 0.10 Evaluacijske mjere za model K-najbližih susjeda

Analiza performansi po klasama (Sl. 0.11) pokazuje da je za klasu "Fake" postignuta F1-mjera od 99 %, uz preciznost od 99 % i odziv od 100 %. Za klasu "Real" postignuta je F1-mjera od 99 %, s preciznošću od 100 % i odzivom od 99 %.

	precision	recall	f1-score	support
0	0.99	1.00	0.99	77
1	1.00	0.99	0.99	76
accuracy			0.99	153
macro avg	0.99	0.99	0.99	153
weighted avg	0.99	0.99	0.99	153

Sl. 0.11 Klasifikacijski izvještaj za model K-najbližih susjeda

Matrica zabune (Sl. 0.12) pokazuje 77 pravilno klasificiranih "Fake" (True Negatives) i 75 pravilno klasificiranih "Real" uzoraka (True Positives), bez pogrešno klasificiranih "Real" (False Positives) i 1 pogrešno klasificirani "Fake" uzorak (False Negatives).



Sl. 0.12 Matrica zabune za model K-najbližih susjeda

Važno je uzeti u obzir ne samo konačne rezultate, već i prosječne metrike dobivene korištenjem kros-validacije (cv = 5). Prosječne metrike iz kros-validacije za KNN uključuju prosječnu točnost od 99.08 %, prosječnu preciznost od 98.95 %, prosječni odziv od 99.21 % i prosječnu F1-mjeru od 99.08 % (Sl. 0.13).

```
Average Accuracy from Cross-Validation: 0.9908
Average Precision from Cross-Validation: 0.9895
Average Recall from Cross-Validation: 0.9921
Average F1-Score from Cross-Validation: 0.9908
```

Sl. 0.13 Prosječne evaluacijske mjere za KNN

Na temelju razlike između prosječnih metrika iz kros-validacije i rezultata dobivenih pri trenutačnom pokretanju KNN klasifikatora, može se primijetiti da su rezultati u trenutačnom pokretanju vrlo blizu prosječnim vrijednostima iz kros-validacije. Naime, točnost u trenutačnom modelu iznosi 99.35 %, dok je prosječna točnost iz kros-validacije bila 99.08 %, što ukazuje na minimalne razlike u performansama modela. Preciznost od 100 %, odziv od 98.68 % i F1-mjera od 99.34 % također su vrlo blizu prosječnim vrijednostima iz kros-validacije, čime se potvrđuje stabilnost modela u trenutačnom pokretanju.

Ovi rezultati pokazuju da je KNN model dobro balansiran i izuzetno precizan u razlikovanju između klasa. GridSearchCV osigurao je optimalne hiperparametre, što je doprinijelo minimiziranju pogrešnih klasifikacija. Model se može smatrati pouzdanim i učinkovitim alatom za klasifikaciju u ovom zadatku.

Stroj potpornih vektora

Model SVM treniran je korištenjem GridSearchCV metode s peterostrukom unakrsnom validacijom, pri čemu su optimalni hiperparametri uključivali 11 značajki, C parametar postavljen na 10, gamma postavljen na "scale" i RBF kao jezgrena funkcija. Evaluacija modela na testnom skupu pokazala je savršenu točnost, preciznost, odziv i F1-mjeru (Sl. 0.14).

```
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1-score: 1.0
```

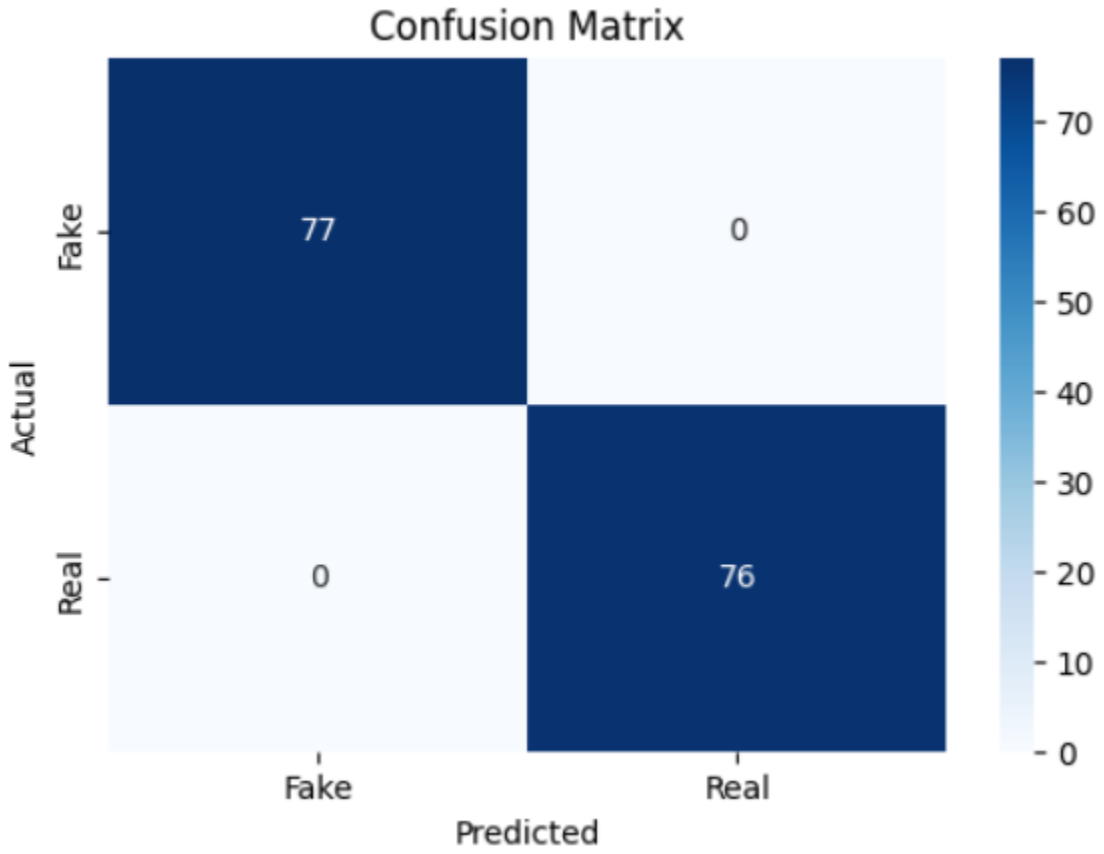
Sl. 0.14 Evaluacijske mjere za stroj potpornih vektora

Performanse po klasama (Sl. 0.15) ukazuju na potpunu uspješnost modela: za obje klase "Fake" i "Real" postignute su preciznost, odziv i F1-mjera od 100 %.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	77
1	1.00	1.00	1.00	76
accuracy			1.00	153
macro avg	1.00	1.00	1.00	153
weighted avg	1.00	1.00	1.00	153

Sl. 0.15 Klasifikacijski izvještaj za stroj potpornih vektora

Matrica zabune (Sl. 0.16) pokazuje da je svih 77 uzoraka klase "Fake" i svih 76 uzoraka klase "Real" ispravno klasificirano, bez ijednog pogrešno klasificiranog uzorka.



Sl. 0.16 Matrica zabune za stroj potpornih vektora

Korištenjem kros-validacije (cv=5) dobivena je prosječna točnost od 98.82 %, prosječna preciznost od 98.95 %, prosječni odziv od 98.69 % i prosječna F1-mjera od 98.81 % (Sl. 0.17)

```
Average Accuracy from Cross-Validation: 0.9882
Average Precision from Cross-Validation: 0.9895
Average Recall from Cross-Validation: 0.9869
Average F1-Score from Cross-Validation: 0.9881
```

Sl. 0.17 Prosječne evaluacijske mjere za SVM

Na temelju razlike između prosječnih metrika iz kros-validacije i rezultata dobivenih pri trenutačnom pokretanju SVM klasifikatora, može se primijetiti da su rezultati u trenutačnom pokretanju vrlo blizu prosječnim vrijednostima iz kros-validacije. Naime, točnost u

trenutačnom modelu iznosi 100 %, dok je prosječna točnost iz kros-validacije bila 99.82 %, što ukazuje na minimalne razlike u performansama modela. Preciznost, odziv i F1-mjera od 100 % također su vrlo blizu prosječnim vrijednostima iz kros-validacije, čime se potvrđuje stabilnost modela u trenutačnom pokretanju.

Ovi rezultati potvrđuju da je SVM model iznimno precizan i pouzdan za klasifikacijske probleme. Kombinacija pipelinea, odabira značajki i optimizacije hiperparametara rezultirala je modelom koji savršeno razlikuje klase. Visoka točnost i balansirane performanse između klasa čine ovaj model idealnim za zadatke gdje je ključna visoka pouzdanost klasifikacije.

Logistička regresija

Model logističke regresije optimiziran je korištenjem GridSearchCV, gdje su optimalni parametri identificirani kao: regularizacijski parametar C = 0.1, penalizacija l2, solver "saga", te broj značajki 13. Evaluacija na testnom skupu rezultirala je ukupnom točnošću od 84.31 %, te preciznošću, odzivom i F1-mjerom od 84.21 % (Sl. 0.18).

```
Accuracy: 0.8431372549019608
Precision: 0.8421052631578947
Recall: 0.8421052631578947
F1-score: 0.8421052631578947
```

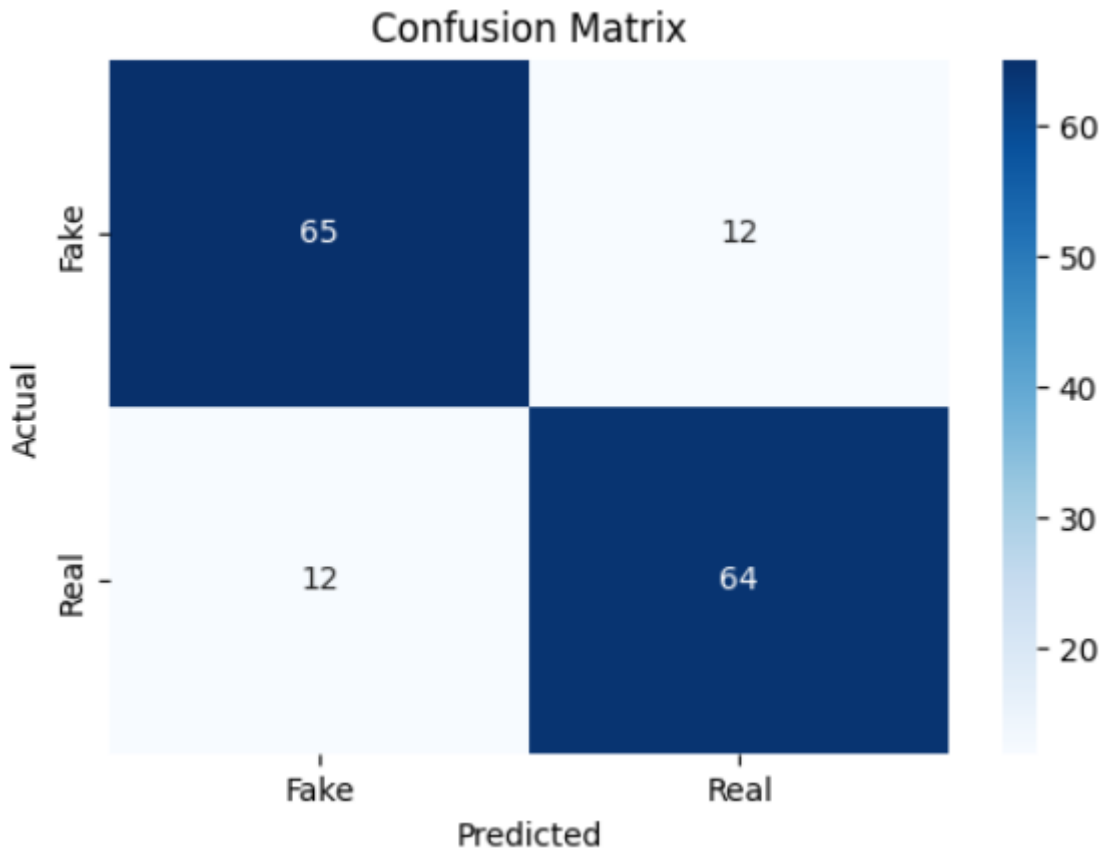
Sl. 0.18 Evaluacijske mjere za logističku regresiju

Analiza performansi po klasama (Sl. 0.19) ukazuje da za obje klase "Fake" i "Real" postignuta je preciznost, odziv i F1-mjera od 84 %.

	precision	recall	f1-score	support
0	0.84	0.84	0.84	77
1	0.84	0.84	0.84	76
accuracy			0.84	153
macro avg	0.84	0.84	0.84	153
weighted avg	0.84	0.84	0.84	153

Sl. 0.19 Klasifikacijski izvještaj za logističku regresiju

Matrica zabune (Sl. 0.20) ukazuje na to da je 65 uzorka klase "Fake" i 64 uzoraka klase "Real" ispravno klasificirano, dok je po 12 uzoraka iz obje klase pogrešno klasificirano.



Sl. 0.20 Matrica zabune za logističku regresiju

Korištenjem kros-validacije (cv=5) dobivena je prosječna točnost od 84.78 %, prosječna preciznost od 86.53 %, prosječni odziv od 82.43 % i prosječna F1-mjera od 84.37 % (Sl. 0.21).

```
Average Accuracy from Cross-Validation: 0.8478
Average Precision from Cross-Validation: 0.8653
Average Recall from Cross-Validation: 0.8243
Average F1-Score from Cross-Validation: 0.8437
```

Sl. 0.21 Prosječne evaluacijske mjere za logističku regresiju

Na temelju razlike između prosječnih metrika iz kros-validacije i rezultata dobivenih pri trenutačnom pokretanju logističke regresije, može se primijetiti da su rezultati u trenutačnom pokretanju (točnost, preciznost, odziv i F1-mjera) vrlo blizu prosječnim vrijednostima iz kros-validacije. Ove male razlike mogu ukazivati na to da trenutačni skup podataka nije sadržavao specifične izazove koji bi značajno utjecali na performanse modela.

Iako je logistička regresija pokazala solidne rezultate, njezina učinkovitost može biti ograničena u slučajevima kada su značajke međusobno nelinearno povezane ili kada podaci sadrže složenije obrasce.

Slučajne šume

Optimizacijom klasifikatora slučajnih šuma identificirani su optimalni hiperparametri: 13 značajki, 200 stabala, maksimalnu dubinu 10 i isključivanje bootstrap uzorkovanja. Na testnom skupu model je postigao visoku točnost od 98.69 %, s preciznošću od 100 %, odziv od 97.37 % i F1-mjerom od 98.67 % (Sl. 0.22).

```
Accuracy: 0.9869281045751634
Precision: 1.0
Recall: 0.9736842105263158
F1-score: 0.9866666666666667
```

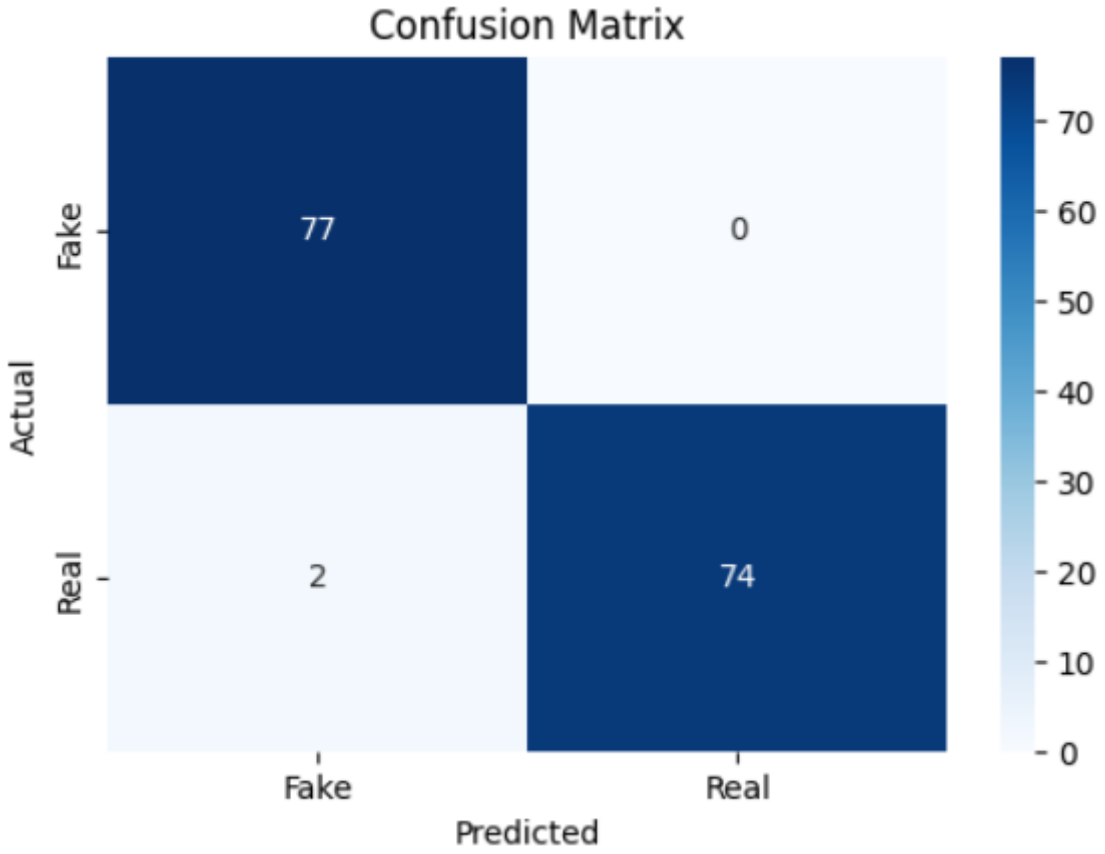
Sl. 0.22 Evaluacijske mjere za klasifikator slučajnih šuma

Analiza po klasama (Sl. 0.23) pokazuje da klasa "Fake" ima preciznost od 97 %, odziv od 100 % i F1-mjeru od 99 %, dok klasa "Real" ima preciznost od 100 %, odziv od 97 % i F1-mjeru od 99 %.

	precision	recall	f1-score	support
0	0.97	1.00	0.99	77
1	1.00	0.97	0.99	76
accuracy			0.99	153
macro avg	0.99	0.99	0.99	153
weighted avg	0.99	0.99	0.99	153

Sl. 0.23 Klasifikacijski izvještaj za klasifikator slučajnih šuma

Matrica zabune (Sl. 0.24) otkriva da je 77 uzorka klase "Fake" i 74 uzoraka klase "Real" ispravno klasificirano, bez pogrešno klasificiranih "Real" (False Positives) i 2 pogrešno klasificirana "Fake" uzorka (False Negatives).



Sl. 0.24 Matrica zabune za klasifikator slučajnih šuma

Korištenjem kros-validacije (cv=5) dobivena je prosječna točnost od 96.85 %, prosječna preciznost od 97.43 %, prosječni odziv od 96.34 % i prosječna F1-mjera od 96.82 % (Sl. 0.25).

```
Average Accuracy from Cross-Validation: 0.9685
Average Precision from Cross-Validation: 0.9743
Average Recall from Cross-Validation: 0.9634
Average F1-Score from Cross-Validation: 0.9682
```

Sl. 0.25 Prosječne evaluacijske mjere za slučajne šume

Rezultati potvrđuju da je Random Forest model visoko učinkovit, s minimalnim brojem pogrešnih klasifikacija, te se može smatrati pouzdanim i učinkovitim alatom za klasifikaciju u ovom zadatku.

Stablo odluke

Decision Tree model treniran je s optimalnim hiperparametrima: 11 značajki, maksimalna dubina stabla bez ograničenja, minimalni broj uzoraka u listu 1, minimalni broj uzoraka za podjelu čvora 2 i kriterij Gini. Na skupu za testiranje model je postigao točnost od 93.46 %, s preciznošću od 90.24 %, odzivom od 97.37 % i F1-mjerom od 93.67 % (Sl. 0.26).

```
Accuracy: 0.934640522875817
Precision: 0.9024390243902439
Recall: 0.9736842105263158
F1-score: 0.9367088607594937
```

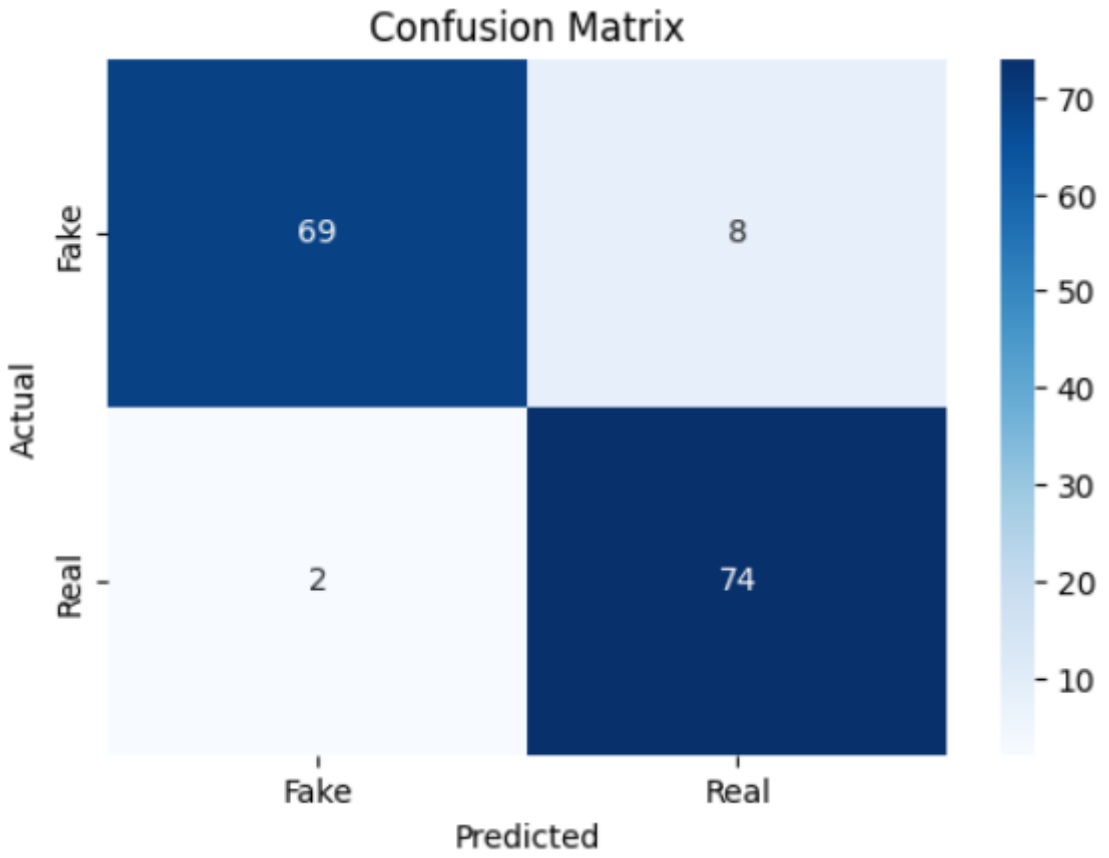
Sl. 0.26 Evaluacijske mjere za stablo odluke

Analiza performansi po klasama (Sl. 0.27) pokazuje da klasa "Fake" ima preciznost od 97 %, odziv od 90 % i F1-mjeru od 93 %, dok klasa "Real" ima preciznost od 90 %, odziv od 97 % i F1-mjeru od 94 %.

	precision	recall	f1-score	support
0	0.97	0.90	0.93	77
1	0.90	0.97	0.94	76
accuracy			0.93	153
macro avg	0.94	0.93	0.93	153
weighted avg	0.94	0.93	0.93	153

Sl. 0.27 Klasifikacijski izvještaj za stablo odluke

Matrica zabune (Sl. 0.28) pokazuje da je 69 uzoraka klase "Fake" i 74 uzoraka klase "Real" ispravno klasificirano, dok su 2 uzoraka pogrešno klasificirana kao "Fake" (False Positives) i 8 kao "Real" (False Negatives).



Sl. 0.28 Matrica zabune za stablo odluke

Korištenjem kros-validacije (cv=5) dobivena je prosječna točnost od 91.08 %, prosječna preciznost od 91.14 %, 91.09 % i prosječna F1-mjera od 91.09 % (Sl. 0.29).

```
Average Accuracy from Cross-Validation: 0.9108
Average Precision from Cross-Validation: 0.9114
Average Recall from Cross-Validation: 0.9109
Average F1-Score from Cross-Validation: 0.9109
```

Sl. 0.29 Prosječne evaluacijske mjere za stablo odluke

Model stabla odluke pokazao je dobre rezultate potvrđujući svoju sposobnost razlikovanja stvarnih i manipuliranih audiozapisa.

CatBoost klasifikator

CatBoost klasifikator postigao je na testnom skupu točnost od 99.35 %, s preciznošću od 100 %, odzivom od 98.68 % i F1-mjerom od 99.34 % (Sl. 0.30).

```
Accuracy: 0.9934640522875817
Precision: 1.0
Recall: 0.9868421052631579
F1-score: 0.9933774834437086
```

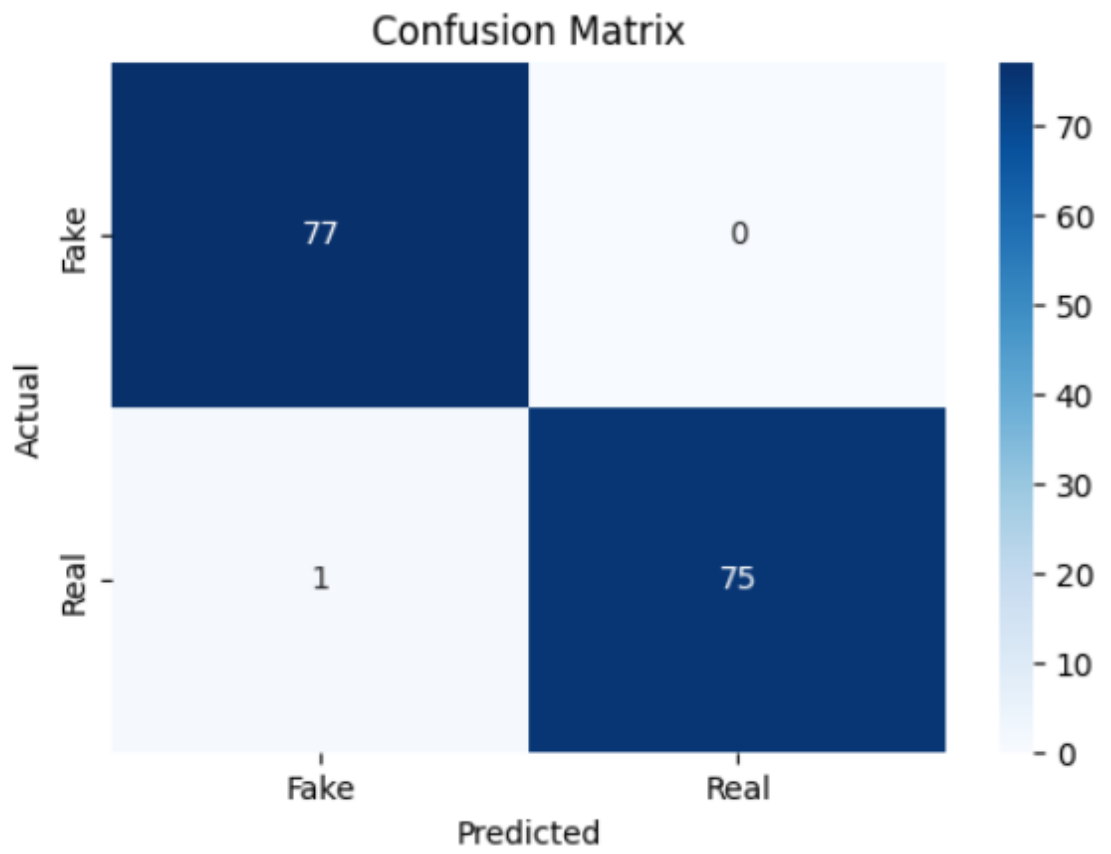
Sl. 0.30 Evaluacijske mjere za CatBoost klasifikator

Analiza performansi po klasama (Sl. 0.31) pokazuje da za klasu "Fake", preciznost iznosi 99 %, odziv 100 % i F1-mjera 99 %, dok za klasu "Real", preciznost iznosi 100 %, odziv 99 % i F1-mjera 99 %.

	precision	recall	f1-score	support
0	0.99	1.00	0.99	77
1	1.00	0.99	0.99	76
accuracy			0.99	153
macro avg	0.99	0.99	0.99	153
weighted avg	0.99	0.99	0.99	153

Sl. 0.31 Klasifikacijski izvještaj za CatBoost klasifikator

Matrica zabune (Sl. 0.32) pokazuje da je model ispravno klasificirao 75 uzorka klase "Real" (True Positives) i 77 uzorka klase "Fake" (True Negatives). Pogrešne klasifikacije uključuju 1 False Negative i 0 False Positive.



Sl. 0.32 Matrica zabune za CatBoost klasifikator

Korištenjem kros-validacije (cv=5) dobivena je prosječna točnost od 98.16 %, prosječna preciznost od 98.69 %, prosječni odziv od 97.65 % i prosječna F1-mjera od 98.15 % (Sl. 0.33).

```
Average Accuracy from Cross-Validation: 0.9816
Average Precision from Cross-Validation: 0.986
Average Recall from Cross-Validation: 0.9765
Average F1-Score from Cross-Validation: 0.9815
```

Sl. 0.33 Prosječne evaluacijske mjere za CatBoost klasifikator

CatBoost klasifikator pokazao je izvanredne rezultate, čime je potvrđena njegova sposobnost visokokvalitetne klasifikacije. Visoka točnost, preciznost, odziv i F1-mjera ukazuju na to da je model vrlo dobro generalizirao na skupu za testiranje. Ovi rezultati čine CatBoost vrlo pogodnim za zadatke koji zahtijevaju visoku pouzdanost i preciznost, kao što je razlikovanje stvarnih i manipuliranih audiozapisa.

Ekstremno povećanje gradijenta

Na testnom skupu XGBoost postigao je točnost od 98.69 %, te preciznost, odziv i F1-mjeru od 98.68 % (Sl. 0.34).

```
Accuracy: 0.9869281045751634
Precision: 0.9868421052631579
Recall: 0.9868421052631579
F1-score: 0.9868421052631579
```

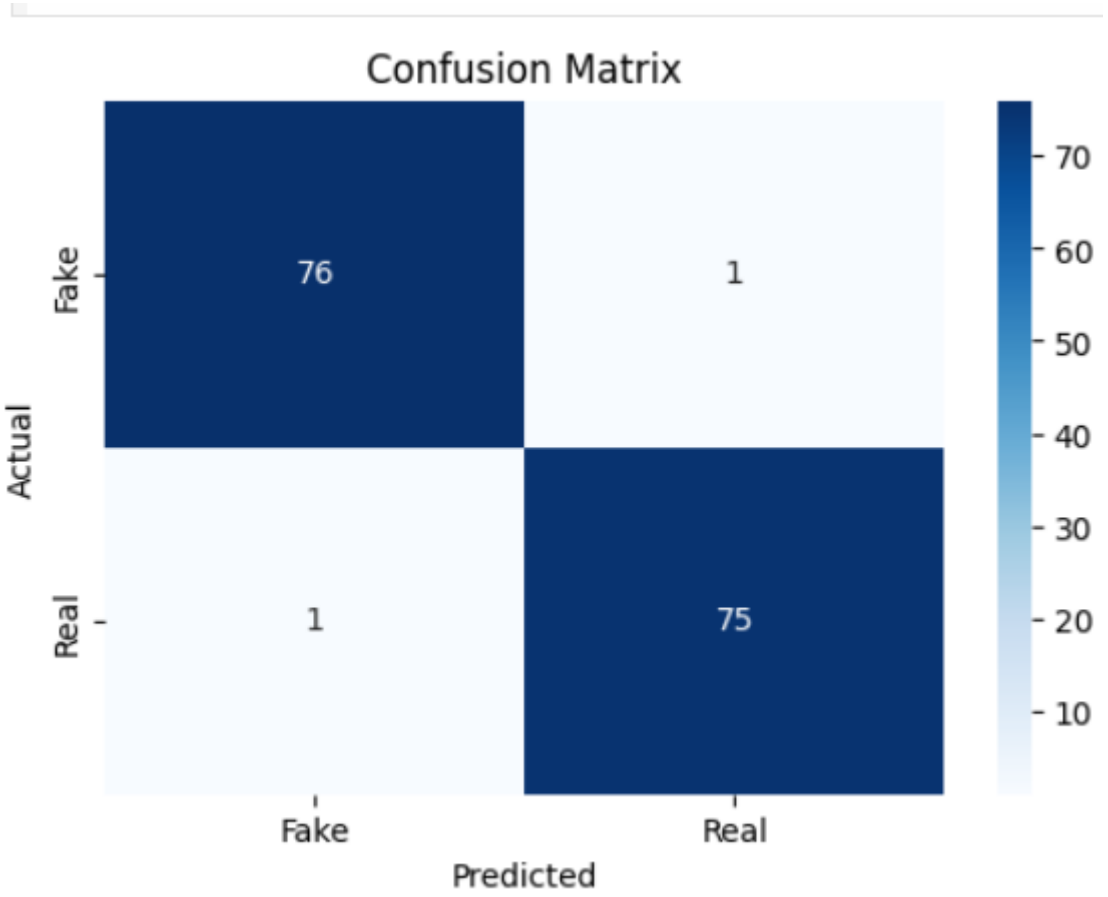
Sl. 0.34 Evaluacijske mjere za ekstremno povećanje gradijenta

Iz izvještaja po klasama (Sl. 0.35) ukazuje da za obje klase "Fake" i "Real" postignute su preciznost, odziv i F1-mjera od 99 %.

	precision	recall	f1-score	support
0	0.99	0.99	0.99	77
1	0.99	0.99	0.99	76
accuracy			0.99	153
macro avg	0.99	0.99	0.99	153
weighted avg	0.99	0.99	0.99	153

Sl. 0.35 Klasifikacijski izvještaj za ekstremno povećanje gradijenta

Matrica zabune (Sl. 0.36) pokazuje da je model ispravno klasificirao 76 uzorka klase "Fake" i 75 uzorka klase "Real". Pogrešno je klasificirano po 1 uzorak iz klase "Fake" i iz klase "Real".



Sl. 0.36 Matrica zabune za ekstremno povećanje gradijenta

Za analizu ekstremnog povećanja gradijenta korištene su i prosječne metrike dobivene kroz petostruku stratificiranu validaciju, umjesto klasične kros-validacije. Ovaj pristup omogućio je precizniju procjenu performansi modela na različitim podskupovima podataka, pružajući uvid u konzistentnost rezultata kroz više iteracija. Dobivena je prosječna točnost od 97.25 %, prosječna preciznost od 98.38 %, prosječni odziv od 96.08 % i prosječna F1-mjera od 97.20 % (Sl. 0.37).

```
Average Accuracy: 0.9725
Average Precision: 0.9838
Average Recall: 0.9608
Average F1-score: 0.9720
```

Sl. 0.37 Prosječne evaluacijske mjere za ekstremno povećanje gradijenta

XGBoost klasifikator pokazao je izvanredne rezultate, čime je potvrđena njegova sposobnost visokokvalitetne klasifikacije. Ovim rezultatima XGBoost se pokazao kao učinkovit model za ovaj zadatak, s visokom preciznošću i osjetljivošću, što ga čini dobrim odabirom za klasifikacijske zadatke.

Model dugoročno – kratkoročne memorije

Model je treniran s pomoću Adam optimizatora i unakrsne entropije kao funkcije gubitka, uz primjenu early stopping mehanizma koji je prekinuo trening zbog nedostatka poboljšanja na validacijskom skupu nakon 5 epoha. Treniranje je zaustavljeno u šesnaestoj epohi i model je postigao točnost na testnom skupu od 94.77 % uz gubitak od 21.32 % (Sl. 0.38).

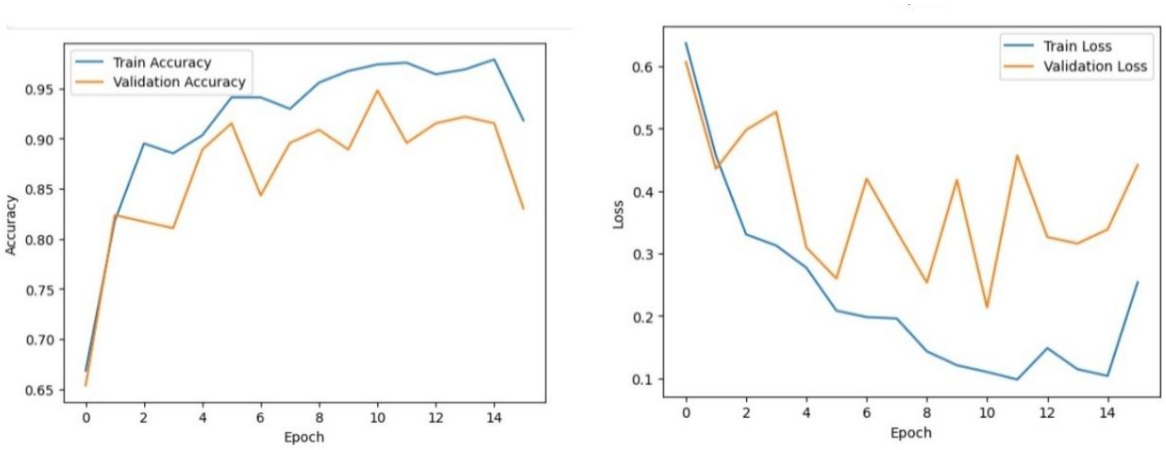
```
Test Loss: 0.21319432556629 18
Test Accuracy: 0.9477124214172363
```

Sl. 0.38 Gubitak i točnost LSTM modela

Na temelju prikazanih grafova (Sl. 0.39) može se zaključiti da LSTM model pokazuje određene znakove pretreniranosti, iako ima visoku točnost na testnom skupu.

Na lijevom grafu, koji prikazuje točnost, vidi se da točnost na trening skupu (plava linija) raste i doseže gotovo 100%. S druge strane, validacijska točnost (narančasta linija) također raste, ali nakon određenog broja epoha počinje oscilirati. To upućuje da model dobro uči na trening podacima, ali ne generalizira jednako dobro na neviđenim podacima.

Na desnom grafu, koji prikazuje krivulje gubitka, vidi se da gubitak na trening skupu opada, ali validacijski gubitak nakon početnog pada počinje oscilirati, što je još jedan jasan znak pretreniranosti.



Sl. 0.39 Krivulje gubitka i točnosti na skupovima za treniranje i validaciju (LSTM)

Testiranjem različitih veličina modela, jačina L2 regularizacije i postotaka Dropout slojeva, ovaj model je ostvario najbolje rezultate među isprobanima te je stoga odabran u konačnoj

implementaciji. Unatoč tome, vidljivo je da i dalje dolazi do pretreniranosti, što sugerira da bi dodatna eksperimentiranja s arhitekturom i hiperparametrima dodatno smanjiti pretreniranost i poboljšati generalizaciju modela.

Konvolucijska neuronska mreža

Model je treniran s pomoću Adam optimizatora i kategorijske unakrsne entropije kao funkcije gubitka, uz primjenu early stopping mehanizma koji je prekinuo trening zbog nedostatka poboljšanja na validacijskom skupu nakon 5 epoha, te iz tih razloga je trening prekinut u 21 epohi.

Krajnji rezultati (Sl. 0.40) pokazali su da je model postigao 90.79 % točnosti na validacijskom skupu s gubitkom od 32.38 %.

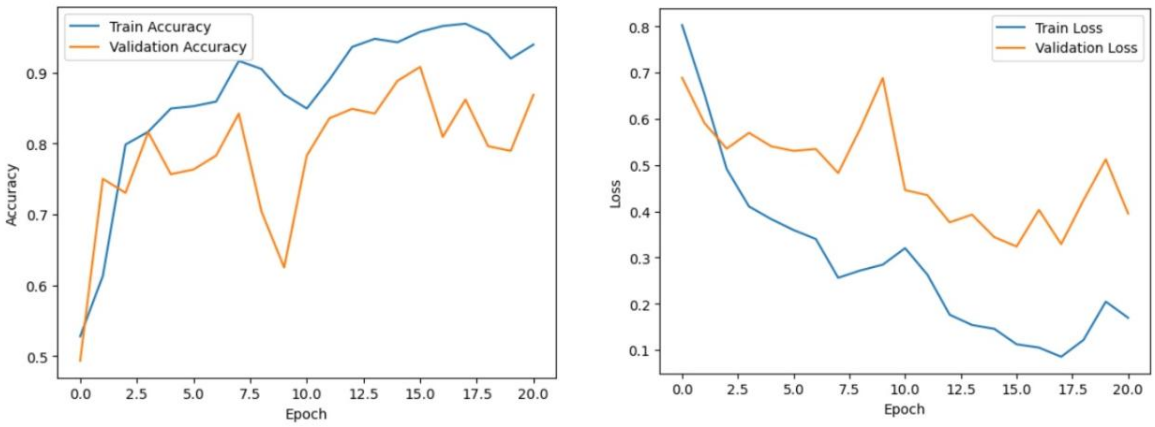
```
Validation Loss: 0.32384034991264343
Validation Accuracy: 0.9078947305679321
```

Sl. 0.40 Gubitak i točnost CNN modela

Iako je model postigao visoku točnost na neviđenim podacima, na temelju prikazanih grafova (Sl. 0.41) može se zaključiti da CNN model pokazuje određene znakove pretreniranja, unatoč korištenju Dropout slojeva i early stopping mehanizma.

Kao i kod LSTM modela, na lijevom grafu, vidi se da točnost na trening skupu (plava linija) raste i doseže vrlo visoke vrijednosti, blizu 100 %, dok validacijska točnost (narančasta linija) također raste, ali nakon određenog broja epoha počinje oscilirati i ne prati više stabilno krivulju točnosti trening skupa. Ove oscilacije upućuju na to da model ne generalizira jednako dobro na neviđene podatke te da dolazi do određene pretreniranosti.

Na desnom grafu, vidi se da gubitak na trening skupu opada, ali validacijski gubitak nakon početnog pada počinje oscilirati. Na nekoliko mjesta validacijski gubitak raste, što je još jedan pokazatelj da model počinje previše učiti specifične obrasce iz trening podataka, umjesto da uči generalizirane značajke koje bi bolje funkcionirale na novim podacima.



Sl. 0.41 Krivulje gubitka i točnosti na skupovima za treniranje i validaciju (CNN)

Eksperimentiranjem s različitim postotcima Dropout slojeva i arhitekturama modela, ovaj model je ostvario najbolje rezultate među isprobanima te je stoga odabran u konačnoj implementaciji. Unatoč tome, vidljivo je da i dalje dolazi do pretreniranosti, što sugerira da bi dodatna eksperimentiranja s arhitekturom i hiperparametrima, proširenje skupa podataka ili implementacija naprednijih arhitektura poput transfer learninga bi mogle poboljšati performanse modela.

Vremenska složenost modela

Vremenska složenost modela strojnog učenja odnosi se na vrijeme potrebno za treniranje i evaluaciju modela. U kontekstu ovog istraživanja, procjena vremenske složenosti bila je ključna za razumijevanje koliko vremena je potrebno za treniranje različitih modela i njihovu provjeru na testnom skupu. Ovo je osobito važno kada je potrebno optimizirati trajanje obrade podataka, posebno kod velikih skupova podataka ili modela koji zahtijevaju dugotrajno treniranje, poput dubokih neuronskih mreža.

Tablica (Sl. 0.1) prikazuje vrijeme treniranja svakog modela, kao i vrijeme za evaluaciju na testnom skupu. Ove informacije omogućuju bolje razumijevanje vremenske učinkovitosti svakog modela i pomažu u odabiru optimalnog pristupa s obzirom na trajanje obrade.

Model	Vrijeme treniranja (s)	Vrijeme evaluacije (s)
Naivni Bayes	0.0034	0.3017
K-najbližih susjeda	8.9659	0.3236
Stroj potpornih vektora	63.5524	0.2169
Logistička regresija	9.6638	0.7713
Slučajne šume	221.2837	0.2820
Stablo odluke	20.7950	0.2338
CatBoost klasifikator	4.1665	0.2088
Ekstremno povećanje gradijenta	0.1509	0.2258
Model dugoročno – kratkoročne memorije	103.2923	2.0907
Konvolucijska neuronska mreža	431.9923	2.7305

Sl. 0.1 Vremenska složenost modela

Rezultati

U ovom istraživanju testirani su različiti modeli za klasifikaciju s ciljem optimizacije točnosti predviđanja manipuliranih audiozapisa. Evaluacija je provedena na temelju nekoliko ključnih metrika, uključujući točnost, preciznost, odziv i F1-mjeru. Osim toga, analizirana je i vremenska složenost modela, odnosno vrijeme potrebno za treniranje i evaluaciju, što je od posebne važnosti pri radu s velikim skupovima podataka.

Iz rezultata vremenskih složenosti modela (Sl. 0.1) je vidljivo da jednostavniji modeli, poput Naivnog Bayesa i metode ekstremnog povećanja gradijenta (XGBoost), imaju zanemarivo vrijeme treniranja i evaluacije. S druge strane, kompleksniji modeli poput LSTM-a i CNN-a zahtijevaju značajno duže vrijeme treniranja, što može predstavljati izazov pri radu s velikim skupovima podataka.

U nastavku je prikazana tablica s trenutačnim i prosječnim točnostima modela (Sl. 0.1):

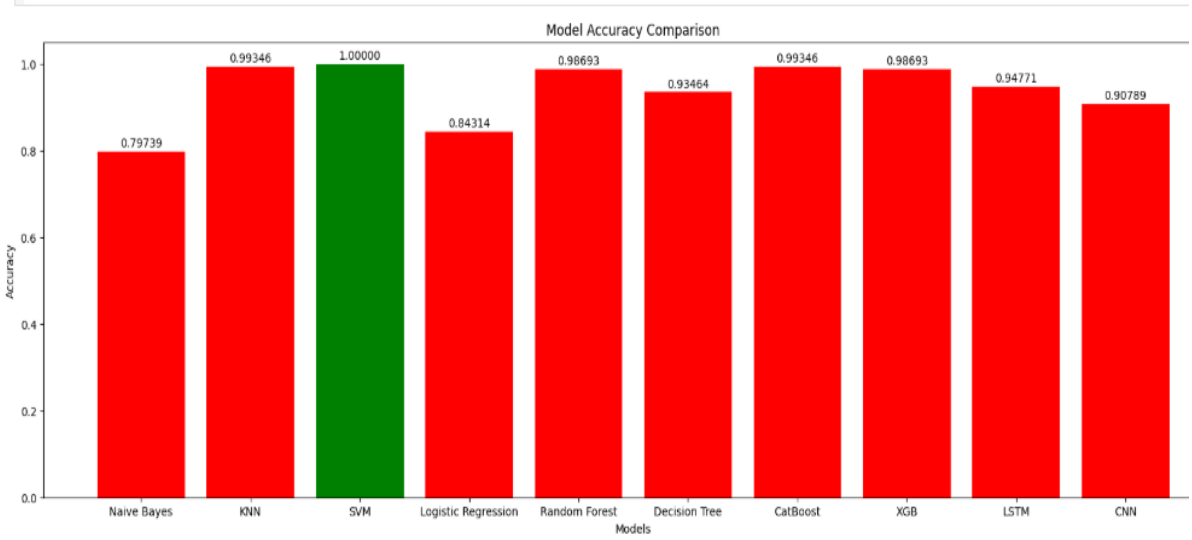
Model	Trenutačna točnost (%)	Prosječna točnost (%)
Naivni Bayes	79.74	86.87
K-najbližih susjeda	99.35	99.08
Stroj potpornih vektora	100	98.82
Logistička regresija	84.31	84.78
Slučajne šume	98.69	96.85
Stablo odluke	93.46	91.08
CatBoost klasifikator	99.35	98.16
Ekstremno povećanje gradijenta	98.69	97.25
Model dugoročno – kratkoročne memorije	94.77	-
Konvolucijska neuronska mreža	90.79	-

Sl. 0.1 Tablica trenutačnih i prosječnih točnosti modela

Analizom točnosti modela, najbolji rezultat u trenutačnoj evaluaciji postigao je SVM model, ostvarivši točnost od 100 %, dok je model K-najbližih susjeda (KNN) imao nešto bolju prosječnu točnost kroz kros-validaciju od prosječne točnosti SVM-a. Ovi rezultati ukazuju na izuzetnu sposobnost oba modela u prepoznavanju obrazaca i generalizaciji podataka, pri čemu SVM postiže vrhunsku preciznost na testnom skupu, dok KNN pokazuje veću konzistentnost performansi kroz različite podjele podataka.

Za LSTM i CNN modele nisu izračunate prosječne točnosti zbog njihove visoke vremenske složenosti, koja bi značajno produžila trajanje eksperimenta.

Za bolju ilustraciju performansi modela, prikazan je stupčasti grafikon koji uspoređuje trenutačne točnosti svih evaluiranih modela (Sl. 0.2). SVM model, označen zelenim stupcem, jasno se ističe kao model s najvišom trenutačnom točnošću u usporedbi s ostalim modelima.



Sl. 0.2 Prikaz trenutačnih točnosti modela

Izazovi u razvoju sustava za detekciju manipuliranih audiozapisa

Razvoj sustava detekcije manipuliranih audiozapisa susreće se s brojnim izazovima i ograničenjima koja značajno utječu na učinkovitost i točnost detekcijskih sustava.

Najznačajniji izazovi su:

- Neprestani napredak sustava za generiranje manipuliranih audiozapisa
  - Moderne tehnike generiranja audiozapisa omogućuju stvaranje zvučnih zapisa koji zvuče gotovo identično stvarnim snimkama. Brzi napredak ovih tehnologija neprestano smanjuje prepoznatljive zvučne nepravilnosti, čineći detekciju sve izazovnijom.
- Ograničenja u pristupu podacima
  - Jedan od glavnih izazova je nedostatak kvalitetnih skupova podataka koji sadržavaju sintetičke i stvarne glasove na kojima se modeli mogu učiti. Bez dovoljno raznovrsnih podataka, trenirani modeli mogu biti pristrani i loše generalizirati.
- Računalna i energetska složenost
  - Analiza većih količina podataka, zahtijeva značajne računalne resurse. Osim toga, dodatni izazov predstavlja optimizacija modela za postizanje ravnoteže između preciznosti detekcije i brzine izvođenja, osobito u real-time sustavima.
- Etička i pravna pitanja
  - Jedan od etičkih izazova u detekciji sintetički generiranih audiozapisa je mogućnost lažno pozitivnih rezultata, gdje stvarni govor može biti nepravedno označen kao sintetički, što može imati ozbiljne posljedice. Osim toga, prikupljanje stvarnih i sintetičkih glasova za obuku sustava može izazvati zabrinutost pojedinaca za zaštitu njihove privatnosti i sigurnosti.

Interdisciplinarni pristupi

Detekcija sintetički generiranih audiozapisa predstavlja izazov koji zahtijeva primjenu tehnika i metoda iz različitih znanstvenih disciplina, uključujući računalnu viziju, obradu prirodnog jezika i duboko učenje. Suradnja između tih područja omogućuje razvoj snažnijih alata za prepoznavanje manipuliranih audiozapisa.

Računalna vizija, iako prvenstveno fokusirana na analizu slika i videozapisa, može značajno doprinijeti detekciji deepfake sadržaja, čak i kada se problem odnosi na zvučne zapise. Jedna od ključnih metoda koja se koristi u tom kontekstu je analiza spektrograma koja je i korištena u ovom istraživanju. Spektrogrami su vizualni prikazi amplituda zvučnih valova prema vremenu i frekvenciji, što omogućuje njihovu obradu kao slika. Korištenjem tehnika računalne vizije, poput konvolucijskih neuronskih mreža, moguće je identificirati specifične vizualne nesavršenosti koje proizlaze iz generiranja manipuliranih audiozapisa.

Obrada prirodnog jezika (NLP) također igra ključnu ulogu u prepoznavanju deepfake sadržaja. Analiza sintaktičkih i semantičkih obrazaca u govoru omogućuje otkrivanje nepravilnosti ili neuobičajenih jezičnih struktura koje se ne podudaraju s tipičnim načinima izražavanja. Korištenjem naprednih NLP modela, moguće je detektirati nelogičnosti, gramatičke pogreške ili neobičnu intonaciju u sintetičkom govoru. Ove tehnike također pomažu prepoznati monotoniju, neprirodno brz ili spor tempo govora ili neskladnosti u govornim uzorcima koji odstupaju od prirodnih ljudskih reakcija.

Budući pristupi

Moguće buduće pristupanje ovoj analizi moglo bi uključivati uključivanje ljudskih sudionika u zadatak prepoznavanja stvarnih i lažnih audiozapisa, s ciljem usporedbe ljudske percepcije zvuka sa sposobnostima stroja. Ovaj pristup omogućava direktnu usporedbu između ljudskog i strojnog prepoznavanja, istražujući razlike u točnosti i brzini koje svaka strana pravi, odnosno pomaže u razumijevanju prednosti i slabosti svakog sustava.

Ljudski sudionici bi prvo bili podučeni na određenom broju audiozapisa, u kojem bi im bilo jasno označeno koji su zapisi stvarni, a koji lažni. Ova početna faza obuke pruža ljudima potrebne informacije i kontekst za uspješno prepoznavanje zvukova, što je ključno za njihovu kasniju evaluaciju. Nakon faze obuke, sudionici bi se testirali na neviđenim primjerima, gdje bi njihov zadatak bio odrediti je li zvuk stvaran ili lažan.

Paralelno s tim, strojni model bio bi testiran na istim primjerima. Korištenje istih primjera za oba pristupa omogućava pravednu usporedbu rezultata, što je ključno za preciznu analizu njihovih performansi. Nakon što su podaci prikupljeni, rezultati bi bili analizirani kako bi se dobio uvid u snage i slabosti ljudske percepcije u odnosu na strojne metode prepoznavanja. Analizom točnosti odgovora i brzine prepoznavanja mogli bismo bolje razumjeti kako ljudi i strojevi pristupaju zadatku prepoznavanja zvuka, što može pružiti vrijedne smjernice za unapređenje metoda prepoznavanja u budućnosti.

Zaključak

Cilj ovog istraživanja bio je razviti sustav koji može učinkovito razlikovati stvarne ljudske glasove od sintetičkih, s posebnim naglaskom na povećanje sigurnosti digitalne komunikacije. Korištenjem različitih tehnika strojnog učenja, postignuta je zadovoljavajuća razina točnosti u prepoznavanju manipuliranih audiozapisa. Time se omogućava budući razvoj učinkovitih alata koji mogu biti ključni u zaštiti privatnosti pojedinaca, smanjenju rizika od dezinformacija i zloupotrebe u digitalnom okruženju. S obzirom na ubrzani napredak tehnologija za generiranje sintetičkih glasova, sustavi za detekciju postaju nužni kako bi se očuvao integritet i povjerenje u komunikaciji. Iako su postignuti zadovoljavajući rezultati, potrebno je kontinuirano unapređivanje sustava i implementacija novih tehnika, kako bi se osigurala učinkovitost i preciznost u prepoznavanju novih oblika manipulacija u budućnosti.

Literatura

[1] I. Mutica, S. Mihalache and D. Burileanu, "Synthetic Speech Detection Using Deep Neural Networks," *2024 47th International Conference on Telecommunications and Signal Processing (TSP)*, Prague, Czech Republic, 2024, str. 53-57, doi: 10.1109/TSP63128.2024.10605922.

[2] I. Altalahin, S. AlZu'bi, A. Alqudah and A. Mughaid, "Unmasking the Truth: A Deep Learning Approach to Detecting Deepfake Audio Through MFCC Features," *2023 International Conference on Information Technology (ICIT)*, Amman, Jordan, 2023, str. 511-518, doi: 10.1109/ICIT58056.2023.10226172.

[3] G. S, S. G. A, J. D and A. J, "Deepfake Video Prediction Using Attention-Based CNN and Mel-Frequency Cepstral Coefficients," *2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Trichirappalli, India, 2024, str. 1-6, doi: 10.1109/ICEEICT61591.2024.10718393

[4] M. Dua, S. Meena, Neelam, Amisha and N. Chakravarty, "Audio Deepfake Detection Using Data Augmented Graph Frequency Cepstral Coefficients," *2023 International Conference on System, Computation, Automation and Networking (ICSCAN)*, PUDUCHERRY, India, 2023, str. 1-6, doi: 10.1109/ICSCAN58655.2023.10395679.

[5] R. Anagha, A. Arya, V. H. Narayan, S. Abhishek and T. Anjali, "Audio Deepfake Detection Using Deep Learning," *2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART)*, Moradabad, India, 2023, str. 176-181, doi: 10.1109/SMART59791.2023.10428163.

[6] Walczyna T., Piotrowski Z., *Overview of Voice Conversion Methods Based on Deep Learning,* MDPI, (2023, veljača). Poveznica: Overview of Voice Conversion Methods Based on Deep Learning; pristupljeno 23. studenog 2024.

[7] Laumann F., *Text-to-Speech 101: The Ultimate Guide,* Medium, (2023, studeni). Poveznica: Text-to-Speech 101: The Ultimate Guide | by Felix Laumann, PhD | NeuralSpace | Medium; pristupljeno 23. studenog 2024.

[8] Kothadiya D, Pise N, Bedekar M., *Different Methods Review for Speech to Text and Text to Speech Conversion*, 175, 20 (2020), str. 9-12.

[9] Bird J. J., *DEEP-VOICE: DeepFake Voice Recognition,* Kaggle (2024). Poveznica: https://www.kaggle.com/datasets/birdy654/deep-voice-deepfake-voice-recognition; pristupljeno 17. listopada 2024.

[10] Anantharaman R, *Real-Time Voice Cloning: A Deep Dive into Revolutionary TTS Technologies,* Medium (2024, srpanj). Poveznica: Real-Time Voice Cloning: A Deep Dive into Revolutionary TTS Technologies | by Anantharaman R | Medium; pristupljeno 24. studenog 2024.

[11] *Deepfake Technology Unmasking: The Rise of Synthetic Media and its Implications,* Boston Institute of Analytics (2024, travanj). Poveznica: Unmasking Deepfake Technolgy: Evolution Of Synthetic Media; pristupljeno 28. studenog 2024.

[12] Malhotra R., *What is Deepfake AI and Its Potential,* ValueCoders. Poveznica: Understanding Deepfake AI: Technology, Potential, and Ethics; pristupljeno 28. studenog 2024.

[13] *Retrieval-based Voice Conversion,* Wikipedia (2024, prosinac). Poveznica: Retrieval-based Voice Conversion - Wikipedia; pristupljeno 05. siječnja 2025.

[14] Kiran U., *MFCC Technique for Speech Recognition,* Analytics Vidhya (2023, kolovoz). Poveznica: MFCC Technique for Speech Recognition - Analytics Vidhya; pristupljeno 05. siječnja 2025.

[15] Sunil R., *Naive Bayes Classifier Explained With Practical Problems,* Analytics Vidhya (2025, siječanj). Poveznica: Naive Bayes Classifier in Machine Learning; pristupljeno 06. siječnja 2025.

[16] *Naive Bayes Classifiers,* GeeksforGeeks (2025, siječanj). Poveznica: Naive Bayes Classifiers - GeeksforGeeks; pristupljeno 06. siječnja 2025.

[17] *K-Nearest Neighbor(KNN) Algorithm,* GeeksforGeeks (2025, siječanj). Poveznica: K-Nearest Neighbor(KNN) Algorithm - GeeksforGeeks; pristupljeno 07. siječnja 2025.

[18] Yasar K., Tabsharani F., *What is a support vector machine (SVM)?,* TechTarget (2024, studeni). Poveznica: What is a Support Vector Machine (SVM)? | Definition from TechTarget; pristupljeno 10. siječnja 2025.

[19] *Logistic Regression in Machine Learning,* GeeksforGeeks (2024, lipanj). Poveznica: Logistic Regression in Machine Learning - GeeksforGeeks; pristupljeno 12. siječnja 2025.

[20] *Random Forest Algorithm in Machine Learning,* GeeksforGeeks (2025, siječanj). Poveznica: Random Forest Algorithm in Machine Learning - GeeksforGeeks; pristupljeno 14. siječnja 2025.

[21] *Decision Tree,* GeeksforGeeks (2025, siječanj). Poveznica: Decision Tree - GeeksforGeeks; pristupljeno 14. siječnja 2025.

[22] *CatBoost in Machine Learning,* GeeksforGeeks (2024, srpanj). Poveznica: CatBoost in Machine Learning - GeeksforGeeks; pristupljeno 15. siječnja 2025.

[23] *ML | XGBoost (eXtreme Gradient Boosting),* GeeksforGeeks (2023, prosinac). Poveznica: ML | XGBoost (eXtreme Gradient Boosting) - GeeksforGeeks; pristupljeno 15. siječnja 2025.

[24] *What is LSTM – Long Short Term Memory?,* GeeksforGeeks (2024, lipanj). Poveznica: What is LSTM - Long Short Term Memory? - GeeksforGeeks; pristupljeno 15. siječnja 2025.

[25] *Convolutional Neural Network (CNN) in Machine Learning,* GeeksforGeeks (2024, ožujak). Poveznica: Convolutional Neural Network (CNN) in Machine Learning - GeeksforGeeks; pristupljeno 17. siječnja 2025.

[26] Malinverno M., *What is a Spectrogram?,* Splice (2024, kolovoz). Poveznica: https://splice.com/blog/what-is-a-spectrogram/; pristupljeno 17. siječnja 2025.

[27] Naitali A., Ridouani M., Salahdine F., Kaabouch N., *Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions,* MDPI (2023, prosinac). Poveznica: Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions; pristupljeno 20. siječnja 2025.

[28] Singal A., *Hearing is Believing: Revolutionizing AI with Audio Classification via Computer Vision*, Hugging Face (2023, listopad). Poveznica: https://huggingface.co/blog/Andyrasika/voice-with-vision; pristupljeno 20. siječnja 2025.

[29] Stryker C., *What is NLP?,* IBM (2024, kolovoz). Poveznica: https://www.ibm.com/think/topics/natural-language-processing; pristupljeno 20. siječnja 2025.

Sažetak

Naslov:

Razvoj sustava za prepoznavanje i identifikaciju sintetički generiranih glasova korištenjem tehnika strojnog učenja

Sažetak:

Ovo istraživanje fokusira se na razvoj učinkovitog sustava za detekciju sintetički generiranih audiozapisa. Glavni cilj bio je razviti i evaluirati različite modele za prepoznavanje manipuliranih audiozapisa, kako bi se mogli koristiti za zaštitu privatnosti, prevenciju manipulacija i zloupotreba. U istraživanju provedeno je nekoliko koraka obrade podataka, uključujući segmentaciju, augmentaciju, ekstrakciju značajki i organizaciju podataka za treniranje, kako bi se podaci pripremili za različite tipove modela strojnog učenja.

Upotrijebljeni su različiti klasifikatori, među kojima naivni Bayes, stroj potpornih vektora, K-najbližih susjeda, logistička regresija, slučajne šume, stabla odluke, CatBoost klasifikator, ekstremno povećanje gradijenta te neuronske mreže: LSTM i CNN. Rezultati pokazuju da je stroj potpornih vektora postigao najbolje rezultate u detekciji, ostvarivši 100%-tnu točnost, preciznost, odziv i F1-mjeru. Drugi testirani modeli su također dali dobre rezultate, njihova točnost bila je nešto niža u odnosu na SVM.

Istraživanje također obuhvaća analizu popularnih metoda generiranja sintetičkih audiozapisa, ističe potrebu za razvojem sustava za njihovu detekciju, analizira tehničke izazove u izradi takvih sustava te istražuje interdisciplinarne pristupe koji mogu doprinijeti njihovoj učinkovitosti.

Zaključno, detekcija sintetički generiranih audiozapisa postaje ključna u zaštiti društva od prijetnji koje ova tehnologija može donijeti. Daljnja istraživanja trebaju se usmjeriti na razvoj naprednih sustava koji će pratiti razvoj alata za generiranje manipuliranih audiozapisa i uspješno ih detektirati. Kako bi se poboljšala detekcija manipuliranih audiozapisa trebaju se koristiti integrirani pristupi sustava za detekciju s naprednim tehnologijama, kao što su obrada prirodnog jezika i računalni vid.

Ključne riječi: deepfake, detekcija sintetički generiranih audiozapisa, MFCC, spektrogrami, SVM, klasifikator, neuronska mreža, manipulirani audiozapis, točnost

Summary

Title:

Development of systems for recognition and identification of synthetically generated voices using machine learning techniques

Summary:

This research focuses on the development of an effective system for detecting synthetically generated audio recordings. The main objective was to develop and evaluate different models for recognizing manipulated audio recordings, so that they could be used to protect privacy, prevent manipulation and abuse. Several data processing steps were carried out in the research, including segmentation, augmentation, feature extraction, and training data organization, to prepare the data for different types of machine learning models.

Various classifiers were used, including Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Random Forest, Decision Tree, CatBoost classifier, eXtreme Gradient Boosting, and neural networks: LSTM and CNN. The results show that the support vector machine achieved the best results in detection, achieving 100% accuracy, precision, response and F1-measure. Other tested models also gave good results, their accuracy was slightly lower compared to SVM.

The research also includes an analysis of popular methods of generating synthetic audio recordings, emphasizes the need to develop systems for their detection, analyzes technical challenges in the creation of such systems and explores interdisciplinary approaches that can contribute to their efficiency.

In conclusion, the detection of synthetically generated audio recordings becomes crucial in protecting society from the threats that this technology can bring. Further research should focus on the development of advanced systems that will accompany the development of tools for generating manipulated audio recordings and successfully detect them. To improve the detection of manipulated audio, integrated detection system approaches with advanced technologies, such as natural language processing and computer vision, should be used.

Keywords: deepfake, detection of synthetically generated audio recordings, MFCC, spectrograms, SVM, classifier, neural network, manipulated audio, accuracy

Skraćenice

SVM	*Support Vector Machine*	stroj potpornih vektora
KNN	*K-Nearest Neighbors*	k-najbližih susjeda
XGBoost	*eXtreme Gradient Boosting*	ekstremno povećanje gradijenta
LSTM	*Long Short Term Memory*	dugoročno-kratkoročna memorija
CNN	*Convolutional Neural Network*	konvolucijska neuronska mreža
RNN	*Recurrent Neural Network*	povratna neuronska mreža
GPU	*Graphics Processing Unit*	grafički procesor
TTS	*Text-to-Speech*	tekst u govor
VC	*Voice Conversion*	pretvorba glasa
RVC	*Retrieval Voice Conversion*	konverzija glasa na temelju dohvaćanja
MFCC	*Mel-Frequency Cepstral Coefficients*	mel-frekvencijski kepstralni koeficijenti
GNB	*Gaussian Naive Bayes*	Gaussov Bayesov klasifikator
SVC	*Support Vector Classifier*	klasifikator potpornih vektora
NLP	*Natural Language Processing*	obrada prirodnog jezika
GTCC	*Gammatone Cepstral Coefficients*	Gammatone kepstralni koeficijenti
GFCC	*Graph Frequency Cepstral Coefficinets*	koeficijenti grafičkog frekvencijskog kepstra
EER	*Equal Error Rate*	stopa jednakih grešaka
t-DCF	*Detection Cost Function*	funkcija troška detekcije
MLP	*Multilayer Perceptron*	višeslojni perceptron

Privitak

Upute za pokretanje .ipynb bilježnice u Google Colabu:

- Otvaranje Google Colaba:
    1. Prvo posjetite Google Colab na poveznici: Welcome To Colab - Colab
    2. Kliknite na gumb "File" (Datoteka) u gornjem lijevom kutu.
- Učitavanje bilježnice:
    1. Odaberite opciju "Upload Notebook" (Učitaj bilježnicu).
    2. Iz predane .zip mape odaberite predanu .ipynb datoteku.
- Pokretanje bilježnice:
    1. Kada se datoteka otvori, kliknite na gumb "Runtime" (Izvršavanje) u gornjem izborniku.
    2. Zatim odaberite opciju "Run all" (Pokreni sve).
- Učitavanje Kaggle.json datoteke:
    1. U ćeliji koja traži učitavanje datoteke, učitajte predani Kaggle.json.