International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Deep Learning Based Sentiment Analysis at the Sentence Level for Afaan Oromoo Text from Social Media

Lelisa Bushu Teso[1], T.Gopi Krishna[2], Mohamed Abdeldaiem Mahboub[3]

[1,2]Adama Science and Technology University, Department of Computer Science and Engineering, Adama, Ethiopia

[3]Department of Information Systems, Faculty of Information Technology University of Tripoli, Libya

[1]lbteso73@gmail.com, [2]gktiruveedula@gmail.com, [3]mabdeldaiem2009@gmail.com

Abstract. This study aims to implement deep learning models for sentence-level sentiment analysis in Afaan Oromoo text, investigating the impact of integrating Emojis into the labeling dataset and model performance. As a crucial aspect of natural language processing (NLP), sentiment analysis involves classifying emotions into positive and negative sentiments. The unstructured nature of data obtained from social media, comprising comments and feedback, posed challenges for real-time customer sentiment analysis. To overcome these obstacles, diverse techniques were applied from data collection to model development. Data, sourced from various Facebook pages and YouTube channels, underwent preprocessing with NLP techniques, forming a dataset of Afaan Oromo text enriched with Emojis. CNN, LSTM, and BiLSTM models were constructed with word2Vec feature extractions. Experimental results revealed a decline in performance—LSTM, BiLSTM, and CNN models dropped from 73.34% to 72.03%, 72.88% to 71.88%, and 74.58% to 72.66%, respectively, when Emojis were incorporated. Evaluation on a binary dataset demonstrated accuracies of 89.60% (LSTM), 88.32% (BiLSTM), and 87.52% (CNN) with skip-gram, and 87.68% (LSTM), 87.36% (BiLSTM), and 88.64% (CNN) with CBOW. The CNN model exhibited superior performance with 74.58% accuracy on multi-class datasets using skip-gram, leading to its selection for sentence-level sentiment analysis in Afaan Oromo text

Keywords: sentiment analysis, sentence level, deep learning, CNN, LSTM, BiLSTM, Afaan Oromoo, NLP, classification.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

I Introduction

The proliferation of social media platforms like Facebook and YouTube has empowered billions of users to share their opinions and feelings online, generating a vast digital library of sentiment. This data, encompassing text, emojis, and even audio/video, offers valuable insights across diverse fields - business leaders can gauge customer satisfaction, researchers can track public attitudes towards social issues, and governments can monitor public opinion on policies. However, effectively analyzing this treasure trove of information requires sophisticated tools like sentiment analysis, which categorize opinions as positive, negative, or neutral, unlocking the potential hidden within. Sentiment analysis, as explored by Bhatia, Sharma, and Bhatia (2018) [2], is a domain within natural language processing (NLP) and data mining that investigates the sentiments, responses, and judgments associated with various products and services. Businesses commonly leverage sentiment analysis to categorize sentiments within social media data, evaluate brand reputation, and gain deeper insights into customer preferences. This analytical approach extends beyond textual content to include opinions expressed through emojis, audio, and videos. When analyzing text, sentiment analysis evaluates polarity (positive, negative, or neutral) along with distinct emotions (such as anger, happiness, sadness), urgency (urgent or not urgent), and intentions (interested or not interested). The fundamental concept of sentiment analysis, as outlined by Dave et al (2003) [3], revolves around classifying the polarity of comments and reactions through the elimination of features and components mentioned in each text. The categorization of sentiment analysis can vary based on the interpretation of customer feedback, and common types encompass graded sentiment analysis, emotion detection, and multilingual sentiment analysis.

Graded sentiment analysis proves valuable when polarity correctness holds significance for the selected business. In this method, polarity is broadened to include classes like very positive, positive, neutral, negative, and very negative. Emotion detection within sentiment analysis surpasses mere polarities, identifying emotions such as happiness, frustration, anger, and sadness. Numerous emotion detection algorithms depend on lexicons or sophisticated machine learning algorithms, while multilingual sentiment analysis requires extra resources and preprocessing.

Facilitating individuals in expressing their thoughts and emotions is a critical function of sentiment analysis. It has rapidly become an indispensable tool for monitoring and understanding sentiment across various data types. Analyzing client feedback on social media facilitates product adaptation to customer needs, addressing satisfaction and dissatisfaction.

Challenges arise in sentiment analysis when attempting to categorize people's feelings into polarity rates, including issues related to different writing styles, the context-dependent nature of words, and the potential for sarcastic sentences. Sentiment analysis is approached from different perspectives, commonly at the document, sentence, and aspect levels.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

In document-level sentiment analysis, the overall opinion of a sentence is computed, operating under the assumption that each document yields a singular sentiment result Fikre, (2020) [5]. Sentence-level sentiment analysis, on the other hand, focuses on what people like or dislike about each entity feature in words or sentences Duwairi, (2015) [4]. This analytical level identifies the pertinent aspects of the mentioned entity and categorizes the sentiment of the associated opinion. Polarity is assigned to each sentence, treating individual sentences as distinct units, thus accommodating diverse opinions within the document. The primary task of sentiment analysis at the sentence level is subjectivity classification Fikre, (2020), where subjective sentences hold a polarity of information, either positive or negative, while objective sentences convey factual information with neutral sentiment or no opinion Duwairi, (2015).

While extensive research has been conducted on sentiment analysis in rich resource languages like English, Arabic, and other European languages, Afaan Oromo sentiment analysis has received limited attention. Few studies have explored Afaan Oromo sentiment analysis using machine learning approaches Kuyu (2021) [7], Rase (2020) [10], Wayessa et al, (2020) [17], Oljira et al, (2020) [9]. The reasons for this limited attention may include morphological complexity, resource scarcity, and lack of labeled data availability, limited research resources, absence of abbreviation rules, and other factors. Several challenges persist in Afaan Oromo language sentiment analysis that remains unaddressed. Key issues include the absence of a standardized dataset for developing sentiment analysis across diverse domains such as politics, economics, entertainment, sports, hotel services, and transportation. Some studies have focused solely on political data for sentiment analysis, limiting the breadth of application. Given the broad scope of sentiment analysis, adopting diverse solutions for these challenges is not advisable due to the need for multiple tools and models tailored to specific contexts.

In response to identified research gaps, this study undertakes sentiment analysis of the Afaan Oromo language. In addressing these challenges, the research employs deep learning models, specifically LSTM, BiLSTM, and CNN, Yao et al. (2016) to conduct sentence-level sentiment analysis based on Facebook and YouTube social media platforms. The study also evaluates the impact of integrating textual data with emojis on labeling data and model performance. To achieve its objectives, the study employs a range of techniques, from identifying data sources to developing models tailored to the investigated problems.

Facebook and YouTube were selected as social media platforms, guided by a study of the prevailing social media usage patterns in Ethiopia during the preceding twelve months, which indicated significant user engagement. The data collection process involved extracting information from several Facebook public pages, such as VOA Afaan Oromo, FBC Afaan Oromo, OBN, and BBC Afaan Oromo, as well as the YouTube channels of Ibrahim Bedane (2015) [6], Vision Entertainment, and Aanaa Entertainment. The collected data underwent preparation and preprocessing using natural language processing (NLP) techniques. The study proposes LSTM, BiLSTM, and CNN Yao et al. (2016) neural models Lipton et.al (2015) [8] with Word2Vec feature extraction. Ultimately, the study classifies sentiment analysis into negative, very negative, neutral, positive, and very positive classes.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

II Literature Review

The emergence of social media has given rise to an extensive sea of user-generated data, abundant in opinions and emotions. Analyzing this data, particularly in under-resourced languages like Afaan Oromoo, offers valuable insights into public sentiment and online conversations. This literature review examines the current state of sentence-level sentiment analysis for Afaan Oromoo text using deep learning approaches, identifying gaps and paving the way for future research. In the existing approach, while sentiment analysis has been extensively explored for dominant languages like English, Afaan Oromoo Ibrahim Bedane, (2015) [6] still faces research limitations. Several studies have attempted Afaan Oromoo sentiment analysis using machine learning techniques like Support Vector Machines and Random Forests. These studies mostly focused on document-level analysis, achieving moderate accuracy. Deep learning for Afaan Oromoo sentiment analysis, recognizing the limitations of traditional methods, recent research has embraced deep learning models for Afaan Oromoo sentiment analysis. A study by Kuyu (2021) [7] employed LSTMs trained on Facebook comments, achieving promising results for positive, negative, and neutral sentiment classification. Similarly, Rase (2020) [10] utilized CNNs Yao et al. (2016) for news articles, demonstrating the potential of deep learning for Afaan Oromoo Ibrahim Bedane (2015) [6] sentiment analysis. Existing research primarily focuses on document-level sentiment analysis, neglecting the nuances of individual sentences within a document. Sentences often express diverse and sometimes conflicting emotions, which are lost in document-level analysis. Therefore, shifting the focus to sentence-level analysis becomes crucial for understanding the intricate nature of online opinions in Afaan Oromoo.

Despite the promising initial results, deep learning-based Afaan Oromoo sentiment analysis faces several challenges. Lack of large-scale labeled datasets, morphological complexity of the language, and limited research resources pose significant hurdles. However, these challenges also present exciting opportunities for developing innovative solutions. Exploring multilingual transfer learning techniques,
leveraging active learning for efficient dataset labeling, and incorporating Afaan Oromoo-specific linguistic features within models can be promising avenues for future research. This literature review highlights the limited research on sentence-level sentiment analysis of Afaan Oromoo Ibrahim Bedane (2015) [6], text using deep learning models. My proposed study aims to address this gap by employing LSTMs, BiLSTMs, and CNNs Yao et al. (2016) for sentence-level sentiment classification on Afaan Oromoo data collected from Facebook and YouTube. We will also investigate the impact of incorporating emojis into the model. We believe this study can contribute significantly to the development of robust and accurate Afaan Oromoo sentiment analysis tools, fostering better understanding of online communication in this under-resourced language

2.1 Stages of Sentiment Analysis

Analyzing sentiment entails determining polarity for the target object across different levels. The key levels encompass document level, sentence level, and aspect-based analysis. The subsequent section provides an overview.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Sentence-Level Sentiment Analysis

Sentimental analysis at the sentence level works with the classification of subjectivity which classifies the sentence given as a subjective or objective sentence (Fikre, 2020). A subjective sentence expresses personal feelings, views, emotions, or beliefs which carry a category of positive or negative polarity, whereas an objective sentence presents some factual information.

Document-Level Sentiment Analysis

This method retrieves sentiment from the entire review and categorizes the entire opinion based on the overall sentiment of the opinion holder. When a document is authored by a single individual and conveys sentiment about a specific entity, document-level classification is most effective. It categorizes the entire sentence as positive, negative, or neutral.

Aspect-Oriented Sentiment Analysis

Aspect-Oriented Sentiment Analysis (AOSA) is a text analysis approach that dissects data into aspects and assesses the sentiment linked to each aspect Bethiha, (2021). Customer opinions can be scrutinized through aspect-based sentiment analysis, connecting distinct emotions with various features of a product or service. This form of sentiment analysis can extract sentiments and aspects separately. Sentiments revolve around positive or negative feelings regarding a particular subject, while aspects delineate a category, characteristic, or topic under discussion.

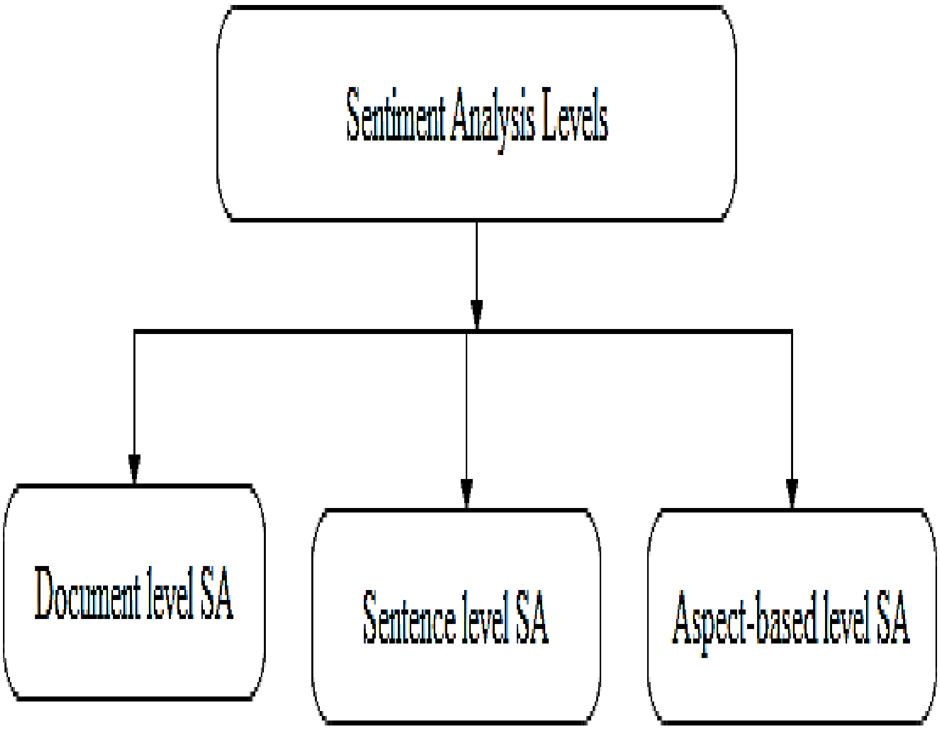


Fig. 1. Level of Sentiment Analysis (Shelar et.al, (2018)) [12]

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

We have delved into sentiment analysis using varied methodologies, encompassing Lexicon (rule-based), machine learning-based, and hybrid approaches (integrating machine learning and rule-based techniques). A more recent technique in sentiment analysis is the deep learning approach, demonstrating a more profound comprehension of data compared to the conventional machine learning approach. Nevertheless, the deep learning approach necessitates substantial volumes of high-quality data for attaining optimal model performance. The following section provides an outline of each approach. Lexicon-based sentiment analysis utilizes a sentiment dictionary containing sentiment words to classify sentiment. This technique relies on a set of facts or data sources along with rules to manipulate the data. It is a rule-based approach to text analysis that doesn't involve training machine learning models. The outcome is a set of rules for labeling text as positive, negative, or neutral. Also known as the lexicon-based approach, notable implementations include SentiWordNet, TextBlob, and VADER, Tesfaye, (2011) [13].

In contrast to lexicon-based sentiment analysis, machine learning employs models to assess sentence sentiment. Several investigations have delved into sentiment analysis employing machine learning algorithms, such as Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), Maximum Entropy (ME), Random Forest, among others Sachdeva et al. (2019) [11].

A hybrid approach system works with the collaboration of machine learning and Lexicon-based techniques to deliver a good result. The result acquired from this method is always more precise.

The paper Vaitheeswaran et al. (2016) [15] has shown that the combination provides improved performance in sentiment classification. As this study stated, sentiment lexicons are used as resources and detect sentiment polarities, and the results from the lexicon-based are used to train machine learning algorithms Vaitheeswaran et al. (2016) [15].

III Related Work

Numerous scholars have undertaken investigations pertaining to sentiment analysis across various languages. Among these studies, some specifically focused on conducting research in the realm of Afan Oromo sentiment analysis, aiming to forecast the emotional polarity of individuals. In the following discussion, we aim to outline and explore the efforts dedicated to Afan Oromo sentiment analysis as well as those involving other languages.

Each research approached sentiment analysis from distinct perspectives, including document-level, sentence-level, and aspect-based levels. Researchers exclusively concentrated on textual content, neglecting the consideration of emojis for predicting polarity. The influence of emojis within the text on labeling datasets and performance measures was not explored in the works of Wayessa and Abas (2020) [16], Kuyu (2021) [7], and Oljira et al. (2020) [9]. A brief overview of the reviewed papers is provided in Table 1 below.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Table 1. Summary of Related Works

Authors	Paper Title	Feature, Method, Accuracy	Gaps
Negessa,(2020), Wayessa et.al [17]	Sentence-level sentiment analysis from Afan Oromo text based on supervised Machine Learning	TF-IDF; 90% of SVM, 89% of RF	did not consider Emojis, consider only political data, this study was done usingML
Megersa Oljira et al. (2020) [9]	Analyzing Sentiment in Afan Oromo using Bidirectional Long Short-Term Memory and Convolutional Neural Network (CNN)	Word embedding; 93.3% CNN, 91.4%, Bi-LSTM 94.1% CNN-Bi-LSTM	Emojis did not considerable, did not check to overfit, Consider only political data
Wondwosen Mulugeta (2014) [17]	Machine Learning Reads the Mood of Amharic Online Posts	N-grams,43.6% of Unigram, 44.3% of a bigram, and 39.3% of trigram using NB	Low performance, workdone using traditional ML
Jamal and Abate, (2019) [1] Ibrahim Bedane (2015) [6]	Unsupervised Opinion Mining Approach for Afaan Oromo Sentiments	Lexical Databases, bigram performance of 70% recall, 82% precision, and 76% F-score	The technique followed is difficult for a large dataset, and used 600 reviews only
Megersa Oljira et al. (2020) [9]	Machine Learning-Based Sentiment Analysis of Afan Oromo (Document-level)	TF-IDF, 93% of MNB	Used binary class, did not consider emojis, worked on traditional ML algorithms, consider only political data

IV Methods

This investigation adopts an experimental research approach wherein the outcomes are derived from an experiment designed to forecast sentiment analysis. The research design encompasses key activities such as identifying the research area, conducting a literature review, formulating gaps, collecting and preprocessing data, applying feature extraction, constructing models, and predicting comment polarity. The procedures of the Study Design as shown in Fig. 2 below.

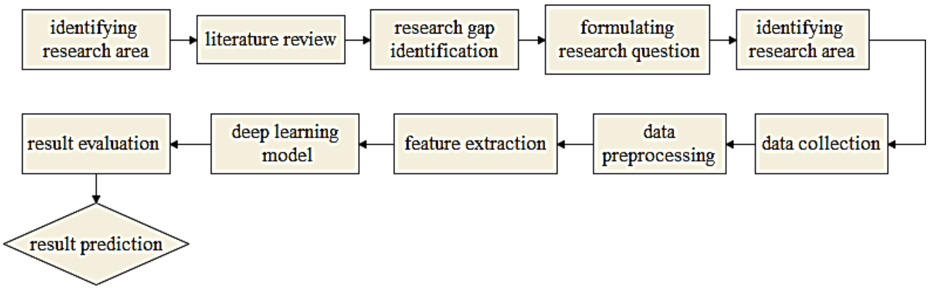


Fig. 2. Experimental Design

4.1 Data Collection

This study collected comments on Afaan Oromoo texts and emojis from Facebook public pages and YouTube platforms using the Face Pager software application and YouTube online comment scrapers respectively. The reason why we selected Facebook and YouTube was that they have huge followers, subscribers, and the availability of resources. The below chart in Fig.3 shows that the selected social media have a greater number of followers than the other social media for the past twelve months. In this research, comments on Afaan Oromoo texts and emojis were gathered from public pages on Facebook and YouTube platforms. The Face Pager software application and YouTube online comment scrapers were employed for data collection. The choice of Facebook and YouTube was driven by their substantial follower and subscriber bases, as well as the accessibility of resources. The chart below illustrates that these selected social media platforms have consistently maintained a larger number of followers compared to other platforms over the past twelve months as shown in Fig.3.

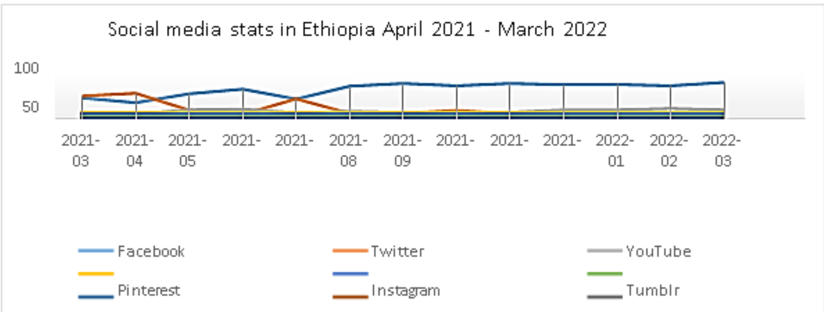


Fig. 3. Social media Users Stats in Ethiopia

This research chose specific public Facebook pages and YouTube channels, guided by the following selection criteria:

- Prominent Facebook public pages consistently publishing content in Afaan Oromo language
- Pages and channels with a substantial follower base.
- Pages featuring news related to politics, the economy, and entertainment.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Table 2. Total Followers on Chosen Facebook Pages and YouTube Channels

Titles of Facebook Pages and YouTube Channels	**Number of followers**
BBC News Afaan Oromoo	1,176,953
OBN Afaan Oromoo	1,126,632
FBC Afaan Oromoo	796K
Vision Entertainment	459K
Aanaa Entertainment	315,109

In the above Table. 2, Political, economic, and entertainment data were collected from Facebook pages and YouTube channels between March 20, 2022, and May 2, 2022. Public pages like BBC Afaan Oromoo and popular YouTube channels like Aanaa Entertainment were chosen for diverse perspectives. The dataset totaled 7404 comments, split across politics (3381), economy (2000), and entertainment (2023).

4.2 Dataset Preparation

After gathering, the data was then prepped: filtered, cleaned, missing values dealt with, and finally combined into one dataset. Excel prepped the CSV files, and then specific criteria readied them for annotation.

- Disregard all non-Afaan Oromo texts, special characters, and numbers.
- Exclude missing values, whitespace, and blank entries from consideration.
- Ensure uniqueness for each text in the dataset and consolidate all data into a unified dataset.

4.3 Dataset Annotation

This study built two sentence-level sentiment datasets for comparison: labeled text only and text with emojis. 7,404 comments were annotated, considering emoji impact with the text-emoji dataset. Annotators overlapped on 3,702 comments for agreement analysis, using five polarity classes (very positive to very negative). Results are detailed in Table 3.

Table 3. Unique annotated dataset

Annotators	**Very positive**	**positive**	**neutral**	**negative**	**Very negative**	**Total dataset**
Annotator 1	338	1099	578	1066	621	3,702
Annotator 2	334	1107	574	1053	634	3,702

Table.4 illustrates the shared annotated dataset used by both annotators, facilitating the assessment of agreement between them.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Table 4. Common annotated dataset

Annotators	**Very positive**	**positive**	**neutral**	**negative**	**very negative**	**Total dataset**
Annotator 1	39	127	109	110	115	500
Annotator 2	42	130	102	117	109	500

The shared dataset was identical for both annotators, enabling the evaluation of inter-annotator agreement as shown in Table.5.

Table 5. Total annotated text only and text with Emoji dataset

Polarity	Total comments	text only dataset	text with Emoji dataset
Very negative	1242	995	247
negative	2132	1925	207
neutral	1156	1004	152
positive	2198	1898	300
Very positive	676	582	94
Total	7404	6404	1000

4.4 Inter Annotator Agreement

This study measured annotation quality using Cohen's kappa (0.685), indicating substantial agreement between annotators on 500 shared comments. While discrepancies remain, overall labeling quality is good.

Table 6. Sample of disagreement of the annotator

Comments	**Annotator 1**	**Annotator 2**
Duulli birmadummaa biyyaa eegsisuu karaa milkaahina qabuunni raawwatamaa jedhame	neutral	Positive
Dhugaan ni tura malee hinuma baati yoo Waaqayyo jedhe.	neutral	Positive
Jabaadhuu Gadi guuri iccitii gantoota Sabaaf osoo taane warragaraaf jiraattu ergamtoota nafxanyaa.	Very negative	Negative
Jaalala na qabsiifte!	positive	Very positive

4.5 Architecture of the Proposed Model

This study encompasses the entire process from identifying data sources to data annotation, preprocessing, feature extraction, and utilizing deep learning models. Various methods and components enhance the effectiveness and efficiency of the model. Fig. 4 illustrates the overall system architecture for the automatic sentiment analysis of Afaan Oromoo text.

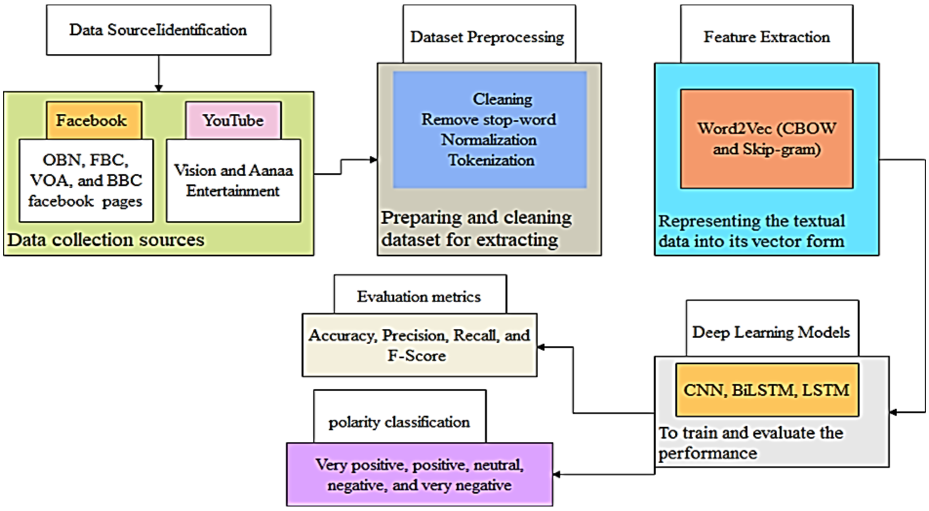


Fig. 4. Architecture of the proposed model

Sentiment analysis was performed using deep learning models, including LSTM, BiLSTM, and CNN due to their accuracy and ability to handle textual data with emojis. LSTM excels at understanding context with its multiple memories, while CNN handles emojis well. BiLSTM uses two LSTMs for even better sequence classification as shown in Fig.5.

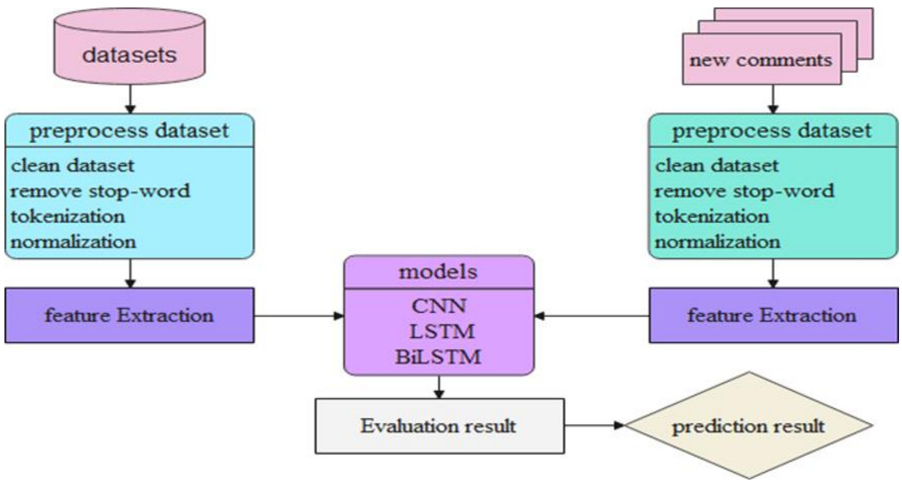


Fig. 5. Model Building of Afaan Oromo text sentiment analysis

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

4.6 Proposed Deep Learning Model

Based on the criteria, this study suggests employing deep learning models like CNN, LSTM, and BiLSTM. The choice of multiple models is deliberate, considering each model's unique strengths and weaknesses. For example, the LSTM model faces challenges in learning sentence order without access to future data. Fig. 6 below illustrates the testing of these models with new data and their evaluation using performance metrics to determine the most suitable model for Afaan Oromoo sentiment analysis.

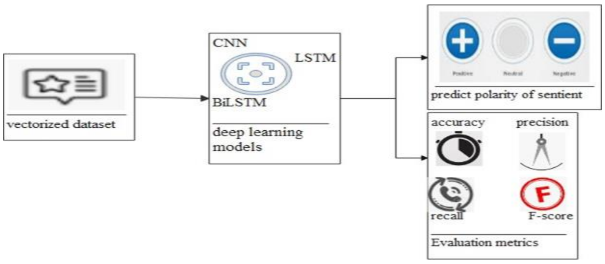


Fig. 6. Proposed deep learning models for Afaan Oromo SA

4.7 Implementation of Proposed Data Preprocessing

Loading Dataset

The Python pandas library employed the pd.read_csv () method to load the dataset in CSV format. The following sample code demonstrates loading the dataset in panda's format.

```
df=pd.read_csv('input.csv', header = 0, encoding='unicode_escape')
```

Cleaning the Dataset

Data cleaning is accomplished by eliminating various things like punctuation, digits, white spaces, and special characters. The cleaning dataset was displayed in the sample codes below.

```
def clean_text_round(text):
    text=re.sub('\[.*?\]', '',text)
    text = re.sub(r'\s+', ' ', text).strip()
    text = re.sub('([@0-9_]+)|[^\w\s]|#|http\S+', '', text)
    return text
round1= lambda x:clean_text_round(x)
df1=pd.DataFrame(df.reviews.apply(round1))
```

Normalization

Normalization is important to reduce the number of unique tokens in a text, get rid of variations, and clean up the text by getting rid of extraneous details. The following provides evidence for the sample codes used to normalize the dataset.

```
def clean(text, stem_words=True):
text = re.sub(" mini", " miti ", text, flags=re.IGNORECASE)
text = re.sub(" baay'ee ", " baayyee ", text, flags=re.IGNORECASE)
text = re.sub(" ta'e ", " tahe ", text, flags=re.IGNORECASE)
text = re.sub(" ba'e", "bahe", text, flags=re.IGNORECASE)
text = re.sub(" ka'e", " kahe ", text)
text = re.sub("yuunbarsiitii", " yuunvarsiitii ", text, flags=re.IGNORECASE)
text = re.sub("umrii", " umurii ", text, flags=re.IGNORECASE)
text = re.sub("bilxiginnaa", " badhaadhina ", text, flags=re.IGNORECASE)
data['reviews'] = data['reviews'].apply(clean)
```

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Tokenization

Tokenization is used to break up phrases into words known as tokens once the dataset has been cleaned, normalized, and stop words have been eliminated.

Removing stop words

Stop words are unnecessary in sentiment sentences and their removal helps to minimize the dimensionality. Below is a sample of code for deleting stop words and tokenization words.

```
def clean_text_round(text):
    text=re.sub('\[.*?\]', '',text)
    text = re.sub(r'\s+', ' ', text).strip()
    text = re.sub('([@0-9_]+)|[^\w\s]|#|http\S+', '', text)
    return text
round1= lambda x:clean_text_round(x)
df1=pd.DataFrame(df.reviews.apply(round1))
```

4.8 Implementation of proposed Word Embedding

Word Embedding aids in transforming textual data into a vector format comprehensible to models. The code snippet below illustrates a sample of word2Vec feature extraction.

```
import gensim
from gensim.models import Word2Vec
from gensim.models import FastText
from gensim.models import KeyedVectors
Min_count=2, Size=100, Workers=4, Window=5
model=gensim.models.Word2Vec(sentences=stopw,vector_size=Size,sg=0, workers=Workers,min_count=Min_count, window=Window)
model.build_vocab(stopw, progress_per=1000)
words=list(model.wv.key_to_index)
```

4.9 Proposed Model Implementation

A dataset represented in vector form was employed for model training and the development of deep learning methods. This study introduced various models, including CNN, BiLSTM, and LSTM.

LSTM Model Implementation

Two hidden layers were used in the LSTM model for sequential data. To produce the findings, the model was trained with 150 filters, 0.2 dropouts, activation relu, and dense 128. Here is a sample LSTM model program.

```
modells = Sequential()
modells.add(embedding_layer)
modells.add(LSTM(128,  return_sequences=True, activation = 'relu' ))
modells.add(Dropout(0.25))
modells.add(LSTM(64,  return_sequences=False, activation = 'relu'))
modells.add(Dropout(0.2))
modells.add(Dense(6, Activation ='softmax'))
```

BiLSTM Model Implementation

BiLSTM utilized two LSTM hidden layers for both directions, enhancing its ability to comprehend the context of both past and future sentences. Below is a snippet of the BiLSTM model code.

```
modells = Sequential()
modells.add(embedding_layer)
modells.add(Bidirectional(LSTM(128,  return_sequences=True, activation = 'relu' ))
modells.add(Dropout(0.25))
modells.add(Bidirectional(LSTM(64,  return_sequences=False, activation = 'relu'))
modells.add(Dropout(0.2))
modells.add(Dense(6, Activation ='softmax'))
```

CNN Model Implementation

In constructing the CNN model for this study, layers with Conv1D convolutions using 64 filters, the Adam optimizer, a maximum dropout of 0.25, maximum pool size of 1, dense layer with 128 units, and the ReLU activation function were utilized. Below is an excerpt of the code for the CNN model.

```
modelcnn = Sequential()
modelcnn.add(embedding_layer)
modelcnn.add(Conv1D(64, 1, 1, input_shape=(8, 1, 1), activation='relu'))
modelcnn.add(Conv1D(32, 1, 1, activation='relu'))
modelcnn.add(Dense(6, activation='softmax'))
```

V RESULTS

This study employed various deep learning models in experiments to demonstrate its findings. The inclusion of emojis within text comments was also found to influence both model performance and dataset labeling. The subsequent studies elucidate the results of each model, the impact of increasing the number of classes, and the role of emojis in the analysis.

5.1 The effect using of Emojis on Labeling Afaan Oromo texts

This study conducted experiments to investigate the impact of emojis on labeling Afaan Oromo sentiment analysis. As previously discussed in the introduction, emojis are regarded as reliable indicators of sentiment, enhancing sentiment classification. The study concludes with an exploratory experiment illustrating labeled emojis alongside text as shown in Table.7 in the dataset and the number of comments labeled with each class.

Table 7. The effect using Emojis on dataset distribution

Comments	Number of comments
Comments are written with text only	6404
Comments are written with text and Emojis	1,000
Total	7, 404

In Table.7, out of the total 7,404 comments in the dataset, 14.3% included emojis. Many comments on selected Facebook pages and YouTube channels consisted solely of emojis. As our focus was on assessing the impact of emojis used with text, we did not exclusively consider comments with only emojis. The prevalence of emoji comments indicates a widespread use of emojis for expressing feelings.

The utilization of emojis for comment labeling influences sentiment polarity and the categorization of comments as positive, neutral, or negative. For instance, a statement like "Mootummaa naannoo Oromiyaa" (Oromia regional government) could be neutral,

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

but adding emojis alters its classification. Therefore, "Mootummaa naannoo Oromiyaa ❤❤" is classified as positive. Similarly, a sentence like "sirba haaraa Andualem Gosa gadhiifame" is neutral, but adding emojis as in "Sirba haaraa Andualem Gosa gadhiifame ❤🙏🙏" classifies it as positive. This example illustrates the influence of emojis in changing comment classifications.

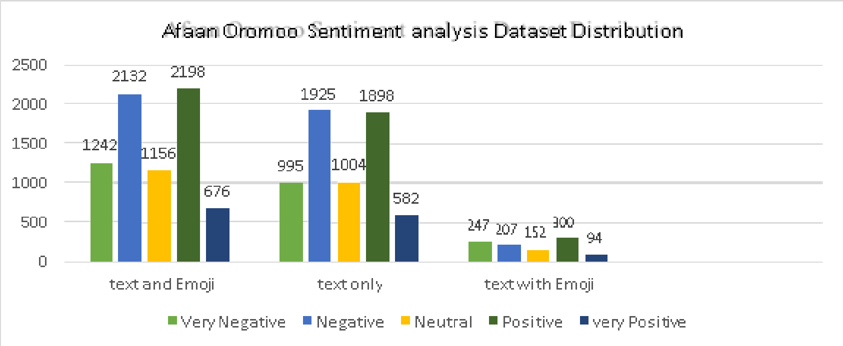


Fig. 7. Afaan Oromoo Sentiment Analysis Dataset Distribution

In Fig.7, the distribution of categorized comments, including very positive, positive, neutral, negative, and very negative, differed across text and emojis, text-only, and text with emojis. Out of 7,404 labeled comments, 6,404 were labeled as text-only, while 1,000 were labeled as text with emojis. This highlights the significant influence of emojis in the labeling of data.

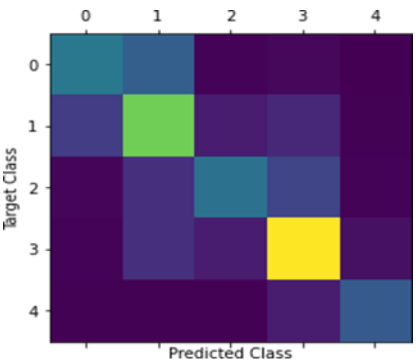
5.2 Result of LSTM Model

The construction of the LSTM model involved training hyperparameters on the prepared dataset within a neural network. The model demonstrated an accuracy of 73.73% and 70.31% on the text-only dataset and the text-with-emoji dataset, respectively, using CBOW. In skip-gram feature extraction, the model attained accuracies of 72.03% and 71.66% on the text-only dataset and text-with-emoji dataset, respectively. The accompanying figure displays the accuracy and loss trends throughout the training and validation of the LSTM model across the five multi-class categories, as depicted in Fig. 8.

Fig. 8. Accuracy and loss during training and validation of LSTM in a multi-class (five) scenario

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

As discussed in the aforementioned study on evaluation metrics, the proposed models were evaluated using metrics such as precision, recall, and f-score for the LSTM model across various classes. The results of this assessment for accuracy are illustrated in the Fig.9 below.



Classification Report

	precision	recall	f1-score
0	0.65	0.54	0.59
1	0.57	0.68	0.62
2	0.68	0.50	0.58
3	0.70	0.78	0.74
4	0.82	0.75	0.78
accuracy			0.66
macro avg	0.69	0.65	0.66
weighted avg	0.67	0.66	0.66

Fig. 9. Evaluation metrics of the LSTM model

5.3 Result of BiLSTM Model

BiLSTM, a neural network model that processes sequential data, deviates from the conventional LSTM by considering information from both forward and backward directions. Utilizing the BiLSTM model, this study attained accuracies of 72.88% and 71.88% on the text-only dataset and the text-with-emoji dataset, respectively, with skip-gram. With CBOW, the model achieved accuracies of 70.34% and 72.66% on the text-only dataset and text-with-emoji dataset. The Fig.10 below depicts the accuracy, training and validation loss of the BiLSTM model across the five multi-class categories

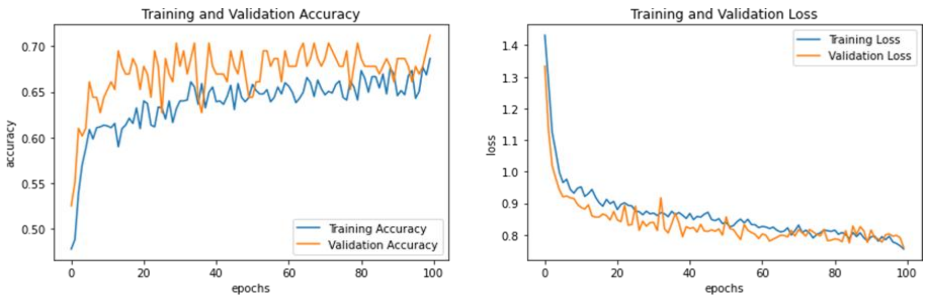
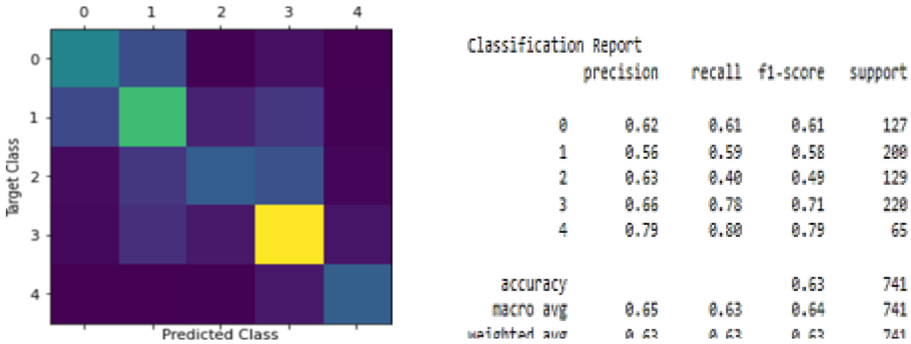


Fig. 10. Metrics for BiLSTM Model Performance: Accuracy and Loss in a Multi-Class Context

In Fig. 11 below, the results of evaluation metrics—precision, recall, and f-score—for the BiLSTM model across multiple classes are depicted.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752



Classification Report

	precision	recall	f1-score	support
0	0.62	0.61	0.61	127
1	0.56	0.59	0.58	200
2	0.63	0.40	0.49	129
3	0.66	0.78	0.71	220
4	0.79	0.80	0.79	65
accuracy			0.63	741
macro avg	0.65	0.63	0.64	741
weighted avg	0.63	0.63	0.63	741

Fig. 11. Result of the evaluation metrics of the BiLSTM model on multi-class

5.4 Result of CNN Model

The CNN model was constructed using optimal network hyperparameters trained on the prepared dataset. The model attained an accuracy of 74.58% on the text-only dataset and 71.09% on the text-with-emojis dataset with skip-gram. With CBOW, the accuracy reached 73.73% on the text-only dataset and 72.66% on the text-with-emojis dataset. Fig.12 below illustrates the accuracy, training, and validation loss of the CNN model across the five multi-class categories.

Fig. 12. Evaluation metric of the CNN model for multi-class (five)

The Fig.13 below displays the evaluation results of precision, recall, and f-score for the CNN model across multiple classes.

Classification Report

	precision	recall	f1-score	support
0	0.61	0.59	0.60	127
1	0.58	0.58	0.58	200
2	0.57	0.53	0.55	129
3	0.70	0.72	0.71	220
4	0.79	0.83	0.81	65
accuracy			0.64	741
macro avg	0.65	0.65	0.65	741
weighted avg	0.64	0.64	0.64	741

Fig. 13. Evaluation metrics of the CN

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

5.5 The effect of using Emojis on the Model Performance

One of the objectives of this study is to examine how incorporating emojis in text datasets affects the performance of deep learning models. The evaluation involved feeding distinct datasets, one with text only and another including emojis, into CNN, LSTM, and BiLSTM models to assess the influence of emojis. The results illustrating the impact of emojis on experiments are presented in Table.8 below.

Table 8. Comparison of deep learning model's performance with and without Emoji

Datasets	**LSTM**	**BiLSTM**	**CNN**
Text only	73.34	72.88	**74.58**
Text and Emoji	72.03	71.88	72.66

Results from Table.8 indicate that the accuracy of the CNN model decreases from 74.58% to 72.66%, the BiLSTM model drops from 72.88% to 71.88%, and the LSTM model declines from 73.34% to 72.03% when emojis are used for comment labeling with Afaan Oromo text. The experimental findings demonstrate that incorporating emojis in Afaan Oromo text reduces the performance of the CNN, BiLSTM, and LSTM models by 1.92%, 1.00%, and 1.31%, respectively. Consequently, the addition of emojis to Afaan Oromo text diminishes the models' effectiveness.

5.6 Comparison of the result of skip-gram and CBOW

The performance measurements of the proposed deep learning models were derived using word2vec's skip-gram and CBOW feature extraction techniques. The outcomes of the experiments are analyzed and provided in the following Table 9.

Table 9. Comparison result of a skip-gram and CBOW on the multi-class

Word2Vec	**LSTM**	**BiLSTM**	**CNN**
Skip-gram	72.03	72.88	**74.58**
CBOW	73.73	70.34	73.73

As indicated in Table.9, the experimental results reveal that skip-gram outperforms CBOW on the CNN and BiLSTM models, achieving accuracies of 74.58% and 72.88%, respectively. Conversely, CBOW surpasses skip-gram on the LSTM model with an accuracy of 73.73%.

5.7 The effects of increasing the number of classes on model performance

Another objective of this study was to evaluate the impact of increasing the number of classes on the model's performance. To investigate these effects, the study conducted experiments by reducing the number of classes from the original multi-class (five) to three and binary classes.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

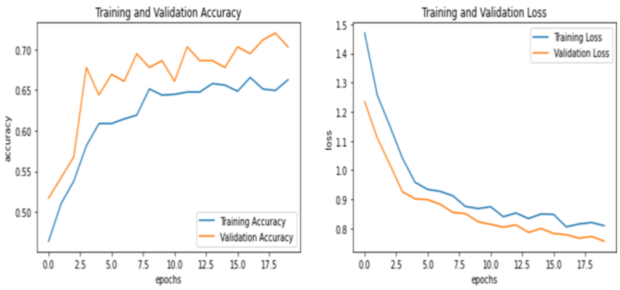


Fig. 14. Assessment results of the CNN model in a multi-class (five) setting

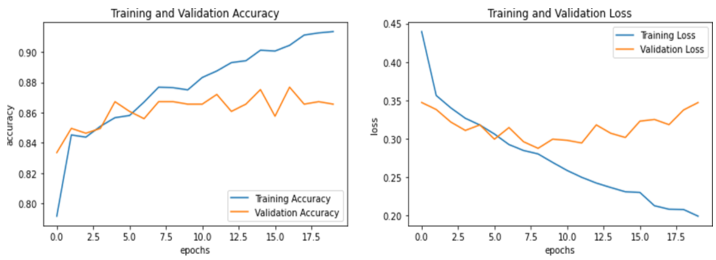
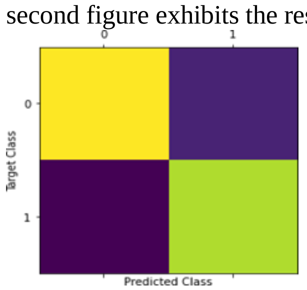


Fig. 15. Assessment outcomes for the CNN model in a binary class scenario

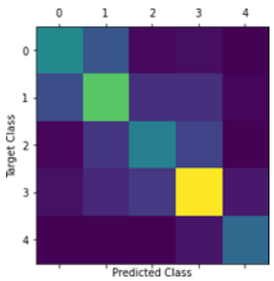
The influence of introducing additional classes is illustrated in Figures 14 and 15 above. The first figure displays the results of the multi-class experiment for CNN, while the second figure exhibits the results of the binary class experiment.



Classification Report

	precision	recall	f1-score	support
0	0.91	0.84	0.87	341
1	0.82	0.90	0.86	284
accuracy			0.87	625
macro avg	0.87	0.87	0.87	625
weighted avg	0.87	0.87	0.87	625

Fig. 16. The evaluation result of the CNN model on binary classes



	precision	recall	f1-score	support
0	0.61	0.59	0.60	127
1	0.58	0.58	0.58	200
2	0.57	0.53	0.55	129
3	0.70	0.72	0.71	220
4	0.79	0.83	0.81	65
accuracy			0.64	741
macro avg	0.65	0.65	0.65	741
weighted avg	0.64	0.64	0.64	741

Fig. 17. The assessment outcomes of the CNN model in the multi-class (five) scenario

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Figures 16 and 17 above display the evaluation results of the CNN model on binary and multi-class scenarios, highlighting the effects of increasing the number of classes.

Table 10. Result summary of using a different number of classes

Number of classes	**Skip-gram**			**CBOW**		
	LSTM	BiLSTM	CNN	LSTM	BiLSTM	CNN
Binary class	89.60	88.32	87.52	87.68	87.36	88.64
Three class	82.05	79.35	87.04	79.08	79.89	79.22
Multi-class (five)	72.03	72.88	74.58	73.73	70.34	73.73

Based on the findings in Table 10, the experimental results reveal a performance decline in the proposed model as the number of dataset classes increases. For instance, the accuracy of the BiLSTM model with skip-gram feature extraction on the same dataset was 88.32%, 79.35%, and 72.88% for binary, three classes, and multi-class (five), respectively. Comparatively, the binary class result increased by 15.44% in contrast to the multi-class (five) result. The detailed results are presented in Table 10 above. In summary, the rise in the number of classes adversely affected the performance of the proposed models.

5.8 Hyperparameter Tuning

In pursuit of optimal hyperparameters for the proposed neural network models, this study employs the grid search hyperparameter tuning method. The exploration involves searching various combinations of potential optimal hyperparameters, encompassing parameters such as activation, optimization, number of epochs, batch size, and others. The hyperparameters for the proposed models are detailed in Table 11 below.

Table 11. Hyperparameter for the proposed model

Models	**Filters**	**Hidden layers**	**Activation**	**Dropout**	**Optimizer**	**Epochs**	**Batch size**
LSTM	128	2	Relu	0.25	Adam	100	42
BiLSTM	128	2	Relu	0.25	Adam	50	42
CNN	64	2	Relu	0.25	Adam	50	64

VI Discussions

This study aimed to conduct sentiment analysis for Afaan Oromo text from Facebook and YouTube using deep learning approaches, specifically CNN, LSTM, and BiLSTM models with skip-gram and CBOW feature extraction. Comparisons were drawn with previous research efforts that focused on specific dataset areas, lacked consideration for emojis, and often categorized sentiment into binary or three classes. This study addressed these gaps by considering emoji effects, incorporating data from diverse areas (political, economic, and entertainment), and undertaking a multi-class sentiment classification. The evaluation encompassed the impact of emojis on labeling comments, the influence of increasing class numbers on model performance, the selection of optimal feature extraction, and the identification of superior models for sentence-level sentiment analysis in Afaan Oromo text. Research questions were formulated and addressed, with

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

the study providing comprehensive solutions. The discussion covers the findings and insights gained from the experimental investigations conducted throughout the research.

Feature extraction significantly enhance the performance of Afaan Oromo text sentence-level sentiment analysis with the proposed deep neural networks? This research inquiry assesses the efficacy of two word2Vec feature extraction methods (skip-gram and CBOW) in conjunction with the proposed neural networks. The experimental results indicate that skip-gram outperforms CBOW, achieving accuracy rates of 74.58% and 72.88% on the BiLSTM and CNN neural models, respectively. However, on the LSTM model, CBOW performs better than skip-gram, achieving an accuracy of 73.73%. Based on these findings, the researcher recommends using skip-gram feature extraction for the BiLSTM and CNN models in classifying sentiments for Afaan Oromo text at the sentence level, and CBOW for the LSTM model.

Deep learning models effectively create an improved classification model for sentence-level Afaan Oromo sentiment analysis? This research question assesses the performance of proposed neural networks on a prepared Afaan Oromo text-based sentiment analysis dataset. The experiment results reveal that the CNN model outperforms the other two models, achieving a remarkable accuracy rate of 74.58%. Consequently, this study recommends the use of CNN neural models for sentiment classification in Afaan Oromo text at the sentence level.

This research investigated how the inclusion of emojis affects the performance of the proposed models when applied to Afaan Oromo texts. According to the experiment, the CNN model's accuracy drops from 74.58% to 72.66%, the BiLSTM model from 72.88% to 71.88%, and the LSTM model from 73.34% to 72.03% when utilizing a dataset containing text with emojis. The findings indicate that the addition of emojis to Afaan Oromo textual data results in a reduction in performance for the CNN, BiLSTM, and LSTM models by 1.92%, 1.00%, and 1.31%, respectively. Consequently, based on the experimental results, this study concludes that the incorporation of emojis diminishes the overall performance of the models applied to Afaan Oromo textual data.

VII Conclusion

Sentiment analysis, a subset of NLP, assesses people's feelings, categorizing them as positive, neutral, or negative across various mediums. Previous studies in Afaan Oromo sentiment analysis predominantly utilized traditional machine learning, with only one document-level study employing deep learning for binary classification. There's a notable gap in research covering different data categories, emoji impact on datasets and model performance in Afaan Oromo text, and the influence of class size on model performance. This research aims to conduct sentence-level sentiment analysis for Afaan Oromo text using deep learning models, specifically CNN, LSTM, and BiLSTM. Data was gathered from public Facebook pages and YouTube channels, resulting in a preprocessed dataset of 7,404 comments. Three proposed models, incorporating word2Vec feature extraction, were implemented with hyperparameter tuning and dropout regularization to prevent overfitting. The inclusion of emojis alters dataset classifications, shifting from neutral to positive or negative when used with Afaan Oromo texts. The experiment reveals that using emojis alongside Afaan Oromo text decreases the LSTM, BiLSTM, and CNN models' performance by 1.31%, 1.00%, and 1.92%, respectively, compared to the dataset containing only Afaan Oromo text. In conclusion, the use of emojis with Afaan Oromo text negatively impacts the models' performance, as evidenced by the experimental findings. The experiment indicates that the model's performance is influenced by class size. Different datasets for binary, three and five classes were created, revealing a performance decline with an increasing number of classes. Effectiveness evaluation using two word2Vec types, skip-gram and CBOW, favored the CNN model over the others. Consequently, the study opted for the CNN model for Afaan Oromo text-based sentence-level sentiment analysis. The study's contributions include dataset preparation with 7,404 instances, examining the impact of emojis on labeling and model performance in Afaan Oromo texts, assessing the effect of dataset class size on model performance in multiclass sentiment classification, and ultimately, developing Afaan Oromo text sentiment analysis using CNN, BiLSTM, and LSTM models. This study focused on text-based sentiment analysis, with plans for future expansion into audio and video data for sentiment classification. Emojis were examined in this study to understand their impact on labeling datasets and model performance when integrated with Afaan Oromo text. Future considerations involve using Emojis for sentiment classification as positive or negative.

Data Availability

This study collects comments on Afaan Oromoo texts and emojis from Facebook public pages and YouTube platforms using the Face Pager software application and YouTube online comment scrapers respectively. https://gs.statcounter.com/social-media-stats/all/ethiopia

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Funding Statement

This works don’t have any funding assistance.

Authors Contributions

LBT, Investigating and executing the entire objective of the work, using proper research methodology and Development of Algorithms as per the research work and executing the program. TGK, Topic selection, supervising, executing the research as per the scientific principles. MAM, Overall Review and Mathematical understanding, programs verifications,

References

1. Abate, Jemal, and Meshesha Million. 2019. “*Unsupervised Opinion Mining Approach for Afaan Oromoo Sentiments*”.
2. Bhatia, Surbhi, Manisha Sharma, and Komal Kumar Bhatia. 2018. “Sentiment Analysis and Mining of Opinions.” Studies in Big Data 30(May): 503–23.Author, F., Author, S., Author, T. “Book title”. 2nd edn. Publisher, Location (2009).
3. Dave, Kushal, Steve Lawrence, and David M. Pennock. 2003. “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews.” Proceedings of the 12th International Conference on World Wide Web, WWW 2003: 519–28.
4. Duwairi, Rehab M. 2015. “Journal of Information Science.” (March).
5. Fikre, Turegn, and Addis Ababa. 2020. “Effect of Preprocessing on Long Short Term Memory Based Sentiment Analysis for the Amharic Language”.

International Journal of Science, Engineering and Technology
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

6. Ibrahim Bedane. 2015. "The Origin of Afaan Oromo : Mother Language." Double Blind Peer Reviewed International Research Journal 15(12).
7. Kuyu, Sultan Jenbo. 2021. "Developing an Automated Machine Learning Based Sentiment Analysis for Afaan Oromoo".
8. Lipton, Zachary C, John Berkowitz, and Charles Elkan. 2015. "A Critical Review of Recurrent Neural Networks for Sequence Learning ArXiv : 1506. 00019v4 [Cs.LG] 17 Oct 2015." 1–38.
9. Oljira, Megersa, and Kula Kekeba. 2020. "Sentiment Analysis of Afaan Oromo Using Machine Learning Approach." International Journal of Research Studies in Science, Engineering and Technology 7(9): 7–15.
10. Rase, Megersa Oljira. 2020. "Sentiment Analysis of Afaan Oromoo Facebook Media Using
11. Deep Learning Approach." New Media and Mass Communication: 7–22.
12. Sachdeva, Saransh Jitendra, Raj Abhishek, and Dr. Annapurna V.K. 2019. "Comparing Machine Learning Techniques for Sentiment Analysis." Ijarcce 8(4): 67–71.
13. Shelar, Amrita, and Ching-Yu Huang. 2018. "Sentiment Analysis of Twitter Data." Proceedings-2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018 3(1): 1301–2.
14. Tesfaye, Debela. 2011. "A Rule-Based Afan Oromo Grammar Checker." 2(8): 126–30. "Understanding LSTM Networks-Colah's Blog." 2016,1–8.
15. Vaitheeswaran, G, and L Arockiam. 2016. "Machine Learning Based Approach to Enhance the Accuracy of Sentiment Analysis on Tweets." Journal of Advance Research in Computer Science and Management Studies 6(6): 185–90.
16. Wayessa, Negessa, and Sadik Abas. 2020. "Multi-Class Sentiment Analysis from Afaan Oromo Text Based On Supervised Machine Learning Approaches." International Journal of Research Studies in Science, Engineering and Technology 7(7): 10–18.
17. Wondwossen Mulugeta. 2014. "A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts." HiLCoE Journal of Computer Science and Technology 2(2): 80–87.