

Research Article

Aspect-Based Sentiment Analysis for Afaan Oromoo Movie Reviews Using Machine Learning Techniques

Obsa Gelchu Horsa ¹ and Kula Kekeba Tune ²

¹Information Science Department, College of Informatics, Bule Hora University, Oromia, Ethiopia

²Software Engineering Department, Director of Center of Excellence for HPC and Big Data Analytics, AASTU University, Addis Ababa, Ethiopia

Correspondence should be addressed to Obsa Gelchu Horsa; obsa.gelchu@bhu.edu.et

Received 27 March 2023; Revised 19 July 2023; Accepted 20 November 2023; Published 7 December 2023

Academic Editor: Bhargav Appasani

Copyright © 2023 Obsa Gelchu Horsa and Kula Kekeba Tune. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aspect-based sentiment analysis (ABSA) is the subfield of natural language processing that deals with essentially splitting data into aspects and finally extracting the sentiment polarity as positive, negative, or neutral. ABSA has been widely investigated and developed for many resource-rich languages such as English and French. However, little work has been done on indigenous African languages like Afaan Oromoo both at the document and sentence levels. In this paper, ABSA for Afaan Oromoo movie reviews was investigated and developed. To achieve the proposed objective, 2800 Afaan Oromoo movie reviews were collected from YouTube using YouTube Data API. Following the data preprocessing, predetermined aspects of the Afaan Oromoo movie were extracted and labeled into positive or negative aspects by domain experts. For implementation, different machine learning algorithms including random forest, logistic regression, SVM, and multinomial naïve Bayes in combination with BoW and TF-IDF were applied. To test and measure the proposed system, accuracy, precision, recall, and f1-score were used. In the case of random forest, the accuracy obtained in combination with both BoW and TF-IDF was 88%. Using the SVM, the accuracy generated with BoW and TF-IDF was 88% and 87%, respectively. Applying logistic regression, the accuracy generated with both BoW and TF-IDF was 87%. Using multinomial naïve Bayes, the accuracy generated in combination with both BoW and TF-IDF was 88%. To improve the optimal performance evaluation parameters, different hyperparameter tuning settings were applied. The implementation result shows that the optimal values of models' performance evaluation parameters were generated using different hyperparameter tuning settings.

1. Introduction

The growth of the internet and the wide explosion of social media, such as Facebook, Twitter, YouTube, and blogs, have formed many new chances for people to express their attitudes towards any individual services, organizations, political parties, and government policy by using their native languages [1]. The monitoring of these sentiments on social media has piqued the interest of a growing number of researchers. It is difficult to manage such a significant amount of online content because many users express their views, comments, judgments, ideas, and opinions [2].

The rapid development and the popularity of social media networks result in a huge amount of user-generated

content being available online. Determining the polarity of this user-generated content has substantial benefits in different areas. In business industries, it allows companies to automatically gather the opinions of their customers about their products or services and identify areas of improvement [3]. It is a critical need for automated methods for opinion analysis, which allows a short time to process large amounts of data and understand the polarity of users' messages. Sentiment analysis (SA), also called opinion mining, is the task of natural language processing that concerns determining the writers, speakers, or another subject's attitude, concerning a certain topic or event.

According to [4], sentiment analysis (SA) is defined as the process of computationally identifying and categorizing

opinions expressed in a piece of text, particularly to determine whether the writer's attitude towards a specific topic or product is positive, negative, or neutral.

In natural language processing or machine learning, sentiment analysis or opinion mining has been investigated at three levels: sentence-level sentiment analysis (SA), document-level SA, and aspect-based sentiment analysis. The main intention of document- and sentence-level SA is that the overall polarity of the review text is considered irrespective of the attributes of an entity. A single comment can relate to numerous aspects of a single object, thus SA tasks at the sentence and document levels cannot handle sentences with multiple aspects [5].

However, users' comments may contain different aspects/features, such as "*the actor is so good, but this movie is just horrible.*" In this sentence, the objective of the reviewer was to give positive sentiment to the actor but negative sentiment to the movie [6].

The subfield of the study which studies the aspect of each entity with its corresponding sentiment is called aspect-based sentiment analysis (ABSA). It is a text analysis technique that categorizes data by aspect and identifies the sentiment attributes of each one. ABSA can be used to analyze customers' feedback by associating specific sentiments with different aspects of products or services [7,8].

In the previous works examined particularly in Afaan Oromo, the researchers have only focused on the document- and sentence-level sentiment analysis. But document- and sentence-level sentiment analysis cannot express full information about a certain product. In movie reviews, the overall review of whole sentences as positive, negative, or neutral cannot help movie users and movie producers identify what customers like or dislike.

In sentence-based and document-based sentiment analysis, the overall review of the customer did not mean that the customer or reviewer likes or dislikes each aspect in the reviewed sentence or document. But in ABSA, each aspect of an entity is extracted and the corresponding sentiment polarity prediction can be assigned for every extracted aspect terms of an entity as positive, negative, or neutral. In our case, the neutral sentiment prediction was not considered due to a few neutral sentiment predictions in subjective opinions. According to [2,3] in ABSA, the most frequently occurring opinions of users/customers about a particular aspect of an entity are either positive or negative.

In this article, the researchers proposed the ABSA model for Afaan Oromoo movie reviews using machine learning techniques. This ABSA was not examined in the previous works by the Afaan Oromoo SA, particularly for the Afaan Oromoo movie reviews. To achieve the proposed objective, the following research questions were examined by the researchers:

- (1) Which machine learning model is most appropriate for Afaan Oromoo aspect-based sentiment analysis for movie reviews?
- (2) What are the best attributes of Afaan Oromoo movie reviews for ABSA?

- (3) What are the main challenges in building Afaan Oromoo aspect-based sentiment analysis dataset for movie reviews?

- (4) What is the performance of the proposed ABSA system?

The main contribution of our paper is as follows:

- (1) ABSA models were developed for the Afaan Oromoo movie reviews.
- (2) The ABSA dataset was built by collecting 2800 Afaan Oromoo movie reviews.
- (3) ABSA dataset's labeling guideline was prepared for Afaan Oromoo movie reviews.
- (4) The paper can be used as a baseline for aspect-based sentiment mining-related research works for opinionated Afaan Oromoo text.

The great challenge that the researchers faced in our study is the lack of publically availability of datasets for ABSA. The comments extracted from the internet were full of grammatical errors, and it took a lot of time to clean and correct the collected comments manually. The amount of the dataset used for this research is not adequate and further researchers are expected to prepare a large dataset for their further study.

The research is limited to Afaan Oromoo, and the model developed for the Afaan Oromoo movie reviews cannot be generalized for other language movies or genres. To develop the Afaan Oromoo ABSA models, the supervised machine learning algorithms which were stated under Section 4 were used with accuracy, recall, precision, and $f1$ -score for model performance evaluation.

The rest of this paper is organized as follows. Section 2 describes related works in detail. Section 3 presents the methodology and materials used in this paper. Section 4 presents details of experimentation, evaluation, and analysis of generated results. Section 5 presents the conclusion and future works of the study.

2. Related Work

Sentiment analysis (SA) is the field of study that analyses people's opinions, sentiments, evaluations, attitudes, appraisals, and feelings towards entities such as products, companies, persons, issues, events, topics, and their features [3]. In the previous investigation, many researchers investigated their research in this opinion mining and sentiment analysis. Naturally, opinion mining and sentiment analysis has been conducted at three levels as follows: document-based sentiment analysis, sentence-based sentiment analysis, and aspect-based sentiment analysis (ABSA).

In document-level and sentence-level sentiment analysis, the overall polarity of the review text is considered without considering the attributes of the entity. But ABSA deals with extracting a fine-grained level from texts, providing very useful information for companies that want to know what people think about them or their products.

The work in [3] defined aspect-based sentiment analysis (ABSA) as the field of study which analyzes the reviewed text to extract entity, aspects/features of entity, opinion holder, time at which the text was posted to the web, aspect sentiment/polarity (i.e., positive, negative, and neutral), and in general, opinion quintuple found in the posted text to the web. Many scholars have been working on aspect-based sentiment analysis since 2014. But in most research, the researchers have extracted aspects/features of entity and sentiment corresponding to each aspect in the reviewed text.

The authors in [9] conducted a study aimed at Arabic aspect-based sentiment analysis using the deep learning technique with a pretrained BERT model. To achieve the proposed objective, the researcher has used the BERT model by tuning parameters using the Adam optimizer and the n-gram feature selection method was used for models other than pretrained BERT models. The Arabic hotel website was used to extract hotel comments to train the pretrained BERT model and other models which were 8320 corpus size. The report of the researcher revealed that the BERT model has out surpassed the state-of-the-art works and was robust to overfitting.

The work in [10] investigated aspect-based sentiment analysis as fine-grained opinion mining by using supervised machine learning models. To achieve the proposed objective, the researchers have collected 5995 corpus sizes from the two domains. From the total corpus size, 2687 were restaurant comment sentences and 3308 were review comments on the laptop. Extracted datasets were annotated manually by expert English speakers and in agreement was solved by a senior expert. The lexicon-based approach method was used to select features for model training.

The author of [11] proposed an aspect-based sentiment analysis search engine for social media data by using both deep learning techniques and machine learning approaches. To achieve the proposed objective, the researcher has collected data for model training from two domains, namely, restaurants and tourism. The collected dataset was preprocessed, and the top 10 intersecting aspects were extracted by both POS tagging and TF-IDF. For POS tagging feature extraction, NLTK's tag was used and the tagged features were stored in the text file format. But for adjectives, adverbs, and noun, the phrase algorithm was applied for feature extraction. For some models, researchers used TF-IDF for feature selection. From the implementation report, the researchers have seen that from a technical point of view, machine-learning techniques were identified to be more suitable for ABSA. The report of the researcher revealed that the performance of the proposed model was increased by deep learning techniques.

The author of [12] proposed a rule-based approach to aspect extraction performance using the POS tag-dependency pattern aspect-based sentiment analysis relationship. To achieve the proposed objective, the researcher has used the SemEval 2014 restaurant and laptop datasets. The total number of reviews in the dataset from the two domains is 7686. From the total number of datasets, 3845 sentences were laptop reviews and 3841 sentences were

restaurant reviews in the datasets. The generated datasets had already been preprocessed and labeled. POS tagging was used to identify nouns, noun-like words, and adjectives. Verbs and adverbs were also represented as sentiment words in the researcher's work, and in these cases, the researcher used the Penny Treebank English POS tag. Using Stanford Parser, dependency parsing was used to determine the syntactic grammatical dependency relationship between words in the reviewed sentence.

The authors in [13] proposed aspect-based sentiment analysis of higher education using the Java programming language. To achieve the proposed objective, the researcher has collected data from social media Twitter and Facebook through Twitter API and Facebook API. The collected data were preprocessed and relevant features were extracted by using a POS tag. The report of the researcher shows that sentiment classes positive or negative were assigned to every extracted aspect by using the Java Stanford NLP library.

Many sentiment analysis research studies have been conducted in English, Arabic, Chinese, and languages other than Afaan Oromoo. In the research conducted before in Afaan Oromoo and Amharic languages, the researchers focused on document-based sentiment analysis and sentence-based sentiment analysis.

The researcher of [14] proposed an unsupervised opinion approach for Afaan Oromoo sentiment classification using word lexicon. To achieve the proposed objective, the researcher has gathered the dataset from the OPDO's official Facebook page, political bloggers, and the OBN's official website. The researcher has gathered 600 corpus sample sizes. Following the completion of data pre-processing, the collected review text was assigned to one of the following three sentiment polarities: positive, negative, or neutral. To train his models, the researcher used a POS tagger for feature selection. HornMorpho and NLP tools were used by the researchers for Afaan Oromoo POS tagging. For the implementation purpose, the researchers conducted various experiments on unigrams, bigrams, and trigrams. The report of the researchers revealed that the performance obtained by bigram was superior to that of unigram and trigram.

Afaan Oromoo sentiment analysis from social media contents was investigated by the author in [15] using a deep learning approach. To achieve the proposed objective, the researcher has collected the data for the research implementation from the OPD social media webpage. The total review text collected was 1452. Following the dataset pre-processing, the researcher has used manual annotation to label the data into positive, negative, or neutral sentiment polarity. The researchers have used the n-gram feature extraction technique to train the proposed models. The labeled dataset was split into 80%/20% for model training and testing, respectively. The implementation report shows that the researchers have conducted different experimental results using CNN, LSTM, and MNB algorithms in combination with n-gram for feature extraction. The report of the researchers revealed that using MNB with unigram-trigram produced a more promising result than other proposed classifiers.

In [16], the researchers have aimed to investigate Afaan Oromoo multiclass sentiment analysis by using supervised machine learning techniques. To achieve the proposed objective, the researcher has collected primary data from OBN's official Twitter webpage. The collected corpus size was above 10,000, and the researchers have used only 1810 corpus sizes after preprocessing. Following data cleaning and normalization, 1810 reviewed text was annotated into very negative, negative, neutral, positive, or very positive sentiment polarity. The annotated dataset was split into 70%/30% for training and testing, respectively. From the implementation report, the researchers have seen that the support vector machine (SVM) and random forest (RF) algorithms were used in combination with TF-IDF feature selection techniques for model training and testing. The report of the researcher revealed that the accuracy generated by the support vector machine (SVM) was more promising than that of the random forest classifier.

The researcher [15] investigated Afaan Oromoo sentiment analysis at the character level using a deep learning approach. To achieve the proposed objective, the researcher has collected the dataset from Facebook and Twitter for investigation. Individually, from each domain, a 1200 corpus size was collected. The total corpus size which the researcher used for research implementation was 2400 corpus size. Following the data preprocessing, the collected dataset was annotated manually into very positive, positive, very negative, negative, and neutral. The implementation report shows that the researcher has used CNN and bidirectional long short-term memory (Bi-LSTM) deep learning algorithms for the research implementation. The total dataset was split into 70%/30% for training and testing, respectively. The report of the researcher revealed that the proposed algorithms performed promising performance-measuring metrics. The researchers stated that even if promising accuracy was obtained, there were many limitations in the investigation.

The author of [1] proposed Afaan Oromoo sentiment analysis using a machine learning approach. To achieve the proposed objective, the researcher has collected 1452 datasets from the OPD official Facebook page by using face graph API. Following the dataset preprocessing, the researcher annotated the dataset manually into positive or negative sentiment polarity. The labeled dataset was split into 70%/30% for model training and testing, respectively. From the report of the researcher, it can be seen that the researcher has employed multinomial naïve Bayes in combination with different n-gram values. The report of the researcher revealed that the accuracy obtained by unigram + bigram was more promising than other n-gram values.

To the best of the researchers' knowledge, in the previous works examined, there was no work done on Afaan Oromoo aspect-based sentiment analysis using machine learning techniques. Therefore, our proposed work on aspect-based sentiment analysis for Afaan Oromoo movie reviews using machine learning techniques is the first work and original. No other researchers considered it in the previous investigation. As a result, this research attempted to cover this gap in our investigation.

3. Methods and Materials

Under this methods and materials section, the researchers have introduced the methodology or all the steps that have been followed to investigate Afaan Oromoo aspect-based sentiment analysis. The dataset description, experimental design, data analysis, visualization tools, and the proposed Afaan Oromoo aspect-based sentiment analysis architecture for movie reviews are described clearly. The general methodology which was followed for Afaan Oromoo Aspect-Based Sentiment Analysis is depicted in Figure 1.

3.1. Experimental Design. Following the review of the literature on sentiment analysis, the researchers identified some potentially important factors and are likely familiar with the methodologies used to quantify that behavior. The major technique or strategy used to integrate diverse aspects of the research coherently and logically to ensure that the research problem is successfully addressed to the establishment of the variables is the experimental research design. The selection of an appropriate experimental research design is critical to the success of the project.

3.2. Data Source and the Dataset. This research adopted a manually annotated dataset built from Afaan Oromoo movie entertainment using machine learning techniques. This dataset comprised a review of different Afaan Oromoo movie reviews. The elicited reviews comprised the entire information gathered and posted during the years 2019 up to 2021. The dataset was collected from four different Afaan Oromoo movies, namely, *dheebuu*, *qormaata jaalalaa*, *hiree*, and *jaalala lakkuu*. The dataset collected from those Afaan Oromoo movie reviews was used only for research purposes. Figure 2 indicates the sample of an annotated dataset and its format.

3.2.1. Dataset Preprocessing. Data preprocessing is the process of cleaning and preparing data for analysis [17]. Preprocessing can also be used to reduce computational processes and feature space which can improve performance accuracy and classification. In the case of text classification, many preprocessing techniques can be used [18]. In machine learning and natural language processing (NLP), to get high performance of model accuracy, dataset preprocessing is very essential to exclude irrelevant data from the dataset. It is used to reduce the computational time of the model and increase the accuracy of classifiers. To prove this, the collected dataset was cleaned, Afaan Oromo stopword was removed, text was normalized, and the work of spell correction was performed.

(1) Dataset Cleaning. From our dataset, to predict the polarity of each extracted aspect, some irrelevant features such as opinion holder name, likes of the comments, opinion feedback holder, time at which the comment was posted, emoji used to express the sentiment for specific aspect, link in the dataset, lower casing characteristics, and non-Afaan Oromoo texts were removed from the dataset.

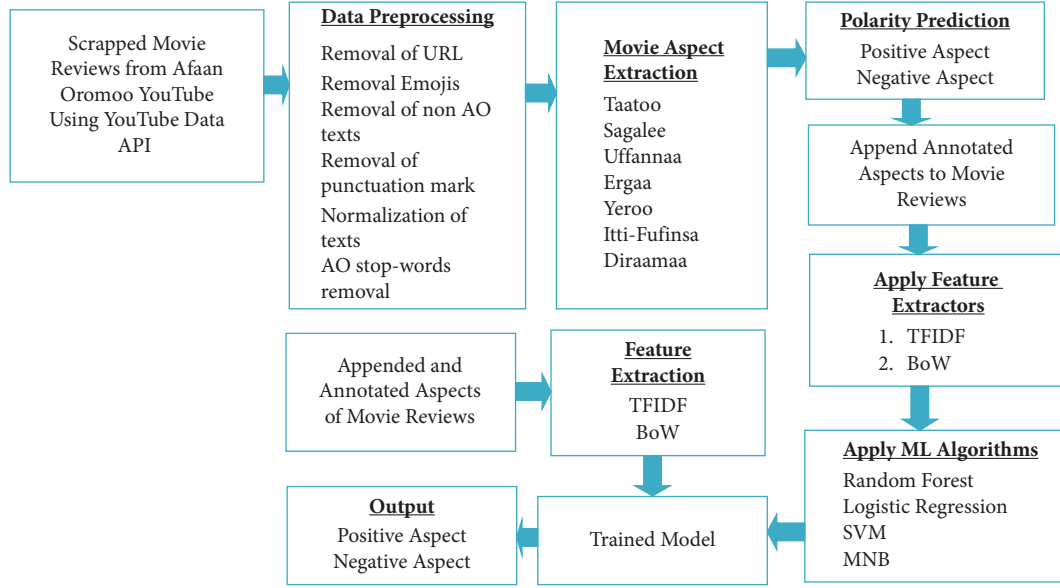


FIGURE 1: Proposed architecture for Afaan Oromoo ABSA for movie reviews.

1	Comments	Aspect-Terms	Polarity
2	maaloo shamarraan keenya safuu uummata keessanii maaliif dagatta	uffannaa	negative
3	Diraamaan namatti tola garuu sagaleen isaa ni hafa	sagalee	negative
4	baga nagaan dhufan oromoo koo eessa baddan? Addaan hin kutina	itti-fufinsa	negative
5	diraamaa akkam jaaldhuu dha maal godhu garuu yeroo isaatti gaba	yeroo	negative

FIGURE 2: Sample of an annotated dataset and its format.

(2) *Stopword Removal*. In English, stopwords are required to be excluded from the dataset. Including stopwords in the dataset will decrease the performance of classifiers. But some stopwords in Afaan Oromoo have a great role to determine the negativity of sentiment to particular subject/services. For example, “hin” and “miti” are Afaan Oromoo stopwords used to indicate the negativity of words/phrases. Also, the stopwords “dhufeera,” “hin dhufne,” “gaarii,” and “gaarii miti” have remained in the reviewed text since they reveal important information about a particular subject/service. Without these two key stopwords, the rest are removed from the reviewed text technically.

(3) *Data Normalization*. Words like “baay’ee” and “baayyee” have the same meaning but they have a different writing style. Only the difference is an apostrophe and the replacement of the apostrophe with “y.” Some reviewers like to elongate the comment they may write on social media. For example, to express their strong feeling about a particular aspect, they use long texts which are un-normal. For example, *waaaawwwwww* is normalized to *waawu*, and normalization of some numbers is into their equivalent words. For instance, “sagaleen kee 100% ni bareeda” is normalized to “sagaleen kee dhibbentaa dhibba guutuu ni bareeda.”

(4) *Spell Correction*. The user-generated content from social media is full of spelling errors. The researchers encountered

many wrongly written texts on social media. Therefore, the researchers highly worked on spell correction to correct misspelled texts to correct text. For example: “sagale ke bareda garu ufana kee siri mitti” which is misspelled and corrected to “sagaleen kee bareedaa, garuu uffannaa kee sirrii miti.” The great challenge of the study was the time taken for data preparation since there was no publically available dataset for Afaan Oromoo ABSA for movie review.

(5) *Tokenization*. Tokenization is the process of dividing a sequence of characters into tokens. It is an important unit in semantic processing. It entails breaking down a phrase, sentence, paragraph, or even an entire text document into smaller pieces, such as individual words or concepts. In this research, the researchers used Python’s NLTK library to tokenize the Afaan Oromoo sentences into smaller pieces of words.

(6) *Stemming*. Stemming is the process of developing morphological variants of root/base words. The stemmer approach reduces morphological variants of words to their base or root form. In morphologically rich languages like Afaan Oromoo, a stemmer will boost sentiment analysis significantly [1, 19]. Thus, for this research work, the researcher used Debela’s stemmer, which takes a word as the input and removes its affixes using a rule-based algorithm [20].

(7) *Lemmatization*. It is the process of merging two or more words into a single word. This method is used to facilitate the process of model training. Our models cannot understand sentences unless they stemmed or lemmatized into their roots which enables our algorithms to understand the insights from the dataset [21]. This analyses the word morphology and eliminates the ending of words; for example, “*rifate*” to “*rif*” and “*qabame*” to “*qab*.”

(8) *Feature Extraction*. In feature extraction, pieces of information are taken from the reviewed text and given to the algorithms for classification purposes. The advantage of feature selection is not only for dimension reduction but also to reduce overfitting. To extract features, the researchers have used bag of words (BoWs) and TF-IDF to compare and contrast the performance evaluation of selected algorithms.

(1) Bag of words (BoWs)

It is one of the simplest types of feature extraction models called as bag of words. The name bag of words refers to the fact that this model does not take the order of the words into account. Instead, one can imagine that every word is put into a bag, where the ordering of the words gets lost. Although there exist a few different variations of this model, the most common one is to simply count the number of occurrences of each word within a document and

keep the result in a vector. This way, the frequencies of the terms remain intact although grammar and order are lost [10]. To convert the text document into the vector representation form, CountVectorizer was used.

(2) Text frequency-inverted document frequency (TF-IDF)

IDF (inverse document frequency) is used to calculate the importance of terms in a document [21]. It is more effective and simple for feature selection scenarios. TF-IDF is easy to avoid hundreds of rare and unhelpful features from the reviewed text. The importance of the features in TF-IDF weighting is computed based on both term frequency and inverse document frequency (Oljira [1]). It is a statistical method of evaluating the importance of words in documents.

In TF-IDF, term frequency (tf) is used to measure the frequency of terms that occurred in a given document. It is a measurement of the number of times a term (t) occurs in a given document (d) against the whole number of words in the document indicated by $tf(t, d)$. Term frequency (tf) is directly proportional to the occurrence of a given term in a given document.

$$\text{Term frequency can be calculated as } TF(t, d) = \frac{\text{frequency of term } (t) \text{ in document } (d)}{\text{total number of terms } (t) \text{ in document } (d)}, \quad (1)$$

where inverse document frequency (IDF) measures the rate of words in a given document (which means the occurrence of a given word is common in the entire document). IDF represented as $idf(t, D)$ is a measure of how much a word

provides information in a given document (D). By a simple definition, it measures the worth weight of a given word in a given document (D). IDF indicates how a word commonly or rarely occurred through all given documents.

$$\text{Idf can be calculated as } IDF(t, D) = \log \left(\frac{\text{total number of documents } (D)}{\text{number of documents with term } (t) \text{ in that document}} \right). \quad (2)$$

The main impression behind TF-IDF follows that the term that appears frequently in a given document is less important than infrequent terms. TF-IDF customizes the vector space modeling system for text document representation. It is the dot product of term frequency and inverse document frequency.

$$TFIDF(t, d, D) = TF(t, d) \cdot IDF(t, D). \quad (3)$$

For implementations, the researchers used both bag of words (BoWs) and TF-IDF in combination with random forest, logistic regression, support vector machine, and multinomial naïve Bayes. The reason why the researchers selected both feature selection techniques is that in previous works examined, the researchers concluded and

recommended that in the sentiment analysis, they produced promising results when they compared and contrasted with other feature selection techniques [11, 22–25].

(9) *Aspect Term Extraction (ATE)*. Aspect term extraction is also known as target expression detection. It is dedicated to finding out an important term in a given sentence, which could be a single word or a phrase. It is an activity that identifies the aspects mentioned within a given sentence or paragraph [26, 27]. In this activity, the main intention is to extract the aspect expressed in the given reviewed sentence. The movie has different aspect terms and different researchers conducted previous research on different movie aspects.

Different ABSAs on movie reviews were reviewed and used as a baseline and benchmark for this work. For this research, the dataset annotators have extracted seven aspects of Afaan Oromo movie reviews, namely, *taatoo* (actor), *sagalee* (sound), *ergaa* (message), *uffannaa* (wearing), *itti fufinsa* (continuity), *diraamaa* (drama), and *yeroo* (duration) using the guideline prepared for this purpose. Figure 3 shows extracted aspect terms and their frequency distribution in the dataset.

(10) *Aspect Sentiment Prediction (ASP)*. Aspect sentiment determines the sentiment orientation associated with the corresponding aspects/opinion target. The fundamental goal of the aspect sentiment prediction (ASP) is to examine the sentiment associated with the aspects/features in the reviewed sentence. The sentiment associated with the particular aspects in the reviewed sentence may be positive, negative, or neutral. In our case, for extracted aspect terms, only positive or negative sentiment was predicted. For example, in the reviewed sentence “*diraamaa baay’ee bareedaa dha; garuu, uffaannaan isaanii gaarii miti*,” the reviewer expressed his/her positive opinion of the “*diraamaa*” aspect but negative opinion of the “*uffannaa*” aspect. In the second reviewed sentence, “*diraamaan baay’ee bareedaa dha; garuu maal godhan baay’ee gabaabbate*,” the reviewer expressed his/her positive opinion of the “*diraamaa*” aspect but a negative opinion of the “*yeroo*” aspect.

Following the aspect term extraction, the polarity of every extracted aspect was given as positive or negative sentiment polarity based on the view of the reviewer. In our case, neutral sentiment polarity was omitted due to few occurrences of it in subjective opinion, particularly in the movie reviews. Table 1 shows the aspect terms and its polarity distribution in the dataset.

In aspect-based sentiment analysis (ABSA), the aspect sentiment prediction (ASP) paves the way for industry, organization, or service providers to improve the features of the product or services that their customers’ like or dislike. This helps the industry, organization, or service providers to be competent with each other in the field of marketing. The organization, industry, or service providers can improve their products/services based on the sentiment associated with each product/service to be more productive in their business activity. In this research, every extracted aspect from the reviewed text was given a positive or negative sentiment polarity based on the view of the reviewer.

3.2.2. Data Annotation. Dataset annotation is the process of labeling the data to its corresponding output for model training. For this purpose, the researchers selected three human annotators (experts) who are native to Afaan Oromoo language from Bule Hora University to extract aspect terms from the collected dataset and to assign positive or negative sentiment predictions for the extracted aspect terms.

3.2.3. Annotation Guideline. A guideline for ABSA of Afaan Oromoo movie reviews was created by reviewing different scientific works in [28–30] and was provided for human annotators to make the annotation process transparent. Using

the provided guideline, the annotators labeled every extracted aspect of the movie into positive or negative sentiment polarity based on the sentiment expressed in the sentences.

For aspect term extraction, we used the following list of guidelines to extract 7 predefined aspect terms from the collected dataset.

- (1) For opinions focusing on the movie actors in general, the aspect would be *taatoo* (actor).
- (2) For opinions focusing on the movie’s message in general, the aspect would be *ergaa* (message).
- (3) For opinions focusing on the wearing style of the movie actors in general, the aspect would be *uffannaa* (message).
- (4) For opinions focusing on the continuity of series movies in general, the aspect would be *itti fufinsa* (continuity).
- (5) For opinions focusing on the duration of the movie in general, the aspect would be *yeroo* (time).
- (6) For opinions focusing on the quality of the movie sound in general, the aspect would be *sagalee* (sound).
- (7) For opinions focusing on the movie by itself in general, the aspect would be *diraamaa* (Drama).

For aspect sentiment polarity determination, we used the following controller and examples to assign it to positive or negative aspects based on the sentiment that is described in the sentence about aspect terms.

- (1) If positive indicators are found in the sentences which describe the aspect mentioned in the sentences, label it as a positive aspect.
- (2) If negative indicator words are there in the sentences which describe the aspect term mentioned in the sentences, label it as a negative aspect.

For example, in the sentence “*taatonni diraamaa kanaa ciccimoo dha; garuu uffaannaan isaanii kan safuu hawaasaa eege miti*,” there are two aspect terms, i.e., *taatoo* (actor) and *uffannaa* (wearing). In this sentence, the sentiment polarity to the aspect *taatoo* (actor) is positive which is indicated by the positive indicator word *ciccimoo* (clever) and the sentiment polarity to the aspect *uffannaa* (wearing) is negative which is indicated by the negative indicator word *miti*.

3.2.4. Inter-Annotator Agreement. Inter-annotator agreement (IAA) is a measure of how multiple annotators can make the same annotation decision for supervised natural language processing algorithms that use a labeled dataset that is often annotated by humans. In our case, the researchers gave equal comments for three annotators to extract movie aspects and label them into positive or negative aspects based on the guideline provided in Section 3.2.3.

To solve interannotator disagreement, the researchers computed Cohen’s kappa coefficient. To compute Cohen’s kappa coefficient, the dataset annotated by three annotators was used. The Cohen’s kappa coefficient value ranges between -1 and 1 .

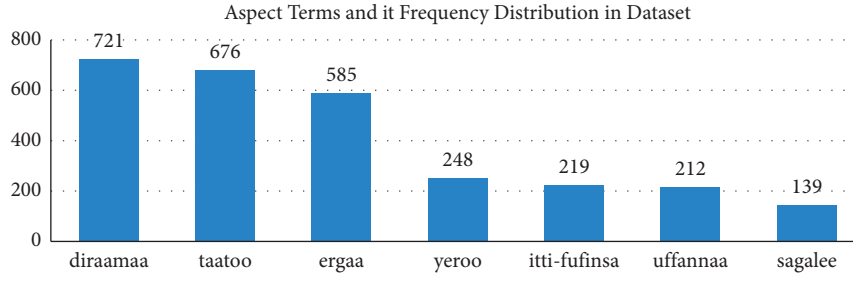


FIGURE 3: Aspect terms and their frequency distribution in the dataset.

(1) *Cohen's Kappa Coefficient (k)*. According to [31], Cohen's kappa coefficient (k) is a statistic that is used to measure the agreement among classifiers who each classify N items into C mutual categories. It is a quantitative measure of reliability for many classifiers that are rating the same thing, corrected for how often the classifiers may agree by chance. In this research, three annotators participated and they used the criterion to make the same assessment on the same extracted comments and good agreement was obtained based on Cohen's kappa coefficient which was 0.78.

Cohen's kappa coefficient can be computed as follows:

$$k = \frac{pa - p\epsilon}{1 - p\epsilon}, \quad (4)$$

where pa = the proportion of observation in agreement and $p\epsilon$ = the proportion in agreement.

Alternatively,

$$k = \frac{na - n\epsilon}{n - n\epsilon}, \quad (5)$$

where n = number of subjects, na = the number of agreements, and $n\epsilon$ = the number of agreements due to chance. The interannotator agreement of kappa value standards is depicted in Table 2.

3.3. The Development Environment and Tools. To implement aspect-based sentiment analysis for Afaan Oromoo movie reviews, the researcher used various tools and packages. The Anaconda Python distribution with version 3.7.11 was used to analyze the data. Python programming language was used for the implementation of the dataset because of its yet powerful, versatility, and ease of programming. It is also a useful tool for data visualization, with each cell utilized to run scripts selectively and independently. From the Anaconda Python distribution, Jupyter Notebook was to visualize the dataset and analyze the results using info-graphs. It is an open-source web application used to create and share documents that contain live code, equations, visualizations, and text on Internet browsers. ABSA of Afaan Oromoo movie reviews was implemented on a personal computer with processor Intel® Core™ i5-7200U, CPU 2.5 GHz, 4 RAM, hard disk with 960 GB storage capacity, and Windows 10 pro, 64 bits.

4. Implementation and Results

To implement the dataset, the researchers used four supervised machine learning algorithms, namely, random

TABLE 1: Aspects terms and their polarity frequency per reviews.

Aspects	Polarity	Frequency
Taattoo	Positive	582
	Negative	94
Sagalee	Positive	65
	Negative	74
Ergaa	Positive	412
	Negative	173
Uffannaa	Positive	119
	Negative	93
itti fufinsa	Positive	0
	Negative	219
diraamaa	Positive	671
	Negative	50
yeroo	Positive	119
	Negative	93

TABLE 2: Interannotator agreement kappa value standards.

Agreement interpretations	Kappa coefficient values
Less than chance agreement	Less than 0.0 slight
Slight agreement	Between 0.0 and 0.2
Fair agreement	Between 0.2 and 0.4
Moderate agreement	Between 0.4 and 0.6
Good agreement	Between 0.6 and 0.8
Very good agreement	Between 0.8 and 1.0
Perfect agreement	1.0

forest (RF) classifier, logistic regression (LR) classifier, support vector machine (SVM) classifier, and multinomial naïve Bayes classifier in combination with bag of words (BoWs) and TF-IDF vectorizer. For model performance evaluation, the researchers used accuracy, precision, recall, and $f1$ -score.

The first supervised machine learning algorithm used to implement the dataset is the random forest (RF) classifier. The performance evaluation metrics obtained by using RF in combination with BoW and TF-IDF were 88% of accuracy, 93% of precision, and 91% of $f1$ -score. But only the model produced different recall values with these different feature selection techniques (i.e., BoW and TF-IDF). The recall value obtained by the random forest classifier with both BOW and TF-IDF were 90% and 89% separately. The only difference that occurred between the two feature extraction techniques (i.e., BoW and TF-IDF) is only the recall metric which was 90% and 89%, respectively. The generated classification

report revealed that using a random forest classifier with either BoW or TF-IDF was promising for Afaan Oromoo aspect-based sentiment analysis for movie review.

The second supervised machine learning algorithm used to implement the dataset is the logistic regression classifier used in combination with both BoW and TF-IDF feature extraction techniques. From the results of implementation, the researchers observed that almost all equal performance evaluation parameters, namely, accuracy, recall, and $f1$ -score of 87%, 87%, and 91%, respectively, were generated. The report depicted that the only difference that occurred by using both feature extraction techniques (i.e., BoW and TF-IDF) using logistic regression was only the precision parameter. The precision values obtained using logistic regression with BoW and TF-IDF were 95% and 96%, respectively.

Based on the generated classification report during implementation, the researchers observed that using logistic regression in combination with either BOW or TF-IDF produced promising evaluation parameters for Afaan Oromoo aspect-based sentiment analysis for movie reviews.

The support vector machine (SVM) is the third supervised machine learning algorithm that is used to develop the ABSA model for the Afaan Oromo movie review. The classification report obtained using the support vector machine learning algorithm revealed that by using BoW and TF-IDF feature selection with the SVM classifier, different performance evaluation parameters were obtained except for the $f1$ -score which was the same for both feature extraction techniques, which is 91%. The performance evaluation parameters obtained from SVM + BoW were 87% accuracy, 95% precision, 86% recall, and 91% $f1$ -score. The accuracy, precision, and recall values obtained by SVM + TF-IDF were 88%, 93%, and 90%, respectively. The classification report shows that the performance evaluation parameters obtained using the support vector machine (SVM) with TF-IDF were better than that of BoW in accuracy. The generated classification report shows that for Afaan Oromoo aspect-based sentiment analysis for movie review, using the SVM in combination with either bag of words (BoWs) or TF-IDF was promising.

The last algorithm used to implement the system is the multinomial naïve Bayes (MNB) classifier. The generated result from the classification report shows that good performance evaluation parameters were obtained by using multinomial naïve Bayes in combination with both bag of words (BoWs) and TF-IDF. The classification report shows that the obtained evaluation parameter using multinomial naïve Bayes in combination with both BoW and TF-IDF was promising. Specifically, the accuracy and $f1$ -score obtained with both BoW and TF-IDF were 88% and 91%, respectively. But other performance evaluations parametric such as precision and recall were produced differently. The precision and recall values generated with TF-IDF were 94% and 88%, respectively, whereas precision and recall values generated with BoW were 93% and 90% individually. The generated classification report revealed that using multinomial naïve Bayes with either BoW or TF-IDF is promising for Afaan Oromoo aspect-based sentiment analysis for movie review. Table 3 shows the comparisons of the models.

The accuracy obtained by logistic regression was slightly lower than the other selected algorithms. The reason why the performance of logistic regression is slightly lower than other selected algorithms was that some machine learning algorithms such as linear and logistic regression can suffer poor performance if there are highly correlated attributes in our dataset [32]. This indicates that in the dataset, attributes are highly correlated, which affects the performance of the logistic regression classifier.

4.1. Hyperparameter Tuning. In machine learning, the term “hyperparameter tuning” refers to a process where the default parameters of the model or algorithm are modified or tuned to increase accuracy and performance. Because the data can vary depending on the issue statement, there are situations when using the algorithms’ default parameters will not work with the data that are now available. To improve the model’s performance, hyperparameter tuning becomes a crucial component of the model construction process. Table 4 shows the configuration made for each selected algorithm to improve performance evaluation parameters.

The main advantage of hyperparameter tuning is to improve the performance of models in machine learning algorithms. Table 4 shows the configuration made to improve the performance of the developed models. For tuning the algorithms, parameters BoW and TF-IDF were applied for all the selected algorithms. Using n _estimators for random forest along with both BoW and TF-IDF, 90% of accuracy, 95% of precision, and 92% of $f1$ -score with different recall values were obtained.

For logistic regression, the solver parameter optimization technique was applied. There are different parameter optimization techniques for solver. In this paper, the *sag* hyperparameter optimization technique was applied along with BoW and TF-IDF feature selection techniques. Using *sag* optimization for logistic regression in combination with BoW and TF-IDF, 89% of accuracy, 96% of precision, and 92% of $f1$ -score were generated with different recall values.

For the support vector machine (SVM), the rbf kernel function was applied. Using the SVM in combination with BoW and TF-IDF along with the rbf kernel function, different values of accuracy, precision, and recall were generated with 92% of $f1$ -score.

For multinomial naïve Bayes (MNB), hyperopt parameter optimization technique was applied in combination with BoW and TF-IDF feature selection techniques. Using MNB in combination with BoW and TF-IDF along with the hyperopt parameter optimization technique, 90% accuracy and 92% of $f1$ -score were generated with different values of precision and recall.

The implementation reports of ABSA for Afaan Oromoo movie reviews shows that using hyperparameter tuning, almost all the same values of accuracy and $f1$ -score were generated with different values of precision and recall along with BoW and TF-IDF feature selection techniques. These results revealed that optimal values of model performance evaluation parameters were generated using different hyperparameter tuning settings.

TABLE 3: Comparison of classifiers with BoW and TF-IDF feature extraction methods.

Models with feature selection methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Random forest + BoW	88	93	90	91
Random forest + TF-IDF	88	93	89	91
Logistic regression + BoW	87	95	87	91
Logistic regression + TF-IDF	87	96	87	91
SVM + BoW	87	95	86	91
SVM + TF-IDF	88	93	90	91
MNB + BoW	88	93	90	91
MNB + TF-IDF	88	94	88	91

TABLE 4: Hyperparameter settings.

Models	Hyperparameter tuning applied
Logistic regression	Solver (sag)
Random forest	$n_estimators$
SVM	Rbf
MNB	Hyperopt

5. Conclusion and Future Work

In the previous works examined particularly in Afaan Oromo, the researchers have only focused on the document- and sentence-level sentiment analysis. But document- and sentence-level sentiment analysis cannot express full information about a certain product. In sentence-based and document-based sentiment analysis, the overall review of the customer did not mean that the customer or reviewer likes or dislikes each aspect in the reviewed sentence or document. But in ABSA, each aspect of an entity is extracted and the corresponding sentiment polarity prediction can be assigned for every extracted aspect terms of an entity as positive, negative, or neutral.

In this paper, aspect-based sentiment analysis for Afaan Oromoo movie reviews using machine learning techniques was proposed and implemented. To achieve the proposed objectives, the researchers collected 2800 datasets from Afaan Oromoo YouTube movie reviews using YouTube Data API. From the collected dataset, the predefined seven aspects of Afaan Oromoo movie reviews which are listed in Table 1 were extracted by three human annotators with the support of the guideline provided for them which is discussed in Section 3.2.3. The extracted seven aspects of the movie reviews were assigned into positive aspects or negative aspects based on the sentiment polarity expressed in the sentences by the selected experts. The disagreement which was among the dataset annotators was solved by using Cohen's kappa coefficient which is described in Section 3.2.4.

The selected classification algorithms used in this study were random forest, logistic regression, support vector machine, and multinomial naïve Bayes (MNB) with two feature selection techniques, namely, bag of words (BoWs) and TF-IDF. In our data preprocessing, after the aspect terms were extracted from the review text, it was appended with the review text, and each algorithm used in combination with bag of words (BoWs) and TF-IDF feature extraction techniques for model training was investigated.

For each proposed algorithm, for implementation, the researchers have used two feature extraction techniques, which were bag of words (BoWs) and TF-IDF. The experimental result revealed that our proposed random forest algorithm with both bag of words (BoWs) and TF-IDF produced the same accuracy, which was 88%. The logistic regression produced 87% accuracy with both bag of words (BoWs) and TF-IDF feature extraction techniques. The accuracy produced by the SVM with BoW and TF-IDF was slightly different. With BoW, 88% accuracy was generated, whereas 87% accuracy was obtained with TF-IDF, and by using multinomial naïve Bayes (MNB), 88% accuracy was generated with both BoW and TF-IDF feature extraction techniques. Table 3 shows the comparisons of ABSA models developed for Afaan Oromoo movie reviews.

In general, in this study, the classification report generated from each proposed classification algorithm revealed that almost all algorithms produced the same performance evaluation metrics, specifically accuracy and $f1$ -score. To improve the optimal performance evaluation parameters, different hyperparameter tuning techniques which were discussed in Section 4.1 were applied for every selected algorithm. The implementation reports of ABSA for Afaan Oromoo movie reviews show that using hyperparameter tuning, almost all the same values of accuracy and $f1$ -score were generated with different values of precision and recall along with BoW and TF-IDF feature selection techniques. These results revealed that optimal values of model performance evaluation parameters were generated using different hyperparameter tuning settings.

But slightly different values of precision and recall were generated by the four classification algorithms with both BoW and TF-IDF. Based on the classification report generated by four classification algorithms with both BoW and TF-IDF, it was prominent for Afaan Oromoo aspect-based sentiment analysis for movie review. As a result, for Afaan Oromoo aspect-based sentiment analysis for movie review, from the proposed classification algorithms, the researchers can use one of it with either BoW or TF-IDF feature extraction technique.

Ultimately, investigation of sentiment analysis can be used in different sectors of online commerce to learn how customers feel about the competitors' products/services, to discuss what individual customers have to say about their products/services, and to control their brand's reputation in the market place, determining its strength and weakness.

For future work, the researchers may consider emoticons and emoji used to express polarity to respective aspects of movie review, multiclass classification, and standard dataset for ABSA for Afaan Oromoo movie reviews within a large size and also using deep learning algorithms would be future works of the investigation.

Data Availability

The datasets used to support the finding of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

The authors Obsa Gelchu Horsa and Kula Kekeba Tune confirm full responsibility for the ideation and design of the study, data gathering, analysis and interpretation of the findings, and writing of the manuscript.

Acknowledgments

We would like to take this opportunity to express their heartfelt appreciation to their research advisor, Kula Kekeba Tune (Ph.D.), Assistant Professor, Director, Center of Excellence for HPC and Big Data Analytics, and Assistant Professor, AASTU University, Department of Software Engineering, for providing invaluable guidance throughout this research investigation. His dynamism, vision, sincerity, and motivation have all left an indelible impression on us. Working and studying under his direction was a wonderful honor and privilege for us. We are deeply grateful to the advisor for what he has provided for us. The research was supported by Bule Hora University.

References

- [1] M. Oljira, "Sentiment analysis for Afaan Oromoo using combined convolutional neural network and bidirectional long short-term memory," *International Journal of Advanced Research in Engineering and Technology*, vol. 11, no. 11, pp. 101–112, 2020.
- [2] M. Ganapathibhotla and B. Liu, "Mining opinions in comparative sentences," in *Proceedings of the 22nd International Conference on Computational Linguistics- COLING'08*, pp. 241–248, Manchester, UK, August 2008.
- [3] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.
- [4] I. Lazrig and S. L. Humpherys, "Using machine learning sentiment analysis to evaluate learning impact," *Information Systems Electronic Journal*, vol. 20, no. 1, p. 20, 2022.
- [5] C. Sun, L. Huang, and X. Qiu, "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence," 2019, <https://arxiv.org/abs/1903.09588>.
- [6] M. Hoang and J. Rouces, "Aspect-based sentiment analysis using bert," in *Proceedings of the 22nd nordic conference on computational linguistics*, pp. 187–196, Linköping University. Electron. Press, Turku, Finland, October 2019.
- [7] Z. Zeng, J. Ma, M. Chen, and X. Li, "Joint learning for aspect category detection and sentiment analysis in Chinese reviews," *Lecture Notes in Computer Science*, vol. 11772, pp. 108–120, 2019.
- [8] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Information Sciences*, vol. 488, pp. 190–204, 2019.
- [9] M. M. Abdelgwad, T. H. A. Soliman, and A. I. Taloba, "Arabic aspect sentiment polarity classification using BERT," *Journal of Big Data*, vol. 9, no. 1, p. 115, 2022.
- [10] G. O. Diaz, X. Zhang, and V. Ng, "Aspect-based sentiment analysis as fine-grained opinion mining," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 6804–6811, Marseille, France, May 2020.
- [11] M. S. Alamanda, "Aspect-based sentiment analysis search engine for social media data," *Core Systems Integration Transactions on Information and Communication Technology*, vol. 8, no. 2, pp. 193–197, 2020.
- [12] A. S. Shafie, "Aspect extraction performance with POS tag pattern of dependency relation in aspect-based sentiment analysis," in *Proceedings of the 2018 Fourth International Conference on Information Retrieval and Knowledge Management*, pp. 1–6, Kota Kinabalu, Malaysia, March 2018.
- [13] L. Balachandran and A. Kirupananda, "Online reviews evaluation system for higher education institution: an aspect based sentiment analysis tool," in *Proceedings of the 2017 11th International Conference on Software, Knowledge, Information Management and Applications*, Malabe, Sri Lanka, December 2017.
- [14] J. Abate, *Unsupervised Opinion Mining Approach for Afaan Oromoo Sentiments*, Haramaya University, Dire Dawa, Ethiopia, 2019.
- [15] M. O. Rase, "Sentiment analysis of afaan oromoo facebook media using deep learning approach," *New Media Mass Commun*, vol. 90, pp. 7–22, 2020.
- [16] N. Wayessa and S. Abas, "Multi-class sentiment analysis from afaan Oromo text based on supervised machine learning approaches," *International Journal of Research Studies in Science, Engineering and Technology*, vol. 7, no. 7, pp. 10–18, 2020.
- [17] S. Vanaja, "Aspect-level sentiment analysis on E-commerce data," in *Proceedings of the 2018 International Conference on Inventive Research in Computing Applications*, pp. 1275–1279, Coimbatore, India, July 2018.
- [18] A. W. Pradana and M. Hayaty, "The effect of stemming and removal of stopwords on the accuracy of sentiment analysis on Indonesian-language texts," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 4, no. 3, pp. 375–380, 2019.
- [19] W. Tariku, *Sentiment Mining and Aspect Based Summarization of Opinionated Afaan Oromoo News Text*, Debre Berhan University, Debre Berhan, Ethiopia, 2017.
- [20] D. Tesfaye, *Designing a Stemmer for Afaan Oromo Text: A Hybrid Approach*, Addis Ababa, Addis Ababa, Ethiopia, 2010.
- [21] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," *Procedia Computer Science*, vol. 152, pp. 341–348, 2019.
- [22] H. M. S. T. Sandaruwan, S. A. S. Lorensuhewa, and M. A. L. Kalyani, "Sinhala hate speech detection in social media using text mining and machine learning," in *Proceedings of the 19th International Conference on Advances in ICT for Emerging Regions*, pp. 1–8, Colombo, Sri Lanka, September 2019.

- [23] X. Wei, "Aspect based sentiment analysis on review data," *FIU Electronic Thesis and Dissertation*, vol. 3721, 2017, <https://digitalCommons.fiu.edu/etd/3721>.
- [24] N. B. Defersha and K. K. Tune, "Detection of hate speech text in afan Oromo social media using machine learning approach," *Indian Journal of Science and Technology*, vol. 14, no. 31, pp. 2567–2578, 2021.
- [25] H. W. Kululo, *Information Filtering of Social media Amharic Texts Based on Sentiment Analysis*, Addis Ababa University, Addis Ababa, Ethiopia, 2020.
- [26] S. Poria, D. Hazarika, N. Majumder, and R. Mihalcea, "Beneath the tip of the iceberg: current challenges and new directions in sentiment analysis research," *IEEE Transactions on Affective Computing*, vol. 14, pp. 1–30, 2020.
- [27] W. Xue, *Aspect Based Sentiment Analysis On Review Data*, FIU Digital Commons, Florida, USA, 2017.
- [28] O. De Clercq and V. Hoste, *SemEval 2016 Task 5 Aspect Based Sentiment Analysis (ABSA-16) Annotation Guidelines*, Association for Computational Linguistics, San Diego, CA, USA, 2016.
- [29] P. M. D. J., "SemEval 2014 task 4: aspect based sentiment analysis annotation guidelines," *SemEval*, vol. 37, pp. 610–614, 2014.
- [30] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: aspect based sentiment analysis," in *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pp. 486–495, Denver, CL, USA, June 2015.
- [31] J. M. Bland, *Measurement in Health and Disease: Cohen's Kappa*, York University Department Health Science Department, Toronto, Ontario, Canada, 2008.
- [32] J. Brownlee, *Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models and Work Projects End-To-End*, Machine Learning Mastery, Melbourne, Victoria, Australia, 2016.