ASPECT BASED SENTIMENT ANALYSIS FOR AFAAN OROMOO TEXT USING BERT

FETIYA FURI

HAWASSA UNIVERSITY, HAWASSA, ETHIOPIA

Oct, 2024

ASPECT BASED SENTIMENT ANALYSIS FOR AFAAN OROMOO TEXT USING BERT

FETIYA FURI

MAJOR ADIVISOR: MICHAEL MELESE (PhD)

CO-ADIVISOR: GIZAW M. (MSc)

A THESIS SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
HAWASSA UNIVERSITY
INSTITUTE OF TECHNOLOGY

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER SCIENCE

HAWASSA, ETHIOPIA

Oct, 2024

APPROVAL SHEET-I

This is to certify that the thesis entitled, “Aspect based sentiment analysis for Afaan Oromoo text using BERT” submitted in partial fulfillment of the requirements for the degree of Master's with specialization in Computer Science, the Graduate Program of the Faculty of Informatics, and has been carried out by Fetiya Furi. Therefore we recommend that the student has fulfilled the requirements and hence hereby can submit the thesis to the department.

___________________________	_________________	_________________
Name of major advisor	Signature	Date

___________________________	_________________	_________________
Name of co-advisor	Signature	Date

EXAMINERS' APPROVAL SHEET

SCHOOL OF GRADUATE STUDIES

HAWASSA UNIVERSITY EXAMINERS' APPROVAL SHEET

We, the undersigned, members of the Board of Examiners of the final open defense by Fetiya Furi have read and evaluated her thesis entitled "Aspect based sentiment analysis for Afaan Oromoo text using BERT", and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree.

______________________	________________	___________
Name of Major Advisor	Signature	Date
______________________	________________	___________
Name of Internal Examiner-I	Signature	Date
______________________	________________	___________
Name of Internal Examiner-II	Signature	Date
______________________	________________	___________
Name of External examiner	Signature	Date
______________________	________________	___________
SGS Approval	Signature	Date

Final approval and acceptance of the thesis is contingent upon the submission of the final copy of the thesis to the School of Graduate Studies (SGS) through the Department/School Graduate Committee (DGC/SGC) of the candidate's department.

Stamp of SGS Date:____________

STATEMENT OF THE AUTHOR

I hereby declare that this MSc thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis have been duly acknowledged.

Name: Fetiya Furi Signature: ……….……………

Place: Institute of Technology, Hawassa University, Hawassa

Date of Submission: ………………………………………….

Acknowledgment

First of all praise is to Allah, the almighty for guiding me at every stage of my life. I would like to express my deepest gratitude to my thesis advisor, Dr.Michael Melese, whose guidance, support, and encouragement have been invaluable throughout this journey. Your insights and expertise have greatly influenced the direction of my research, and I am profoundly grateful for your advice. I would like to thank my Co-Advisor, Mr. Gizaw M. for your support and encouragement. In addition, I want to thank Dilla University for sponsoring me. I am also deeply thankful to Hawassa University and the Department of computer science for providing the resources and environment necessary to complete this research. I also extend my thanks to my parents for their support every time and everywhere. Finally, I would like to thank everyone who, in one way or another, has contributed to the successful completion of this thesis. Your support, no matter how small, has been deeply appreciated.

List of acronyms

ABSA	Aspect-based sentiment analysis
AE	Aspect extraction
ACD	Aspect Category detection
ATE	Aspect Term Extraction
ASC	Aspect Sentiment Classification
BERT	Bidirectional Encoder Representations from Transformer
BiLSTM	Bidirectional Long Short-Term Memory
BIO	Beginning Inside Outside
CNN	Convolutional Neural Network
CLS	Classification token
NLP	Natural language processing
RNN	Recurrent Neural network
SEP	Separator token

Contents

List of figure

List of table

List of equation

Abstract

Aspect-based sentiment analysis (ABSA) is a more important and advanced task of sentiment analysis which determine both the sentiments and the aspects within the text. It is an essential research field within natural language processing, especially for languages that lack extensive resources. This study focuses on developing an ABSA model for Afaan Oromoo language, one of the widely spoken languages in Ethiopia. Despite the rich linguistic diversity of Afaan Oromoo, there is a scarcity of computational tools and datasets for sentiment analysis in this language. Our research addresses this gap by creating a comprehensive dataset annotated with BIO annotation scheme for aspect terms and integrates CNN and BiLSTM for aspect extraction, and BERT for aspect sentiment classification. We fine-tuned pre-trained BERT model on our annotated Afaan Oromoo dataset to perform aspect based sentiment analysis. The total of 2550 review text collected from FBC Afaan Oromoo Facebook page, BBC Afaan Oromoo and other relevant social media are used for this study. After data collection, two annotators' annotated data manually into three classes (i.e., positive, negative and neutral). The aspect terms used for study are extracted from three domain, coffee, gold and flower. Basically ten aspect terms namely (qulqullinna bunaa, oomisha bunaa, foolii, dandhama, worqee baasuu, galii, gatii, diinagdee, agarsiisa worqee and al-ergii) are used for the study. CNN-BiLSTM is used for aspect extraction and performed 92.8% of accuracy. BERT model performed accuracy of 87% for aspect sentiment classification. This work not only contributes to the development of sentiment analysis for Afaan Oromoo but also provides a framework for applying advanced NLP techniques to other low-resource languages.

Keywords*: Aspect-Based Sentiment Analysis (ABSA); BIO Tagging; Bidirectional Long Short-Term Memory ; Bidirectional Encoder Representation from Transformers (BERT); Convolutional Neural Networks; Sentiment Classification; Aspect Term Extraction*

Chapter 1

Introduction

1.1 Background

Language is a means of human communication, expressed through spoken and written forms involving structured and covenant use of words. Language is used not only for communication but also for imparting emotion associated with it[1]. Natural language processing (NLP) is the study of mathematical and computational modeling of various aspects of language and the development of a wide range of systems [2]. These include spoken language systems that integrate speech and natural language; cooperative interfaces to databases and knowledge bases that model aspects of human interaction. Since its inception in 1950, NLP Research has been focusing on tasks such as machine translation, information retrieval, text summarization, Question and answering, topic modelling and more recently opinion mining [3]. Natural Language Processing holds great promise for making computer interfaces that are easier to use for people, since people will hopefully be able to talk to the computer in their own language, rather than learn a specialized language of computer commands [4].

E-commerce is a thriving industry that plays vital role in the global economy. With the rapid growth of social media, different users begin to express their sentiments on different online platforms. These comments express the sentiments of different users and consumers and provide government and sellers with more valuable feedback on the quality of goods or services [5], [6], [7]. A large number of public comments are collected by governments and companies directly from the Internet and analyze opinion of users and their satisfaction to meet their needs. For this reason, as a basic task of Natural Language Processing (NLP), sentiment analysis has attracted attention from the theoretical and practical circles [8].

Sentiment Analysis is the task of natural language processing that deals with determining the speakers, writers and another subject attitude with respect to a specific topic. It is the computational study of people's opinion toward events, topics and their attributes [9]. The rapid growth and widespread use of social media platform have led to a vast amount of user-generated content available online[10]. Analyzing the sentiment of this content offer significant advantages across various domain. In business, it helps companies to automatically collect opinions of customer about their products or services and helping them determine areas of improvement. In politics, it can provide insights into public opinion and political event reactions, aiding decision-making processes. As a fundamental task in natural language processing (NLP), sentiment analysis has gained significant attention from both academic and industry [10]. Its popularity has surged due to the increasing volume of opinionated content generated by users of internet.

A key task in sentiment analysis is determining the polarity of a given text at the document, sentence, or feature/aspect level. This involves identifying whether the opinion expressed in a document, sentence, or specific entity feature/aspect is positive, negative, or neutral. Sentiment analysis can be carried out at three different levels: document level, sentence level and aspect level [9]. Document-level sentiment analysis is perhaps the most extensively studied topic in the field of sentiment analysis especially in its early days [11], [12]. The goal is to categorize an opinion document (such as a product review) as conveying either a positive or negative sentiment, which are referred to as sentiment orientations or polarities. This task is known as document-level analysis since it treats each document as a whole, without examining specific entities or aspects within the document or identifying the sentiments directed toward them.

At the sentence level, each sentence is treated as a short document, and it can be classified as either subjective or objective. As author [13], argues subjective is an opinionated sentence that expresses sentiment. Sentence-level sentiment analysis aims to identify the sentiment or opinion expressed in an individual sentence, rather than the overall sentiment of the entire text. Standard sentiment analysis focuses on determining the overall sentiment of a text, but it does not capture crucial details such as the specific entity, topic, or aspect the sentiment is directed towards. Aspect-based sentiment analysis (ABSA), also known as target-based sentiment analysis, is a more advanced task that involves identifying both the sentiments and the aspects or targets within the text. While more detailed and complex, ABSA offers greater value in addressing the specific needs of customers and organizations

Aspect-Level Sentiment Analysis is the task of identifying fine-grained sentiment polarity toward specific aspect associated with a given target [12]. Aspect-based sentiment analysis (ABSA) recognizes that the sentences may describe more than single objects or present different perspectives on a single object. This type of analysis is designed to evaluate the sentiment polarity directed at a specific object or viewpoint, rather than treating the entire sentence as having a single overall sentiment. In the sentence "I bought this flower at expensive price, but I loved it because of its beautiful color". There are two aspects in above sentences; price and color, the sentiment polarity of price is negative whereas the sentiment polarity of color is positive.

Aspect Terms Extraction (ATE) is a fundamental task in aspect-based sentiment analysis [14]. It is applicable and widely used in research areas such as aspect-based sentiment analysis. The aim of ATE task is to extract opinion targets from the review sentences. The motivation behind aspect extraction is that users often express varying opinions about different aspects of a product[10]. Aspect extraction allows sentiment analysis to be conducted at a more detailed and granular level by identifying these aspects. Therefore, Aspect based sentiment analysis can provide a deeper understanding opinion of user.

Aspect Sentiment Classification (ASC) is an interesting and challenging research task to identify the sentiment polarities of aspect words in sentences [15]. ASC, machine learning models a series of features, e.g., a set of words and sentiment dictionaries, were set up to train classifiers [16]. Their classification effect heavily depended on the features' quality. However, those methods rely on carefully designed manual features on large-scale datasets, resulting in a lot of waste of manpower and time[5]. The neural network model can automatically learn the low dimensional representation of reviews without relying on artificial feature engineering. Recently, neural network methods have dominated the study of ABSA since these methods can be trained end-to-end and automatically learn important features [5].

Transformers started a new era in the NLP field. Transformers are deep learning models that can handle sequential data but they don't require sequential data to be processed in order, unlike recurrent neural networks (RNNs)[17]. Many studies have showed that Transformers are highly successful in many NLP tasks, including summarization, translation, and classification[18]. BERT is one of the architectures that utilizes Transformers and the model, trained in an unsupervised manner on large datasets, can be utilized in many other NLP tasks[19]. BERT, which stands for Bidirectional Encoder Representations from Transformers is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers [19].

In recent years, several studies have proposed on deep-learning-based sentiment analyses, which have differing features and performance. This work looks at the latest studies that have used deep learning models, such as deep neural networks (DNN), recurrent neural networks (RNN), and convolutional neural networks (CNN), to solve different problems related to sentiment analysis (e.g., aspect extraction and aspect sentiment classification)[20]. We applied CNN-BiLSTM and fine tune BERT models to our datasets and implemented the state-of-the-art of aspect based sentiment analysis based on deep learning.

1.2. Motivation

Sentiment analysis plays a vital role in understanding public opinion, particularly in the digital age, where vast amounts of user-generated content are constantly produced. Although considerable progress has been made in sentiment analysis for widely spoken languages like English, there is limited work available for under-resourced languages such as Afaan Oromoo, despite it being the second most spoken language in Ethiopia and a major language among the Cushitic languages. The scarcity of natural language processing (NLP) tools and resources for Afaan Oromoo limits the ability of native speakers to utilize these technologies for sentiment analysis, which could be highly valuable in various sectors including social media monitoring, customer feedback, and public opinion research.

Aspect-based sentiment analysis (ABSA) goes a step further by determining the sentiment expressed in a text and identifying the specific aspects or features being discussed. This research aims to contribute to the field by developing an ABSA model for Afaan Oromoo texts, combining the strengths of BERT and deep learning techniques like CNN-BiLSTM to effectively extract aspects and accurately classify sentiments. The successful completion of this Research has the potential to enhance the computational resources available for Afaan Oromoo and promote further research in the language.

1.3. Statement of problem

With the rapid growth of internet social platforms, buyer reviews (such as comment texts) have become a key resource for consumers to understand products and make purchasing decisions[10]. Early sentiment analysis methods, primarily at the document and sentence levels, assumed that a text had only one overall sentiment. This made it difficult to fully capture the comprehensive opinions of consumers, often leading to misguided decisions[10]. Since analyzing opinions at the sentence or document level does not provide enough detailed information, more fine-grained aspect-level sentiment analysis is necessary.

Aspect-Based Sentiment Analysis (ABSA) is essential for understanding sentiments associated with specific aspects of entities, such as products or services, in textual data. Traditional sentiment analysis often fail to address the granularity needed for aspect-level insights, because they typically focus on classifying the overall sentiment of a text without considering the distinct sentiments attached to individual aspects within that text. For instance, a review of a product might express positive sentiment about quality but negative sentiment about price. A general sentiment model would struggle to capture these nuanced differences.

Studies on sentiment analysis of Afaan Oromoo language are conducted at sentence level and document level using different approaches. The first study on sentiment mining and aspect-based summarization of Afaan Oromoo news text was conducted by authors [21] using a rule-based approach. According to the authors' report, the method yielded promising results. However, the lack of linguistic resources, such as Part of Speech (POS) tagging and a lexical database, posed significant challenges to the work.

study in [22] focused on sentiment analysis for Afaan Oromoo by combining a CNN with bidirectional long short-term memory (BiLSTM). This research utilized a deep learning approach for character-level sentiment analysis, aiming to improve the accuracy and effectiveness of sentiment detection in Afaan Oromoo texts. Another study by [23] applied an unsupervised method for sentiment mining in Afaan Oromoo, conducting experiments using unigram, bigram, and trigram features. Their results showed that the unigram feature achieved high recall, while the bigram feature yielded high precision. Additionally, a study by [24] focused on document-level sentiment analysis for Afaan Oromoo using machine learning approaches.

They experimented with Naïve Bayes, Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN) algorithms.

Existing studies on sentiment analysis for Afan Oromoo have been conducted mainly at the sentence and document levels, utilizing different approaches such as rule-based methods, convolutional neural networks (CNNs) combined with bidirectional long short-term memory (BiLSTM), and unsupervised learning techniques. There are only two study on sentiment analysis for Afaan Oromoo Language at aspect level. The first works was focus on sentiment analysis and aspect level summarization and the other was aspect based sentiment analysis for Afaan Oromoo review text. The former was conducted using rule based approach and the later was based on machine learning approach. While these studies have made notable contributions to sentiment analysis for Afan Oromoo, they fail to address aspect-based sentiment analysis (ABSA), which provides more fine-grained insights by associating sentiments with specific aspects of a product or service. There is a significant gap in research on Afan Oromoo that combines the fine-grained aspect-level sentiment analysis with the powerful capabilities of modern language models like BERT.

This study aims to fill this gap by proposing a novel CNN-BiLSTM and BERT-based ABSA framework for Afan Oromoo, addressing the following key issues: first, efficient data annotation with BIO Tagging and Sentiment Labels. ABSA requires accurate labeling of both aspect terms and their corresponding sentiments within text. Using BIO tagging (Begin, Inside, Outside) is crucial because it allows for the precise identification of aspect terms. This structured annotation helps to distinguish different parts of a sentence that refer to aspects and assign appropriate sentiment labels to each aspect, ensuring the model captures the needed granularity. The second important issue is aspect term extraction. In ABSA, identifying the relevant aspect terms within a sentence is essential. Unlike traditional sentiment analysis that look at the entire text, ABSA models need to recognize where specific aspects (e.g., "coffee quality ") are mentioned. Designing models such as CNN-BiLSTM allows for the effective extraction of aspect terms by utilizing both convolutional layers for local feature detection and LSTM for long-range dependencies in the text.

Once the aspect terms are extracted, accurately classifying the sentiment (positive, negative, neutral) for each aspect is the next challenge. BERT, with its deep bidirectional understanding of language, is preferable for this task. By leveraging BERT's contextual embeddings, the model can capture the nuanced sentiment associated with each aspect, even in complex sentence structures. This structured approach, combining efficient BIO tagging, aspect identification using CNN-BiLSTM, and sentiment classification using BERT, leads to a robust ABSA model that overcomes the limitations of traditional sentiment analysis, providing more fine-grained and actionable insights.

To this end, this study attempts to address and answer the followings research questions.

1. How effective are pre-trained BERT model in performing aspect-based sentiment analysis on Afaan Oromoo text, considering the language's low-resource status?
2. What is the effectiveness of CNN-BiLSTM in extracting aspects from Afaan Oromoo text?

1.4. Objective of the study

1.4.1. General Objective

The main objective of this study is to design and develop aspect based sentiment analysis model for Afaan Oromoo text using CNN-BiLSTM and BERT.

1.4.2. Specific Objectives

To accomplish the general objective of the research specific objectives are as follow:

- ✓ To review literature to understand state-of-the-art in pre-trained model, deep learning and sentiment analysis.
- ✓ To collect, preprocess and prepare representative aspect based sentiment analysis data.
- ✓ To design and implement a Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) based model for extracting aspect terms from Afaan Oromoo text.
- ✓ To Fine-tune BERT for aspect sentiment classification on Afaan Oromoo text, assessing its performance in terms of accuracy, precision, recall, and F1-score.
- ✓ To report the finding of study for the upcoming research area in the field.

1.5 Methodology

A well-structured research methodology is essential for addressing a research problem systematically. The strategy for this research is outlined as follows: the researcher discusses the literature review, the sources of data utilized, the tools and algorithms applied in the research, and the approach used for analyzing and evaluate the findings. This methodology ensures a comprehensive and organized approach to uncovering new insights.

1.5.1. Literature review

For this study the researcher reviewed related work connected with aspect based sentiment analysis. We referred journal of article, conferences, white papers and aspect extraction and sentiment classification method created for different language.

1.5.2. Research design

This study adopt an experimental research design, where various deep learning models (CNN-BiLSTM) was developed and tested for aspect extraction and BERT for aspect sentiment classification.

1.5.3. Data Collection

For the Data source selection, the researcher used review text from FBC Afaan Oromoo Facebook page, BBC Afaan Oromoo news, and other social media as the primary data source. The researcher used Total of 2550 Afaan Oromoo dataset collected from three domain such as coffee, flower and gold for training and testing model. The data was labeled using the BIO tagging scheme for aspect terms, and sentiments are labeled in to three classes as positive, negative and neutral.

1.5.4 Preprocessing

Data preprocessing steps are applied in order to clean and prepare data set for aspect based sentiment analysis. The Preprocessing steps such as tokenization, normalization, removal of stop words, and removal of special characters are performed. Another advanced preprocessing such as Bert-tokenization and adding special tokens (CLS, SEP) applied to prepare dataset in BERT understandable way.

1.5.5. Design Approach

To develop aspect based sentiment analysis on Afaan Oromoo text, hybrid of CNN-BiLSTM is used for aspect extraction and Bidirectional Encoder Representations from Transformers (BERT) model is used for aspect sentiment classification.

1.5.6. Tools and techniques

This research involves preprocessing, aspect extraction and sentiment classification. The preprocessing contains removal of unnecessary texts, symbols and characters and comprises character and short form normalization. This task will performed using Python program. CNN is used for feature extraction and BiLSTM for understanding context and long-range dependencies, BERT for Aspect sentiment classification. We use python with Tensor Flow frame work for model implementation, and Hugging Face's transformers library for BERT. Pandas and NumPy are used for handling datasets and preprocessing.

1.5.7. Evaluation techniques

Evaluation of designed model is important as research focus on designing and developing model using different approaches, technique and tools. Different evaluation metrics such as accuracy, precision, recall and F1-Score are assess the performance of the sentiment classification models.

1.6. Significance of Study

Today, organizations use Sentiment Analysis to gain insights into public opinions and emotions regarding their products, services, and brand. This helps them better understand customer perceptions and make informed decisions. Aspect-based sentiment analysis can be used to analyze customer feedback by associating specific sentiments with different aspects of a product or service. It can help companies automatically analyze customer data, automate processes like customer support tasks, and gain powerful insights on the go. Customers are more vocal than ever, actively engaging with brands and providing feedback, both positive and negative. Every interaction, whether a mention or a comment, offers valuable insights into what businesses are doing well and where they can improve. These insights can help shape marketing strategies, inspire product development, and support competitive analysis.

This study will cover up the details, and express people' s fine-grained and comprehensive sentiments fully , leading people to right decisions and allows businesses to automatically analyze large amounts of data in detail. It contributes towards the realization of more advanced sentiment analysis and overcome the problem of standard sentiment analysis in addition to being an academic exercise to complete the program's requirements. The findings of the study can also use to help construct a full-fledged opinion that can employed in a mining system for Afaan Oromoo and any other Ethiopian language.

1.7. Scope and Limitation of study

1.7.1 Scope of study

The scope of this study is limited to designing and developing model for aspect based sentiment analysis for Afaan Oromoo text. Basically three product review of coffee, flower and gold are used. The aspect terms are extracted and sentiment is classified for each aspect as positive, negative, and neutral. Grammatically accurate sentiment texts are used in this study. Two important task of aspect based sentiment analysis: aspect extraction and aspect sentiment classification are covered in detail. This approach will only applicable to the Afaan Oromoo text domain.

1.7.2. Limitation of Study

Aspect based sentiment analysis consists two basic sub-tasks: aspect term extraction and aspect sentiment classification. Aspect term extraction aim to identify specific words or phrases, known as aspect terms, in a text that refer to particular entities or attributes about which opinion is expressed. Aspect expressions are specific words or phrases in a text that describe particular attributes about which an opinion is expressed. These expressions are highlight the aspects that are the subject of sentiment analysis. There are two types of aspect expressions: explicit aspect expressions and implicit aspect expression. Explicit aspect expression are directly mentioned in the text, while implicit aspect expressions are implied rather than directly mentioned. This study only focuses on explicit aspect expressions and does not cover implicit one. Additionally, the opinion holder, the person or source expressing the sentiment, and aspect category detection is not covered in this study.

1.8. Organization of the Research

This thesis have five chapters and its organization is as follow: chapter one cover introduction part that includes statement of problem, objective, methodology, Limitation and scope of study. Chapter Two provides a review of various types of literature on sentiment analysis, along with an overview of machine learning, deep learning techniques used in this field, Afaan Oromoo language and it’s writing system. Chapter three study's about methodology, model design and general approach, including corpus preparation and preprocessing, system design, classification strategies, and performance measurement. Chapter Four presents the experimental results and findings, detailing how the tests and procedures were implemented. Finally, Chapter five discusses the conclusions drawn from the research and outlines potential future work.

Chapter 2

Literature review

2.1. Overview

The different works that have been done before in the research area are discussed in this chapter. The key concepts that concern with sentiment analysis, level of sentiment analysis, aspect based sentiment analysis task which basically divided into aspect term extraction and aspect sentiment classification, different approaches of sentiment analysis such as lexical based, machine learning, deep learning approach, basic concept of bidirectional encoder representation from transformer (BERT) and The overview of language such as Afaan Oromoo writing system, characteristics of Afaan Oromoo, its linguistic structure, language word classes and any challenges related to sentiment analysis in the language is discussed in detail.

2.2. Sentiment Analysis.

Sentiment analysis (SA) is the field of study that examine people's opinions, emotions, evaluations, attitudes, appraisals and feelings toward entities like products, companies, individuals, issues, events, topics, and their characteristics[9]. Sentiment Analysis Measure the polarity of text by identifying and assessing the expression people use to evaluate or apprise persons, entities or events [11]. Sentiment analysis involves the automated detection of emotions or opinions conveyed in text, playing a crucial role in making informed decisions. As societies become increasingly familiar with platforms such as blogs, forums, Twitter, Facebook and other social media, people now have opportunity to express their opinions and emotions by posting reviews of products or services online[9].

According to [25] the word sentiment is defined as 'an attitude, thought or judgment prompted by feeling', it is also defined as 'a specific view or notion: opinion' and 'emotion'. The word 'opinion' usually referred to as 'a view, judgment or appraisal formed in the mind about a particular matter'. sometimes Sentiment analysis is referred to as opinion mining, and mostly, these two terms have identical meaning [26]. Nevertheless, some researchers explain that there exist small differences in notions of these two terms [26]. For example, [27] noted that opinion mining originated from the information retrieval (IR) community, with a focus on extracting and processing opinions about an entity. In contrast, sentiment analysis emerged from the natural language processing (NLP) community and is focused on identifying the sentiment expressed in a given text.

Moreover, [11] Explain the terminology and historical origins of these two terms, but emphasize that in a broader context, they refer to the same field of study. In [28] stated that, there are also other different names and slightly different tasks, for example, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining etc. The field sentiment analysis or opinion mining has recently gained a lot of attention from the researchers and markets; there has been a steady undercurrent of interest of analyzing opinions [25] .

Before the rise of Web 2.0, only a limited amount of opinionated text was available, which was one of the primary reasons for the lack of research in the field of sentiment analysis. As a result, much of the early work in text processing focused on mining and retrieving factual information, such as in information retrieval systems, text classification or text clustering[14]. Author [12] noted that people usually asked their friends or family for opinions before making a decision and an organization normally conducted opinion polls, surveys and focus groups to find out the sentiments of the general public about its products or services. Subsequently, numerous platforms were created and developed to bring the vision of Web 2.0 technologies to life.

opinion mining and sentiment analysis has been carried out at three levels as follows: Document-level sentiment analysis, sentence-level sentiment analysis, and aspect-based sentiment analysis (ABSA) [12]. In document-level and sentence-level sentiment analysis, the overall polarity of the text is assessed without taking the specific attributes of the entity into account. However, aspect-based sentiment analysis (ABSA) focuses on extracting more detailed insights from the text, offering valuable information for companies interested in understanding people's opinions about their products or services.

2.3. Levels of sentiment analysis

sentiment analysis can be carried out at three different levels: document level, sentence level and aspect level[12].

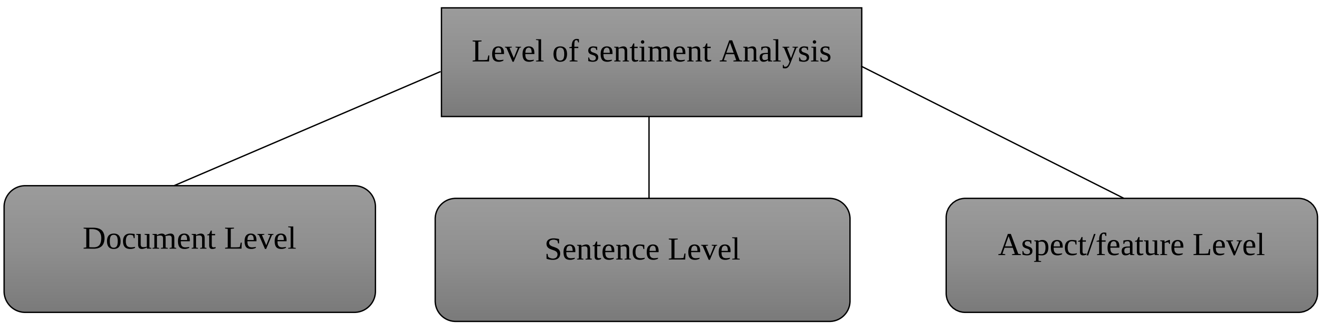

Figure 1: level of sentiment analysis

2.3.1. Document level Sentiment Analysis

Document level sentiment classification or document level sentiment analysis is the widely studied topic in the field of sentiment analysis especially in its early days [11]. The goal is to determine an opinion document (e.g., a product review) as expressing a positive or a negative opinion (or sentiment), which are known as sentiment orientations or polarities. The task at this level is to classify weather a whole document opinion expresses positive or negative sentiment [11]. For example given a product review, the system determine whether the review express an overall positive or negative opinion about the product. This task is known as document-level sentiment classification. Its limitations also motivated the fine-grained task of aspect-based sentiment analysis [29].

2.3.2. Sentence level Sentiment Analysis

Each sentence is considered as a short document, which can be subjective or objective. As author [30], argues subjective sentence is one that conveys opinion and expresses sentiment. Sentence level sentiment mining task is undertaken to classify reviews sentence in to two concepts subjective and objective[12]. Subjective is one that can be classified in to either positive or negative sentiment orientation. This can answer weather the question is the review sentence is positive or negative, while the objective ones are facts therefore, it can be categorized as neutral (i.e. no opinion) [21]. The goal of sentence level sentiment analysis is to identify sentiment or opinion for each individual sentence, instead of overall sentiment. In another word, it involves determining whether each sentence expresses a neutral, positive, or negative opinion. According to [12] the sentences are considered as short documents, which indicate as there is, no basic difference between document-level and sentence-level sentiment analysis.

2.3.3. Aspect level Sentiment Analysis

Both document level and sentence level analysis do not discover what people exactly liked and did not like[12]. Aspect level performs fine grained analysis. Aspect level, sometimes also called feature level studies [11]. The important task in this level is to express the sentiment polarity of a specific aspect by a given sentence. While both document level and sentence level is useful in various cases, they are unsatisfactory to providing the necessary details for an application, as they do not identify sentiment targets or assign opinions to these targets [12]. At the document level, a positive sentiment towards an object does not necessarily indicate that the author holds positive opinions on all aspects of the topic.

Beyond sentence-level analysis, classifying sentiment is often seen as an intermediate step, as it is more valuable to understand which features or entities the opinions are directed toward. For example “although the service is not that great, I still love this food”. Clearly has a positive tone, we cannot say that this sentences is entirely positive. In fact the sentence is positive about the food but negative about its service. In many applications, opinion targets are described by entities and their aspects[12] .

Aspect-Based Sentiment Analysis ABSA is a more advanced task compared to traditional text-level sentiment analysis[10]. It involves identifying specific attributes or aspects of an entity mentioned in a text along with the sentiment expressed towards each aspect. This allows for a more detailed understanding of opinions within the text.

2.4. Types of Opinions

There are two types of opinions, i.e., Regular opinion and comparative opinion[31].

Regular opinion

A regular opinion expresses a direct sentiment or evaluation toward a single entity, such as a product, person, or service. A regular opinion expresses a sentiment only on a particular entity or an aspect of the entity[12]. This type of opinion reflects an individual's positive, negative, or neutral feelings about the entity without comparing it to others. For example "cake tests very good," which express a positive sentiment on the aspect tastes of cake.

Comparative Opinion

A comparative opinion, on the other hand, involves comparing two or more entities and expressing a preference or ranking among them. A comparative opinion compares multiple entities based on some of their shared aspects[12]. For example "cake tests better than bread," which compare cake and bread based on their tastes or an aspect and express a preference of cake. In these cases, the opinion is based on a comparison between entities, showing which one is favored or deemed superior.

2.5. Aspect Based Sentiment Analysis task

The aspect based sentiment analysis contains two subtasks: aspect term extraction and aspect sentiment classification [12]. The task of aspect tem extraction can also been seen as an information extraction task, which aims to extract the aspects that opinions are on.

2.5.1. Aspect Term Extraction

Aspect Extraction (AE) is to determine all aspect terms that exist in each review sentence or comment. AE aim is to extract a specific aspect of a product toward which some sentiment is expressed in a review. It identifies aspects of the entity and more generally it can be seen as an information extraction task. Most of the researchers stated that an aspect can be expressed by a noun, verb, adverb and adjective. Out of which 60% - 70% of aspect terms are explicit nouns [29]. The basic approach of extracting aspects is finding frequent nouns or noun phrases, which are defined as aspects. Then the text containing aspects are classified as positive, negative or neutral[12].

Aspect expression that are noun and noun phrase are called explicit aspect expression[29].for example picture quality in “the picture quality of this camera is great” is an explicit aspect expression. Aspect expression that are not noun or noun phrase are called implicit aspect expression [29]. For example “expensive” is an implicit expression in “this camera is expensive”. It implies the aspect price. Many implicit aspect expression are adjectives and adverbs.

2.5.2. Aspect sentiment classification

The other tasks of aspect based sentiment analysis is aspect sentiment classification (ASC). Aspect sentiment classification is a key part of aspect-based sentiment analysis (ABSA), where the goal is to determine the sentiment expressed toward specific aspects or attributes of an entity, rather than the overall sentiment of the entire text. It involves two main tasks: identifying the aspect or feature being discussed and then classifying the sentiment (positive, negative, or neutral) associated with that specific aspect. ASC determine whether an opinion on an aspect is positive, negative or neutral or assign a numeric sentiment rating to the aspect[12].

2.6. Approaches of Sentiment Classification

Sentiment analysis techniques can be categorized in to three main approaches; lexicon based approach ,machine learning approach and hybrid approach[32]. Lexicon-based techniques were the first to be used for sentiment analysis. They are divided into two approaches: dictionary-based and corpus-based[33].In dictionary-based, sentiment classification is performed by using a dictionary of terms, such as those found in SentiWordNet and WordNet. Corpus-based sentiment analysis does not rely on a predefined dictionary but on statistical analysis of the contents of a collection of documents, using techniques based on k-nearest neighbors (k-NN)[34], conditional random field (CRF) [35], and others.

Machine Learning Approach can be further divided into supervised learning and unsupervised learning. Supervised Learning uses labeled data and includes methods such as rule-based classification, decision trees, linear classification, and probabilistic classifiers. Unsupervised Learning uses unlabeled data and involves techniques that do not rely on predefined labels[32].

Hybrid Approach [36] Combines elements from both machine learning and lexicon-based methods to utilize the strengths of each.

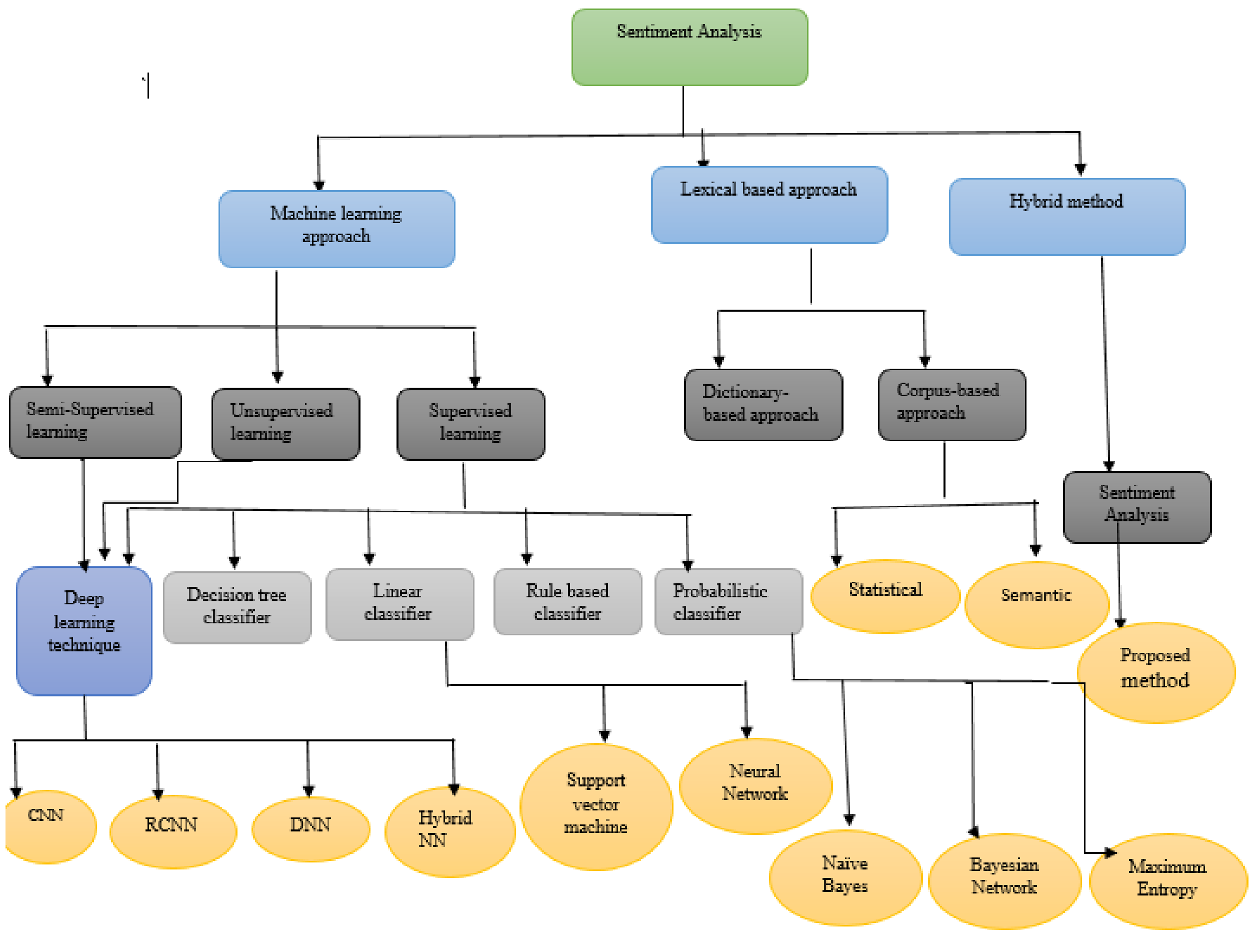

Figure 2: Approaches of Sentiment Classification (source: [20])

2.6.1. A machine learning approach

Machine learning is one part of artificial intelligence, contains many algorithms that aims at examining and evaluating machine behaviors[37]. For the task of sentiment analysis, the success of machine learning also relies on selection and extraction of sentiment features, which are especially from natural language processing (NLP) techniques. Machine learning can be divided into two categories: supervised learning and unsupervised learning. In supervised learning, the algorithm is trained on a large corpus of labeled data. Text can be converted into numerical vectors, allowing computers to process and train models using natural language processing (NLP) techniques. Once trained on labeled data, the model can automatically classify the sentiment of new or unseen text. Machine learning approaches are generally more accurate than rule-based methods because the model learns from vast amounts of text. With a large dataset, the algorithm is exposed to a wide range of patterns, enabling it to classify sentiment accurately.

Unsupervised methods use large datasets to extract numerous mentions, but linking these mentions to specific entities or aspects required for a knowledge base can be difficult. While these methods can effectively identify mentions, the absence of labeled data makes it challenging to ensure that the extracted mentions are relevant or properly aligned with the intended knowledge base. Since these methods can't predict what mentions will be found, human judgment is often needed to determine if a mention is meaningful. The third type of machine learning approach is the semi-supervised approach. This method aims to label the unlabeled samples in a dataset, primarily addressing the challenge of learning from limited labeled data. By doing so, it enhances the model's ability to generalize on the labeled samples. This approach also reduces the reliance on human involvement and the need for extensive manual annotation of the corpus, making the process more efficient.

2.6.1.1. Artificial neural network (ANN)

Artificial Neural Network (ANN) is widely used in the field of machine learning to solve problems such as predictive modeling, classification and function approximation [37].An Artificial Neural Network (ANN) is a collection of interconnected nodes, known as artificial neurons (AN) or perceptrons, which serve as the fundamental units of the network. These neurons are linked by weights, and a simple structure of an artificial neuron is shown in Figure 3. Each neuron receives an input, denoted as *x*, which is associated with a weight *w* and a bias term *b*. The net input is typically calculated as a weighted sum (Equation 2.1), after which an activation function is applied to determine whether the neuron should "fire." The weights either amplify or diminish the input signal. The process within an artificial neuron can be viewed as a nonlinear mapping from an N-dimensional space, *R^N*, to an output, typically ranging between 0 and 1 or -1 and 1[37]. The activation function determines the output range and controls whether the neuron activates or not. The bias term enables the activation function to shift left or right along the x-axis. It is linked to a weight and comes with an additional bias value, allowing for more flexibility in adjusting the model's output [37] as equation 2.2.

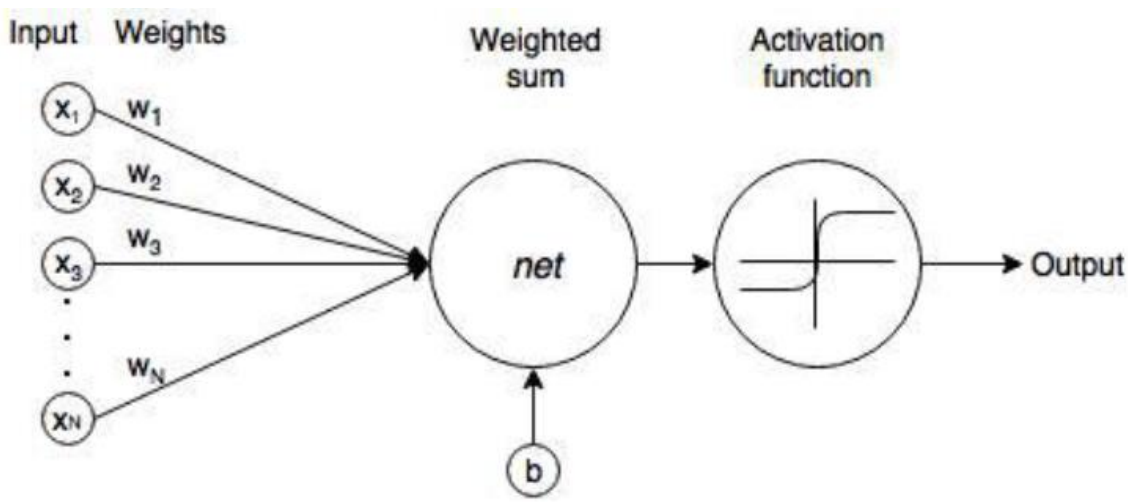

Figure 3 : Artificial neural network (source:[37])

$$net = \sum_{i=1}^{N}(XiWI) + b \text{ where } b \text{ is as equation bellow} \quad (1)$$

$$b = w0*biasvalue, \text{ where biasvalue is usually equal to 1} \quad (2)$$

2.6.2. Deep learning Approach

Deep learning approaches are able to automatically capture, to some extent, the syntactic and semantic feature from text without feature engineering, which is labor intensive and time consuming[38]. Recently, they attract interest of research and achieve state-of-the-art performances in different fields of NLP, including sentiment classification. Deep learning adapts a multilayer approach to the hidden layers of the neural network. In traditional machine learning approaches, features are defined and extracted either manually or by making use of feature selection methods [20].

Deep learning has become highly effective machine learning technique that learns multiple layers representations or features of the data and produces state-of-the-art prediction results. In recent years, deep learning has gained prominence as a powerful computational approach that can automatically discover complex semantic representations of text from data, eliminating the need for manual feature engineering[38]. These approaches have improved the state-of-art in sentiment analysis task including sentiment classification of sentences/documents, sentiment extraction and sentiment lexicon learning. The neural network model can automatically learn the low-dimensional representation of reviews without relying on artificial feature engineering. This feature allows neural networks to be used for aspect-level sentiment analysis tasks and has attracted the attention of researchers [39].

2.6.2.1. Deep Neural Networks (DNN)

A deep neural network [40] is a neural network with more than two layers, some of which are hidden layers. Deep neural networks use sophisticated mathematical modeling to process data in many different ways. A neural network is an adjustable model of outputs as functions of inputs, which consists of several layers: an input layer, including input data; hidden layers, including processing nodes called neurons; and an output layer, including one or several neurons, whose outputs are the network outputs[20].

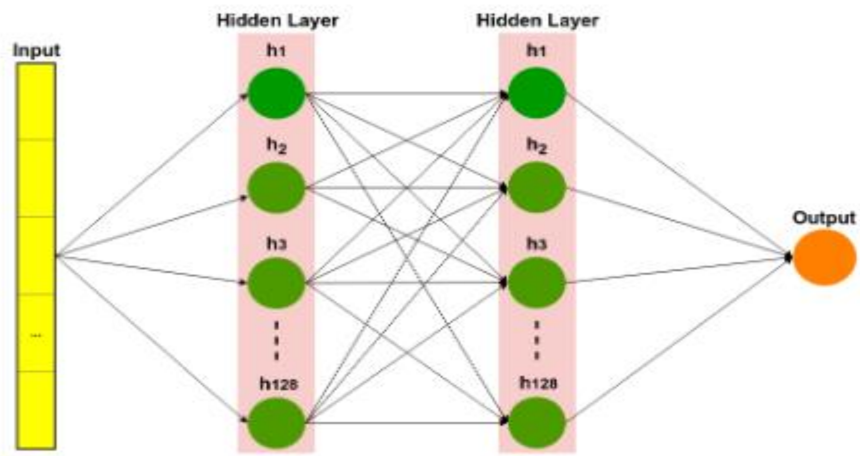

Figure 4: Deep neural network (source:[20]).

2.6.2.2. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) Convolutional neural networks are specialized kinds of neural networks designed to process and analyze data. CNN is a class of deep learning with invariance built into a multi-layer network structure which is more suitable for the recognition of data [35]. It is a deep neural network architecture [41] typically composed of convolutional and pooling or subsampling layers to provide inputs to a fully-connected classification layer. Convolution layers filter their inputs to extract features; the outputs of multiple filters can be combined. Pooling or subsampling layers reduce the resolution of features, which can increase the CNN's robustness to noise and distortion[20]. Fully connected layers perform classification tasks. The input data was preprocessed to reshape it for the embedding matrix.

The figure below shows an input embedding matrix processed by four convolution layers and two max pooling layers. The first two convolution layers have 64 and 32 filters, which are used to train different features; these are followed by a max pooling layer, which is used to reduce the complexity of the output and to prevent the overfitting of the data. The third and fourth convolution layers have16 and 8 filters, respectively, which are also followed by a max pooling layer. The final layer is a fully connected layer that will reduce the vector of height 8 to an output vector of one, given that there are two classes to be predicted (Positive, Negative).

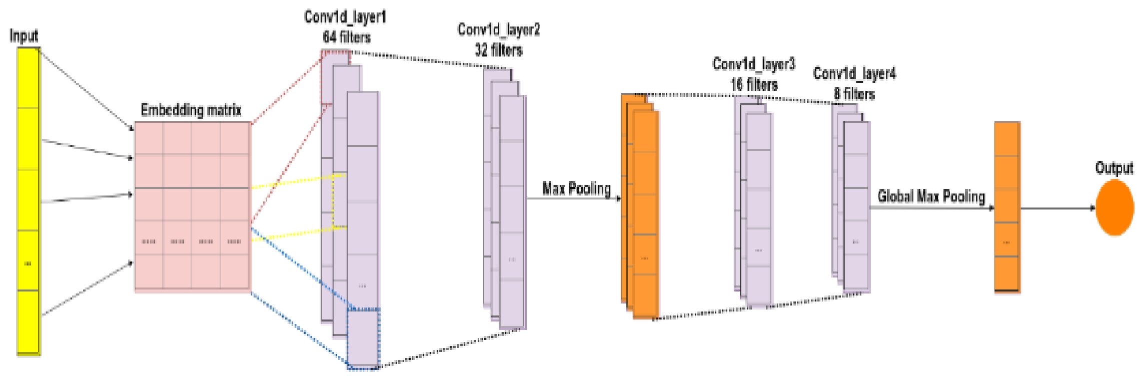

Figure 5: a convolutional neural network (source [20]).

2.6.2.3. Recurrent neural network (RNN)

Recurrent neural network [42] are a class of neural networks whose connections between neurons form a directed cycle, which creates feedback loops within the RNN. The main function of RNN is the processing of sequential information on the basis of the internal memory captured by the directed cycles. Unlike traditional neural networks, RNN can remember the previous computation of information and can reuse it by applying it to the next element in the sequence of inputs.

2.6.2.4. Long short term memory (LSTM)

Long short term memory is special types of recurrent neural network that is capable of long short term memory as the input of activation functions in the hidden layer [20]. To reshape data for embedding matrix, the input data is pre-processed in similar way of CNN. The LSTM layer contains 200 cells and fully connected layer, which is final layer contains 128 cells for text classification. The last layer uses the sigmoid activation function to reduce the vector of height 128 to an output vector of one, given that there are two classes to be predicted (positive, negative)[20].

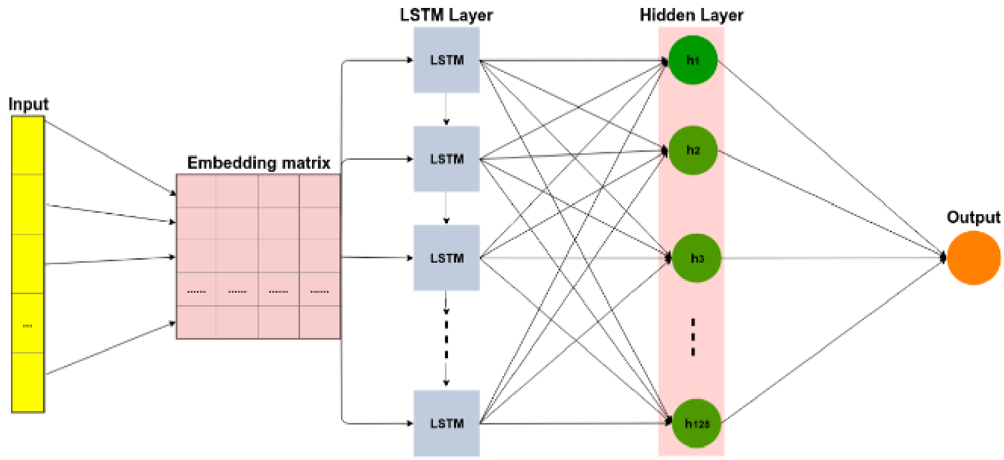

Figure 6: long short term memory

There are gates in LSTM namely an input gate, output gate and a forget gate. This three gates are try to remember when information in memory and how much the information in the memory should be updated during each iteration.

Input gate

This gate updates the cell state by applying an activation function to the previous hidden state and the current input. The activation function determines which values should be updated and identifies the important information for the next step.

Output gate

The output gate is the last gate that used to determines the next hidden state, which holds information from pervious inputs. The previous hidden state and current input are passed through a sigmoid function. The updated cell state is then processed through a tanh function, and the outputs of the tanh and sigmoid functions are multiplied to decide what information the hidden state should retain.

Forget get

Forget get is used to determine the forgotten information from previous state and decide next hidden state.

For single direction the LSTM cell consists of several gates (forget, input, and output gates) that control the flow of information. For each word embedding x_t , hidden state h_t and cell state c_t follows:

Forget gate: Decides what information to discard:

$$f_t = \sigma(Wf \cdot [h_t - 1, x_t] + b_f)$$

Input gate: Decides what new information to store:

$$i_t = \sigma(Wi \cdot [h_t - 1, x_t] + b_i)$$

$$c\tilde{}_t = tanh(W_c \cdot [h_t - 1, x_t] + b_c)$$

$$c_t = f_t . c_t - 1 + i_t . c\tilde{}_t$$

Output gate: Decides the next hidden state:

$$o_t = \sigma(W_o \cdot [h_t - 1, x_t] + b_o)$$

$$h_t = o_t \cdot tanh(c_t)$$

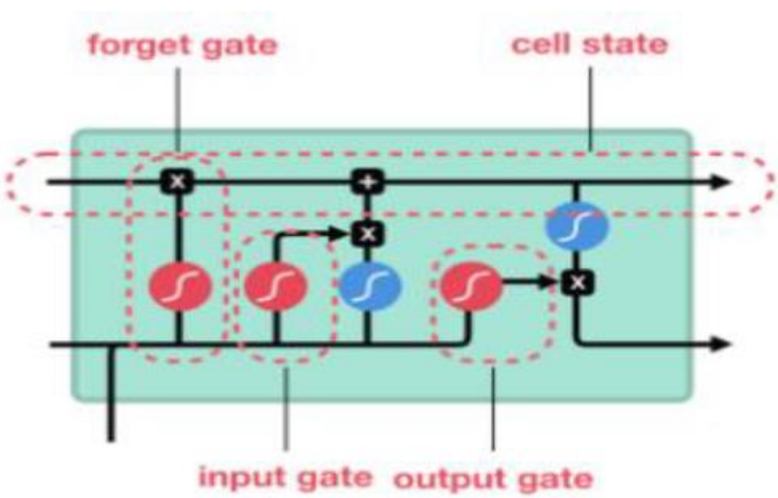

Figure 7: Structure of LSTM

BiLSTM is enhancement of LSTM that process the sequence of data in forward and backward directions. Having bidirectional process enable model to capture context from both past and future tokens. This feature making it effective for tasks like aspect extraction where understanding the context around a word is crucial. BiLSTM bidirectional nature helps capture context from both directions, leading to more accurate extraction. BiLSTM Consists of two LSTM layers: one processes the input sequence in the forward directions left to right while the backward LSTM it in reverse right to left. The outputs from both directions are the combined by concatenation.

$$h_t^{\text{BiLSTM}} = [h_t^{\text{forward}}, h_t^{\text{backward}}]$$

2.6.2.5. Bidirectional Encoder Representation from Transformers (BERT)

BERT stands for Bidirectional Encoder Representations from Transformers. It is a pre-trained language model that is designed to consider the context of a word from both left and right side simultaneously [19]. It improves results at several NLP tasks such as sentiment analysis and question and answering systems. BERT can extract more context features from a sequence compared to training left and right separately. BERT is the first NLP technique to rely solely on self-attention mechanism, which is made possible by the bidirectional Transformers at the center of BERT's design. It is also a neural network based technique for natural language processing task [43].

BERT model is very helpful to the various natural language processing application. So, it uses a special tokenization called WordPiece which can split words to sub-words [23]. This language model is a pre-training deep transformer model that uses encoder from transformer and it is trained based on a self-supervised approach [19]. BERT makes use of transformer, an attention mechanism that learns contextual relations between words (or sub-words) in a text. The encoder consists of 12/24 layers. Those each layer has two sublayers: multi-head self-attention mechanism and feed forward neural network. Architecture layers followed by normalization.

There are two steps in our BERT framework: pre-training and fine-tuning [19]. During pre-training, the model is trained on unlabeled data over different pre-training tasks. For fine-tuning, the BERT model is first initialized with the pre-trained parameters, and all of the parameters are fine-tuned using labeled data from the downstream tasks. Each downstream task has separate fine-tuned models, even though they are initialized with the same pre-trained parameters [19].

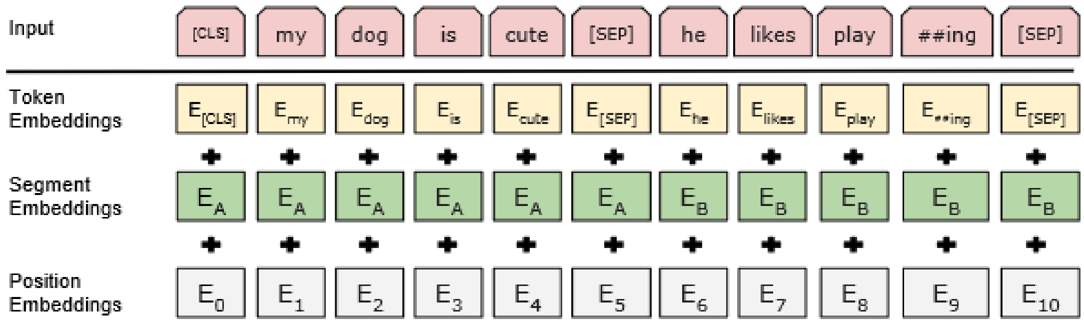

Figure 8: BERT input representation

Masked language model MLM

The masked language model randomly masks some of the tokens from the input, and the objective is to predict the original vocabulary id of the masked word based only on its context[19]. Unlike left-to-right language model, pre-training the MLM objective enables the representation to fuse the left and the right context, which allows us to pre-train a deep bidirectional Transformer[19]. The model randomly masks 15% of the words in the input then run the entire masked sentence through the model and has to predict the masked words[19].

Next sentence prediction (NSP)

NSP (Next Sentence Prediction) is used to help BERT learn about relationships between sentences by predicting if a given sentence follows the previous sentence or not [44]. In training, 50% correct sentence pairs are mixed in with 50% random sentence pairs to help BERT increase next sentence prediction accuracy. The basic assumption of this idea is that the given random text will be separated logically from the first sentence. In order to overcome this problem BERT use NSP [44].

So as in general the main aim of the NSP training process is that whether the given two sentences has a logical, sequential connection or not[44].

Bert fine-tuning

Fine-tuning is straightforward since the self-attention mechanism in the Transformer allows BERT to model many downstream tasks-whether they involve single text or text pairs by swapping out the appropriate inputs and outputs [19]. BERT uses two training strategies: Masked Language Model and Next Sentence Prediction[19]. BERT is the first fine-tuning based representation model that achieves state-of-the-art performance on a large suite of sentence-level and token-level tasks, outperforming many task-specific architectures.

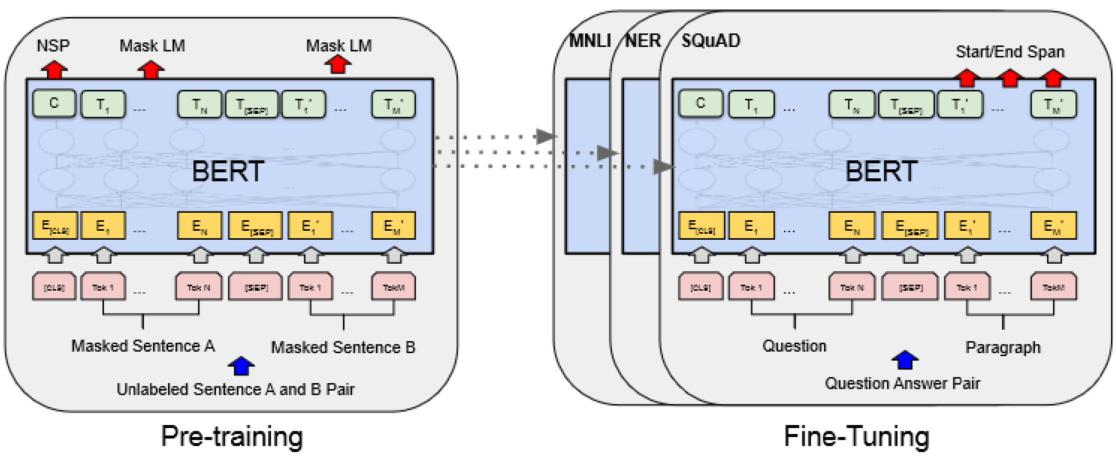

Figure 9: Overall pre-training and fine-tuning procedures for BERT

Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token [19]. The Transformer architecture makes it possible to parallelize ML training extremely efficiently. Massive parallelization thus makes it feasible to train BERT on large amounts of data in a relatively short period of time [44]. Transformers use an attention mechanism to observe relationships between words.

BERT uses a multi-layer bidirectional Transformer to process the input embeddings. Each layer consists of Multi-Head Attention and Feedforward Neural Network. Multi-Head Attention computes attention scores between all token pairs, which capture the relationships between words in a sentence. Feedforward Neural Network**:** Applies a fully connected layer and non-linear activation to each attention output. Each attention head computes the attention scores for each token using the dot-product attention mechanism:

$$\textbf{Attention}(Q, K, V) = \textbf{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where:

Q is the query matrix, K is the key matrix, and Vis the value matrix for the tokens.

d_k is the dimension of the key vectors.

For classification tasks (like sentiment analysis), BERT typically uses the output from the special [CLS] token. After the final transformer layer, the hidden state corresponding to the [CLS] token is passed to a fully connected layer with a softmax function to produce probabilities for each class (e.g., positive, negative, neutral).The output vector from the [CLS] token is passed to a linear classifier:

$$y\text{^}=\text{softmax}(W_{cls}\, h_{cls} + b)$$

Where:

h_{cls} is the hidden state of the [CLS] token from the last transformer layer.

W_{cls} and b are learnable parameters of the classifier.

The softmax function converts the logits into class probabilities.

Softmax function

Softmax is commonly used in neural network models as the activation function for the output layer when dealing with multi-class classification problems. It is particularly useful in tasks like sentiment analysis, where the goal is to classify a text into different sentiment categories (e.g., positive, negative, neutral). The softmax function converts a vector of raw output values (logits) from the network into a probability distribution, where the sum of the probabilities equals 1. Each element in the output represents the probability that the input text belongs to a particular class (e.g., positive, negative, or neutral sentiment). The softmax activation function is essential for multi-class classification tasks in machine learning, providing a probabilistic interpretation of model outputs. Its ability to transform logits into a probability distribution makes it an effective tool for applications such as sentiment analysis, image classification, and more.

2.7. Afaan Oromoo

Language is the ability to acquire and use complex system of communication. It is the human system of communication that uses arbitrary signals, such as voice, sound, gesture and/or written symbols that enable humans to express their feeling, sentiment, thought, idea and experience [45].

Afaan Oromoo is an Afro-asiatic language that belongs to the Cushitic branch. It is native to the Ethiopian state of Oromia and Northern Kenya and is spoken predominantly by the Oromoo people and neighboring ethnic groups in the Horn of Africa [46]. With more than 36 million speaker making up 33.8% of the total Ethiopian population, Oromoo has the largest number of native speakers in Ethiopia, and ranks as the second most widely spoken language in Ethiopia [47] . Afaan Oromoo serves as one of the official working languages of Ethiopia and is also the working language of several of the states within the Ethiopian federal system including Oromia, Harari and Dire Dhawa regional states and of the Oromia Zone in the Amhara Region. However, it is one of the least researched and under resourced language as there is a few computational linguistic related works where done on language.

2.7.1. Afaan Oromoo writing System

With regard to the writing system, since 1991 Latin alphabet called Qubee has been adopted and become the official script of Afaan Oromoo [48]. The Qubee writing system has 33 letters that consists of all the 26 English letters with an addition of 7 combined consonant letters which are known as or “Qubee Dachaa”. These include ch, dh, sh, ny, ts, ph and zy. All the vowels in English are also vowels in “Qubee”. These are ‘a’, ‘e’, ‘i’, ‘o’, and ‘u’. Vowels have two natures in the language, and they can result indifferent meaning. The natures are short and long vowels. A vowel is said to be short if it is one and long if it is two, which is the maximum.

According to [49], Afaan Oromoo is a phonetic language, which means that it is spoken in the way it is written. Unlike English or other Latin based languages there are no skipped or unpronounced sounds/alphabets in the language. Every alphabet is pronounced in a clear short/quick or long/stretched sounds.

Types	Alphabets
Vowels	a,e,i,o,u
Consonants	b,c,d,f,g,h,j,k,l,m,n,p,q,r,s,t,v,w,x,y,z
Digraphs/double consonant	ch, dh, ny, ph, sh,ts,zh.

Table 1: Afaan Oromoo alphabets

2.7.2. Afaan Oromoo Punctuation marks

Analysis of Afaan Oromoo texts reveals that different punctuation marks follow the same punctuation pattern used in English and other languages that follow Latin Writing System [50]. The most usually used punctuation marks in Afaan Oromoo language is as follow:

Full stop “Tuqaa” (.), used at the end of a sentence and in abbreviations. Question mark “Mallattoo Gaafii” (?), used in interrogative or at the end of a direct question. Exclamation mark“Rajeffannoo” (!), used at the end of command and exclamatory sentences. Comma “Qoodduu” (,): is used to separate listing in a sentence or to separate the elements in a series. Colon “Tuqlamee colon” (:), the function of the colon is to separate and introduce lists, clauses, and quotations, along with several conventional uses, and etc. Unlike English language apostrophe (’) is not punctuation mark in Afaan Oromoo, rather it is part of words. For example, ji’a (month), sa’a (cow) etc.

2.7.3. Afaan Oromoo Morphology

Afaan Oromoo is very rich in morphology like many local and African languages. Afaan Oromoo verbs are highly inflected for gender, person, number and tenses[51]. Both Afaan Oromoo nouns and adjectives are highly inflected for number and gender. [51]. Words can be formed from morphemes in two ways: Derivational Morphology and Inflectional Morphology. Derivational Morphology is concerned with the way words are derived from morphemes through processes such as affixation or compounding while inflectional morphology deals with the combination of a word with a morpheme. Afaan Oromoo words have some prefixes and infixes, but suffixes are the predominant morphological features in the language. Almost all Afaan Oromoo nouns in a given text have person, number, gender and possession markers, which are concatenated and affixed to a stem or singular noun form.

In addition, Afaan Oromoo noun plural markers or forms can have several alternatives. For instance, in comparison to the English noun plural marker, s (-es), there are more than ten major and very common plural markers in Afaan Oromoo including: -oota, -oolii, -wwan, -lee, an, een, -eeyyii, -oo, etc.). For instance, mana (house) can take the following different plural forms: manoota (mana +oota), manneen (mana + een), manawwan (mana + wwan). From the noun, nama [man] the following word unit of measurement generated through inflection and affixation. These are Namicha (the man) Nama+ [icha], Namoota [men] Nama+ [oota]. The construction and usages of such alternative affixes and attachments are governed by the morphological and syntactic rules of the language [52].

Afaan Oromoo has two grammatical genders, masculine and feminine, and all nouns belong to either one or the other. Frequent gender markers in Afaan Oromoo include -eessa/-eettii, -a/-ttii or –aa/tuu. Afaan Oromoo displays singular and plural number, but nouns that refer to multiple entities are not obligatorily plural: Nama 'man' namoota 'people', Nama Shan 'five men' namoota Shan 'five people'. Another way of looking at this is to treat the "singular" form as unspecified for number. When it is important to make the plurality of a referent clear, the plural form of a noun is used. Noun plurals are formed through the addition of suffixes

Afaan Oromoo words	Morphology	Gender
Sawwan(cows)	Sa+wwan	Female
Obbolettii(sister)	Obbol+ettii	Female
sangoota(oxes)	Sangaa+oota	Male
Jaarsota(elders)	Jaars+ota	Male

Table 2: Afaan Oromoo morphology

2.7.4. Afaan Oromoo word classes

Nouns

Afaan Oromoo nouns and adjectives are marked for masculine or feminine gender. Nouns have an essential masculine or feminine gender that cannot be determined by the form of the noun, with a few exceptions when biological gender is associated with a particular suffix, such as eessa for masculine and -eettii for feminine nouns, e.g., obboleessa 'brother' and obboleetti 'sister'. Adjectives agree with the nouns they modify in gender. All nouns and adjectives are marked for number: singular and plural, e.g., for masculine nouns nama 'man – namicha 'the Man'; for feminine nouns biyya 'country' – biyyatti 'the country'. All nouns are marked for case.

Nouns and noun phrases play a crucial role in aspect-based sentiment analysis (ABSA) because they often represent the "aspects" or "targets" of sentiment within a text. In sentiment analysis, aspects typically refer to specific components or attributes of a product, service, or topic that are being evaluated. Nouns (e.g., "buna/coffee," "gatii/price," "tajaajila/service") are often the aspect terms that indicate what exactly the sentiment is about.

Nouns directly refer to the objects or entities being discussed, making them primary candidates for aspect terms. For example, in the sentence "The coffee flavor is excellent," "coffee flavor" is the aspect being described.

Adjectives

In Afaan Oromoo, Adjectives are words that describe or modify another person or thing in the sentence [51].Adjectives are very important in Afaan Oromoo because its structure is used in every day conversation. Adjectives are usually placed after the noun in Afaan Oromoo. For instance, in boontuun konkolaataa Adii bitte “Bontu bought a white car” the adjective comes after the noun konkolaata. Adjectives are also play the big role in sentiment analysis as it implies the opinion. Adjectives like good, bad, expensive, etc. are widely used in sentiment analysis.

Verbs

Afaan Oromoo verb consists minimally of a stem, representing the lexical meaning of the verb, and a suffix, representing tense or aspect and subject agreement. For example, in dhufne 'we came', dhuf- is the stem ('come') and -ne indicates that the tense is past and that the subject of the verb is first person plural. Afaan Oromoo makes a basic two-way distinction in its verb system; between the two tensed forms past (or "perfect") and present (or "imperfect" or "non-past"). Each of these has its own set of tense/agreement suffixes. There is a third conjugation based on the present, which has three functions: it is used in place of the present in subordinate clauses, for the jussive ('let me/us/him, etc. V', together with the particle haa), and for the negative of the present (together with the particle hin). For example, deemne 'we went', deemna 'we go', akka deemnu 'that we go', haa deemnu 'let's go', and hin deemnu 'we don't go'.

Conjunctions

Conjunction is a word that can be used to join or connect two phrases, clauses and sentences. Conjunctions can be divided into coordinating and subordinating conjunctions. Coordinating conjunctions are used to connect two independent clauses. Mostly these conjunctions are used when the speaker needs to lay emphasis on the two sentences equally. Some of these conjunctions in Afaan Oromoo include: garuu „but‟, moo„ or‟, kanaafuu„ therefore‟, haata’u malee„however/so‟, tu’ullee„even though‟ etc. Consider the following example: laptoppiin koo qulqullinna iskiriinii gaarii qaba garuu turtiin baatirii isaa gababaadha. This means my laptop screen quality is good but battery life is short.‟garuu‟ in this sentence is coordinating conjunction.

Afaan Oromoo subordinating conjunctions include yoo as if, akka waan as if, wayta/yeenna as when, hamma as until, booda as after‟, dursa as before‟ etc. The following example illustrates the above case: Akka waan nabeektuuu fakkeessite. This means, “She acts as is if she knows me”. Akka waanin this sentence is used as subordinating conjunction. It joins one subordinating clause that is waan nabeektu ‟As she knows me‟ and fakkeessitee ‟She acts‟.

2.8. Afaan Oromoo Sentiment Analysis challenges

There are various challenges that hinder automated methods of understanding, analyzing, and classifying Afan Oromoo text into different sentiment polarities. One of the major challenge of language is lack of available resource for sentiment analysis. There are no resource such as preparation, lexicon database, part of speech tagging used to identify the noun, noun phrase, adjective, verb and adverb, used to detect aspect and polarity of sentiment. Lack of lexicons or morphological analyzers specific to Afaan Oromoo, can hinder the process of identifying and extracting aspect terms. Similarly, there is no organized dataset or corpus available for sentiment analysis. In addition there is no annotated datasets for aspect terms related to product in Afaan Oromoo. This scarcity can make it difficult to train models effectively, leading to challenges in accurately extracting aspect terms.

For impoverished languages such as Afaan Oromoo, there are two fundamental impediments to growth in language processing. To begin with, the diversity of the languages may need the invention of new strategies. Second, the limited availability of existing resources and tools makes building and testing new ones more complex and time-consuming. Afaan Oromoo does not have a standard corpus. For Natural Language Processing experiments, the availability of labeled language resources, such as annotated corpora and domain-specific labeled language resources is crucial. Due to a shortage of resources, manual verification and annotation of electronic text content is typically a requirement for the development of NLP systems.

Afaan Oromoo serves as one of the official working languages of Ethiopia and is also the working language of several of the states within the Ethiopian federal system including Oromia, Harari and Dire Dhawa regional states. This language needs improvement to process various document. To the best of our knowledge, aspect level sentiment analysis for the Afaan Oromoo language using bidirectional encoder representation from transformer is not done so far. In this Thesis, the aspect level sentiment analysis model works based on Afaan Oromoo sentence structure.

Sentiment analysis and Afaan Oromo Sentence Structure

Afan Oromo is a subject-object-verb (SOV) language, meaning the verb typically comes at the end of the sentence. For instance: Afaan Oromoo: *"Mana nyaataa kun nyaata gaarii dhiheessa."* English: "This restaurant serves good food."

Aspect Placement: Aspects (e.g., *nyaata* - food) often appear before the verb (*dhiheessa* - serves).

Sentiment expressions are often tied to adjectives (*gaarii* - good) that precede the verb or aspect. Sentiment-bearing words like adjectives or adverbs often occur away from the verb, requiring models to consider longer-range dependencies.

Linguistic Features and Their Impact

Adjective-Noun Agreement: In Afan Oromo, adjectives agree with nouns in number, gender, and case. For instance, "Nyaata gaarii" *(Good food) vs.* "Nyaatoowwan gaarii" *(Good foods).*

Polarity Modifiers: Words like *hin*, and *miti*, are used to negate or affirm sentiments.

Emphasis and Repetition: Afaan Oromo uses repetition for emphasis, which can amplify sentiment. For instance "Nyaanni isaanii bayyee,bayyee gaarii dha!" *(Their food is very, very good!).* Repetition needs to be detected and interpreted correctly to amplify the sentiment score.

2.9. Related Work.

Many related research were done so far on sentiment Analysis. Among them, we have selected the most important research for different languages, which are related to our work. The approach to sentiment analysis has been mostly follows machine either learning approach or lexicon-based approach. In recent years, researchers proposed Aspect level sentiment analysis for English and other languages using BERT model. This work also uses this technique to obtain prominent experiment result.

2.9.1. International works done on sentiment Analysis

The work in [10] investigated aspect- level sentiment analysis approach via BERT and aspect feature location model. To achieve objective of their work, first they used BERT model to mine aspect-level auxiliary information from comment context. They construct an aspect-based sentiment feature extraction method to understand the expression features of aspect words and the interactive. Finally, they construct evaluation experiments on three benchmark datasets. For experiments, they conduct experiments on three public English review datasets, and these datasets are among restaurant and laptop datasets provided by SemEval 2014. In order to verify the effectiveness of the model, the researcher compare the ALM-BERT approach with different popular aspect-based sentiment analysis model. As the report of researcher, the experimental results of aspect-level sentiment analysis performance of the ALM-BERT significantly better than other comparison methods.

The Author in [53] conduct study on Aspect Term Extraction using various text embeddings methods. They focused on long short-term memory (LSTM) with conditional random field (CRF) enhancement using different pre-trained word embeddings. They also analyzed the influence on the performance of extending the word vectorization step with character based word embeddings. They used 11 different pre-trained word embeddings and evaluated a total of 88 combinations of models and text embeddings in the entire experiment According to researcher report, experimental results on SemEval datasets revealed that bi-directional long short-term memory (BiLSTM) could be used as a very good predictor, comparing to very sophisticated and complex models using huge word embeddings.

The work in[54] proposed Aspect Term Extraction and Term Polarity Classification System using Conditional Random Field (CRF) based classifier for Aspect Term Extraction (ATE) and a linear classifier for Aspect Term Polarity Classification (ATP). For the ATE subtask, they implement a variety of lexicon, syntactic and semantic features and cluster features induced from unlabeled data. To achieve the proposed objective the researcher built aspect-sentiment word pair lexicon from the training set where this lexicon contains 9073 word pairs for the laptop domain and 22171 word pairs for the restaurant domain. The researcher report that they achieves state-of-the-art performances in ATE, ranking 1st (among 28submissions) and 2rd (among 27 submissions) for the restaurant and laptop domain respectively.

The authors [55] proposed Arabic aspect-based sentiment analysis using the deep learning technique with a pre-trained BERT model. The researcher use n-gram feature selection method with pre-trained BERT model. The researcher has used 8320 hotels comment dataset from hotel website for experiment. According to the researcher report, BERT model has outperformed the state-of-the-art.

Author [56] proposed that a BERT -Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text. The algorithm first utilizes the BERT structure to extract sentence-level and aspect-level representation vectors, and enhances local feature extraction through an improved convolutional neural network (CNN). It combines aspect-level and sentence-level corpora to form a sequence sentence pair. Next, the algorithm employs a domain adversarial neural network (DANN) to make the feature representations extracted from different domains as indistinguishable as possible, meaning that features extracted from both the source and target domains become more similar. Finally, by training a sentiment classifier on a source domain dataset with sentiment labels, the algorithm aims to achieve effective sentiment classification in both the source and target domains, handling both sentence-level and aspect-level sentiment classification tasks.

Another study by authors [57] was proposed a combined approach which aims at mining opinions from Arabic documents. They used three methods at sequence in their approach: First, lexicon-based method is used to classify documents. The classified documents used as training set for maximum entropy model which subsequently classified some other documents. After that, k-nearest model is used to classify the rest of the documents. They have done experiments with 1143 posts containing 8793 Arabic Statements. Their system achieved an accuracy of 80.29%. The accuracy almost went from 50% using one method, 60% using two method and 80% using three methods. They also claimed that the experimental results further show recall and precision of positive documents are better than the negative one.

2.9.2. Local works done on sentiment Analysis

Sentiment Analysis done for Afaan Oromoo Language

Many researcher conduct study on sentiment analysis for Afaan Oromoo language using different approach such as rule based approach, machine learning Approach and deep learning approach. These studies was conducted at character level, sentence level, document level and aspect/feature level.

The author of [21] conducted the first study on sentiment mining and aspect-based summarization of Afaan Oromoo news text. The study was carried out based on rule-based approach and follows empirical research design. The researcher has collected the total of 400 dataset from Oromia Radio and Television Organization News. According to the author's report, method shows effective results of performance of 90 % precision and 87.1 % recall in for positive class, 87% precision, 89.7% recall for negative class and achieve 88.3% system accuracy.

The study in [58] was proposed on sentiment analysis model for Afaan Oromoo short message service Text. The study was carried out based on machine learning approach. The study was attempted to design a two-step approach for Afaan Oromoo text sentiment classification model, clustering followed by classification algorithms. A total of 1597, data was collected from Oromia broadcasting corporate (OBN) "8331 SMS database" from three domains (i.e. news, entertainment and general service domain is used to conduct the experiment. Three supervised learning algorithms, including Naïve Bayes (NB), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) in each domain were used to classify the sentiment of short Afaan Oromoo text.

According to the report of researcher, the result shows that SVM out performs NB and KNN with an accuracy of 91.66%, 93.76% and 92.87% for news, entertainment and general service domain respectively.

The authors in [22] conducted on sentiment analysis for Afaan Oromoo Facebook media using deep learning approach . The study was carried out based on combined convolutional neural network and bidirectional long short-term memory. The study was conducted based on character level sentiment analysis. The total of 24000 dataset are collected from two domains namely Facebook and Twitter for the experiment. After the data cleaning process is applied, different annotators annotated data manually into five class namely, 2 ,1, -2, -1, and 0 which represent very positive, positive, very negative, negative and neutral respectively. Experiments were performed on the prepared corpus from Facebook and Twitter by applying Convolutional Neural Network, Bidirectional Long Short-Term Memory and combined Convolutional Neural Network-Bidirectional Long Short-Term Memory with character level word embedding. The experimental result validate that the proposed model outperforms both CNN and Bi-LSTM in both Facebook and Twitter dataset. Based on the implemented Facebook dataset the researcher achieved a performance accuracy of 93.3%, 91.4%, and 94.1% for CNN, Bi-LSTM and CNN-Bi-LSTM respectively. For twitter dataset, they achieved 92.6%, 90.3%, 93.8% for CNN, Bi-LSTM and CNN-Bi-LSTM respectively.

Another study by author [24] conducted on sentiment analysis for Afaan Oromoo using machine learning approach. The study was undertaken on the document level sentiment analysis. The researcher has conducted experiment on Naïve Bayesian, Long short-term memory and convolutional neural network algorithm. According to the report of researcher shows they get 90.7%, 71.1%, 54.6% for unigram, bigram and trigram result for Naïve Bayesian, and they get 92% and 93.6% for CNN and LSTM respectively.

The author of [37] conducted on sentiment analysis for Afan Oromoo Socio-politics Text. The study was undertaken at document level sentiment analysis using deep learning approach. The researcher collect data from FBC official Facebook page of socio-politics domain of Afaan Oromoo news. The total of used 12662 comments with 8675 common vocabulary file for training and testing purpose. As a result accuracy 79.99 % achieved.

The study in [59] was focused on Multi-Class Sentiment Analysis from Afaan Oromoo Text Based on Supervised Machine Learning Approaches. The study was focus basically on sentence level sentiment analysis with five multiple classes- very negative, negative, neutral, positive and very positive. They proposed two supervised machine learning approaches—Support Vector Machine (SVM) and Random Forest algorithms—to classify sentiment polarity from Oromia Broadcasting Network (OBN) Twitter data, using tf-idf for feature extraction. According to the researchers' report, the Support Vector Machine and Random Forest approaches achieved accuracies of 90% and 89%, respectively, on the OBN Twitter dataset with a corpus size of 1,810.

The work in [60] investigate Aspect-Based Sentiment Analysis for Afaan Oromoo Movie reviews. The research. The research is conducted at aspect level using machine learning. The study was follows experimental research design. The researcher collected total of 2800 dataset from Afaan Oromoo YouTube for training and testing purpose. The researcher applied bag of words (BoWs) and TF-IDF for feature extraction. The researcher used for machine learning model namely random forest, logistic regression, support vector machine and multinomial naïve Bayes. According the researcher report, the experimental result obtained that random forest algorithm with both bag of words (BoWs) and TF-IDF produced accuracy of 88%, the logistic regression achieved 87% accuracy with both bag of words (BoWs) and TF-IDF, the accuracy obtained by the SVM with BoW was 88% and 87% with TF-IDF. Multinomial naïve Bayes (MNB) obtained 88% with both BoW and TF-IDF feature extraction techniques.

Sentiment Analysis done for Amharic Language

Different authors was proposed on sentiment analysis for Amharic language using different approach. the author of [61] has studied sentiment-mining Model for opinionated Amharic texts. The study employs a sentiment and subjective lexicon of terms to classify reviews based on the frequency of positive and negative terms present in the text. This rule-based classifier determines sentiment as follows: if the number of positive terms exceeds the number of negative terms, the review is classified as positive; if the number of negative terms surpasses the number of positive terms, it is classified as negative. If the counts of positive and negative terms are equal, the review is considered neutral.

The study in [62] proposed Opinion Mining from Amharic Blog. They used manually crafted rules and lexicon. The proposed model consists of five major components that can extract features, determine opinion words regarding identified features with their semantic orientation, aggregate multiple opinions and generate structured summary. They conduct two experiments that have been for features extraction and sentiment words determination by using 484 reviews from three different domains.

Another study in [63] conducted on Aspect Level Sentiment Analysis. The research was undertaken using at aspect level using deep Learning Approach namely CNN, LSTM, and Hybrid CNN with GRU and hybrid CNN with LSTM. The researcher collect total of 10,000 Comments from Amhara Media Corporation Facebook page for training and testing purpose. They used word2vec word embedding with deep learning model. The researcher report that CNN-GRU model achieved training accuracy of 98% and test accuracy of 98%. Hybrid CNN with GRU is selected as best approach.

Summary of related work

Most of the research were conducted on sentiment analysis at sentence level and document level. As far as our knowledge, there are only two study on sentiment analysis for Afaan Oromoo Language at aspect level. The first works was focus on sentiment analysis and aspect level summarization and the other was aspect based sentiment analysis for Afaan Oromoo review text. The first was conducted using rule based approach and the second was based on machine learning approach.

Traditional machine learning methods, including rule-based methods and statistical based methods. These studies generally relied on laborious manual annotation and feature engineering and then employed traditional machine learning to establish a sentiment classifier. Statistical-based methods rely on carefully designed manual features on large-scale datasets, resulting in a lot of waste of manpower and time. The neural network model can automatically learn the low-dimensional representation of reviews without relying on artificial feature engineering.

This feature allows neural networks to be used for aspect-level sentiment analysis tasks and has attracted the attention of researchers.[39] [53]. However, recurrent neural network (RNN) or convolutional neural network (CNN) mine the semantic information of aspect word and its context, which is easy to ignore the fact that they are insensitive to the location of key components.

To the extent of the knowledge of the researcher, there is no Research Conducted on Afaan Oromoo ABSA using BERT, so there are research gaps which are not included in their work. In this Paper Researcher Proposes Sentiment Analysis at Aspect Level Using BERT and CNN-BiLSTM model.

Author and year	Research Title	Method	Data set	Text granularity	remark
W. TARIKU 2017	"Sentiment Mining and Aspect Based Summarization of Opinionated Afaan Oromooo News Text,"	Rule based	400 reviews	Aspect level	88.3% achieved system accuracy
N. Wayessa and A. Sadik, 2020	Multi-Class Sentiment Analysis from Afaan Oromoo Text Based on Supervised Machine Learning Approaches	Support Vector Machine and Random Forest	1810 from twitter	Sentence level	SVM and RF Achieved 90% and 89% respectively.
A. A. Workineh,2019	Sentiments Analysis for Afaan Oromoo Socio-Politics Text	Deep learning	12662 from Facebook	Document level	79.99%
	Aspect based sentiment analysis for Afaan Oromo movie review	Machine learning	2800, you tube	aspect	RF 88%, LR 87%, SVM 87%&88% MNB 88%

Table 3: summery of related work

Chapter 3

Methodology

3.1. Overview

In this section, the overall design of aspect level sentiment analysis for Afaan Oromoo texts, provide the basic architecture and brief description of different tasks involved in the process are discussed in detail. The overall activity consists of research design, proposed system architecture, data annotation, data preprocessing, aspect Term extraction, aspect sentiment classification and evaluation matrix are discussed under this chapter.

3.2. Research design.

The researcher reviewed different paper on aspect based sentiment analysis to identify and understand methods and methodology used. Basically we used deep learning models like CNN-BiLSTM and BERT for aspect extraction and sentiment classification. We select the experimental research design for successful completion of our study.

3.3. Proposed System Architecture

The proposed system architecture for Aspect-Based Sentiment Analysis (ABSA) involves two main components: Aspect Extraction and aspect Sentiment Classification. This architecture utilize a CNN-BiLSTM model for aspect extraction and a fine-tuned BERT model for sentiment classification. The entire system is designed to work with Afaan Oromoo text. The first component in the architecture is preprocessing that includes tasks such as removing punctuation, numbers, URLs and stop words from a review text. The next task after preprocessing is BIO annotation for tagging BIO for aspect term. The next task after annotation is aspect extraction for extracting relevant aspect. CNN in this case is used for learning important local features and BiLSTM Processes the sequence of embeddings in both forward and backward directions to capture long-term dependencies and context from both sides of the token. A fully connected layer is applied to each time step of the BiLSTM output to predict aspect labels for each token. Then BERT is used for sentiment classification as positive neutral and negative for extracted aspect.

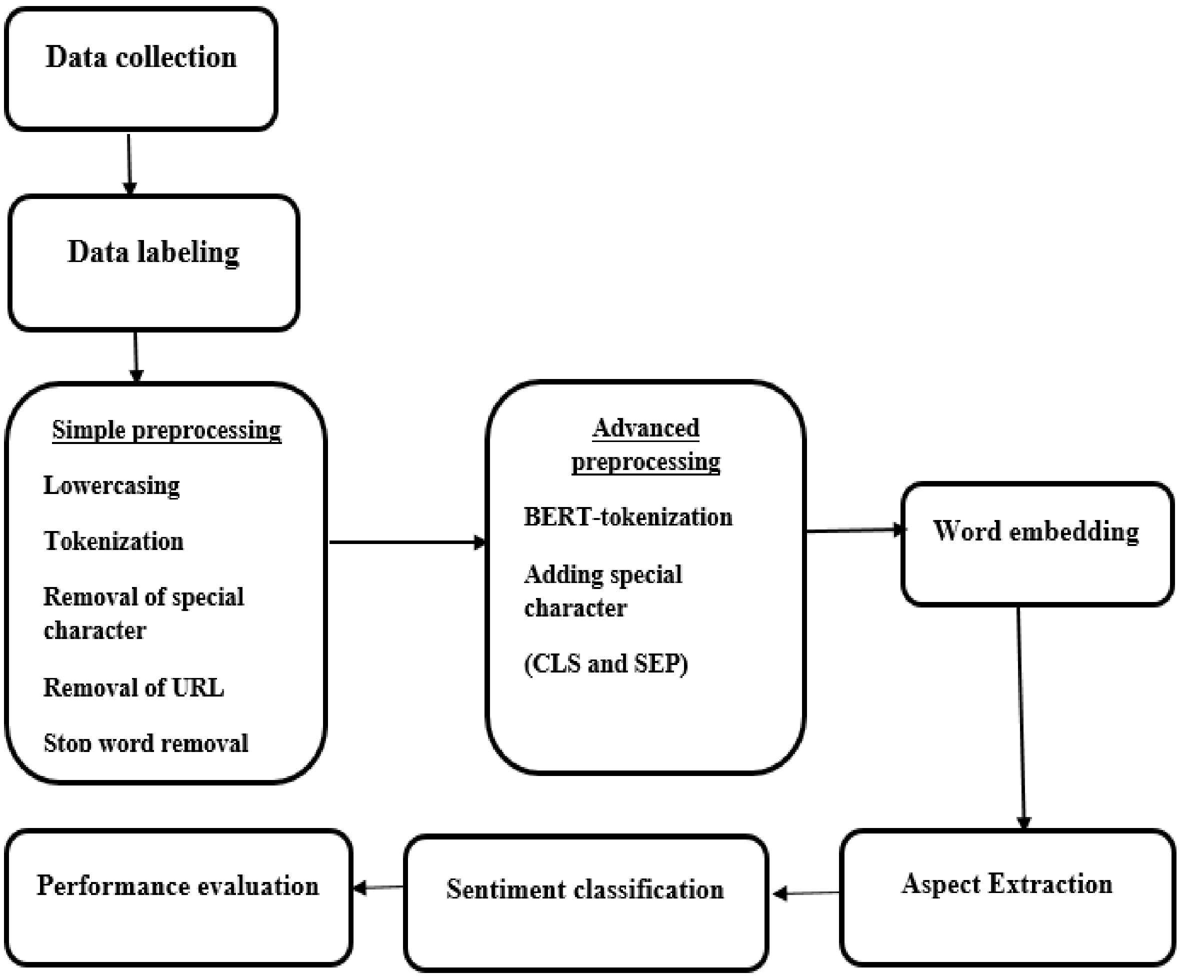

Figure 10: Proposed System Architecture

3.4. Data collection

For this study, the researcher collect Afaan Oromoo Aspect based sentiment analysis dataset from FBC Afaan Oromoo Facebook page, BBC Afaan Oromoo news and other relevant social media. Afaan Oromoo ABSA dataset cover three domains such as coffee, flower, and gold, and the aspects of these domain are extracted for this study.

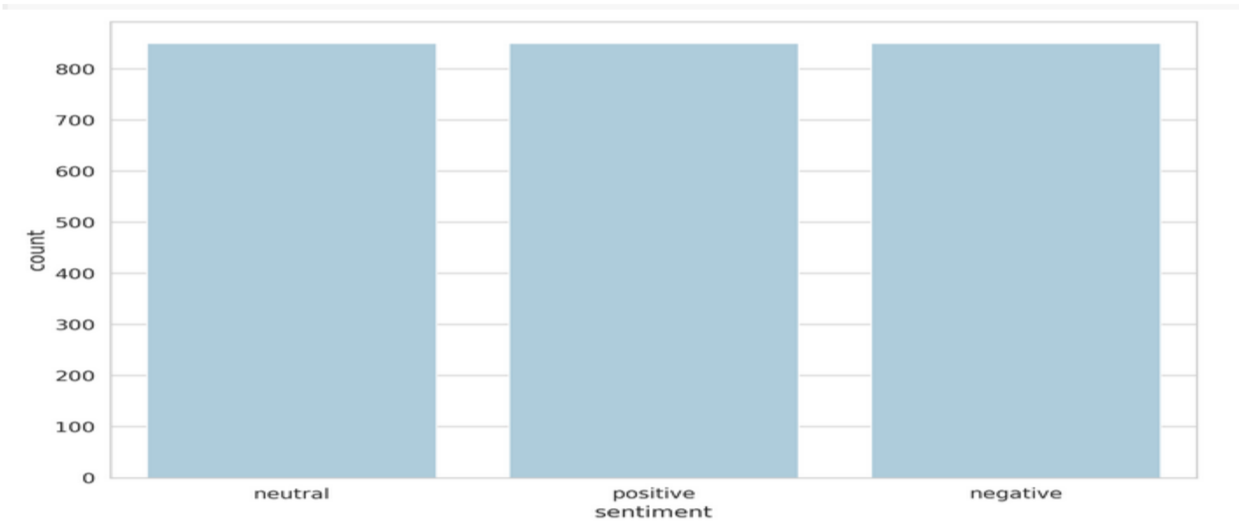

Figure 11: Sentiment value count

In this study, we prepared the data set that contains four columns; “id”, ”review_text”, “Aspect-terms” and Sentiment.

ID: In order to identify unique sentences we used ID Number. As one sentence may have more than one aspects, we gave the same id number for that sentences within the aspect. For instance the sentence ” worqeen halluu nama hawwatu qaba garuu gatii qaaliin bituun ni ulfaata” ,has two aspects “halluu” and “gatii”, we gave the same ID number for the sentences by writing it twice as we have two aspect terms as showed in sample dataset in table below with 102 ID number.

Texts or Reviews: Sentences or paragraphs from domains like coffee, gold, or flowers, expressing opinions or sentiment of a product.

Aspect Terms: Specific words or aspects in the text representing the focus of the sentiment (e.g., **"qulqullinna", "gatii", "galii", "diinagdee", "sanyii", "dhandhama", "foolii", "halluu", "oomisha bunaa" and "worqee baasuu"**)

Sentiments: The sentiment associated with each aspect (e.g., positive, negative, neutral).

Annotations: Explicit labels added during the annotation process, such as BIO tagging for aspect extraction and sentiment labels for classification.

The sample of actual data is represented in table below;

	id	review_text	aspect	sentiment
0	100	hojiin albuuda baasuu oromiyaa kessatti dinagd...	albuuda baasuu	neutral
1	100	hojiin albuuda baasuu oromiyaa kessatti dinagd...	diinagdee	positive
2	101	Itoophiyaan biyya albuuda heddun qabduudha.	albuuda	neutral
3	102	worqeen halluu nama hawwatu qaba akkasuma gat...	halluu	positive
4	102	worqeen halluu nama hawwatu qaba akkasuma gat...	gatii	negative

Table 4: sample of labeled dataset

The data is split into two tasks: Aspect extraction which is identify what is being talked about and Aspect sentiment classification which determine the sentiment related to each aspect. Data is stored in structured formats of CSV file for compatibility with machine learning frameworks. The dataset includes reviews about specific domains (coffee, gold, flowers), ensuring that domain-specific aspects are covered. From these domains ten most frequent aspect terms namely "qulqullinna", "gatii", "galii", "diinagdee", "sanyii", "dhandhama", "foolii", "halluu", "oomisha bunaa" and "worqee baasuu" are used. The data is balanced across domains and sentiment classes (positive, negative, and neutral).

3.4.1. Data annotation

Data annotation the way of labeling or tagging data to make it usable for machine learning model. Data annotation provide structured information that help model to understand particular task. For aspect term, we use BIO format (short for Beginning, Inside, Outside). BIO annotation is a widely used tagging scheme in tasks like named entity recognition (NER) and aspect extraction. The acronym "BIO" stands for Beginning, Inside, and Outside, which are used to label tokens in a sequence according to their position relative to specific entities or aspects of interest, we assign a label for each word in the sentence, where "B" indicates the start of an aspect term, "I" indicates the continuation of an aspect term, and "O" indicates not an aspect term. An example sentences with labels in BIO format. The target is halluu the label *B* indicates the beginning of a target, *I* indicates that the word is inside a target, and *O* indicates a word belongs to no target.

words:	abaaboon	kaleessa	bite	halluu	bareedaa	qaba
Labels:	O	O	O	B	O	O

Table 5: example of BIO tagging

The three domain used for our study namely coffee, flower and gold with number of data are described in table below.

Name of product	No. of data
coffee	970
flower	862
gold	718
Total	2550

Table 6: statistics of aspect labelling

For aspect sentiment classification task, two human annotator were selected for manual annotation purpose. The researcher choose these annotator based on specific knowledge they have on Afaan Oromoo language and experience on annotation. Detail the guidelines was provided to the annotators based on how sentiment classification were defined and how ambiguities were handled.

Annotation guideline

Annotation guideline is important to ensure consistency and accuracy in the labeling process. When the text expresses a favorable/positive such as opinion toward an aspect, it is positive sentiment. When the text expresses a negative opinion toward an aspect, it is negative sentiment and neutral sentiment when text is either factual, without emotional expression, or balanced in its sentiment, neither positive nor negative. For instance the sentence "qulqullinni bunaa baayyee gaariidha garuu tajaajila gad-aanaa kennu/The coffee quality is very good but the service was slow," indicate positive sentiment toward "qulqullinni bunaa/coffee quality aspect," and negative sentiment toward "tajaajila/service" aspect.

Inter-Annotator agreement (IAA)

Inter-annotator is important to measure how consistently different annotators label or classify data. Inter-Annotator Agreement is used to ensure consistency and reliability of annotated data for machine learning models. High agreement among annotators and reliable annotation process indicates a well- defined guidelines. Frequent disagreements, on the other hand, might indicate that the task is subjective or that the guidelines need clarification. We used Cohen's kappa coefficient .Cohen's Kappa measures inter-annotator agreement for categorical items while correcting for chance agreement.

Cohen's Kappa (k) is used when two annotator are involved.it adjusts for the agreement that would happen by chance. Kappa values range from -1 (complete disagreement) to 1 (perfect agreement) as shown in table 7.

Agreement interpretation	Kappa coefficient value
Less than 0.00	Poor agreement
0.00 to 0.20	Slight agreement
0.21 to 0.40	Fair agreement
0.41 to 0.60	Moderate agreement
0.61 to 0.80	Substantial agreement
0.81 to 1.00	Almost perfect agreement

Table 7:Inter-annotator kappa value agreement

From total 2550 two annotator annotate dataset as follow:

	Positive(A2)	Negative(A2)	Neutral(A2)	Total (A1)
Positive(A1)	840	1	1	842
Negative(A1)	2	845	3	850
Neutral(A1)	6	0	852	858
Total (A2)	848	846	856	2250

Table 8: value of annotators labelling

Where: A1: Annotator 1 labels and A2: Annotator 2 labels

Observed Agreement (Po)

This is the percentage of cases where the two annotators agree exactly. Since both annotator labeled 840 as positive, so it is positive agreement. Negative agreement is 845 as both of them it as negative and 852 neutral agreement.

So, the observed agreement of Po = $\frac{840+845+852}{2550} = 0.998$

Expected Agreement (Pe): expected agreement is determined by calculating likelihood that both annotators would assign the same label to a category by chance.

Probability of positive (A1): $842/2550 = 0.330$

Probability of positive (A2): $848/2550 = 0.332$

The expected agreement for positive: Pe (positive) = 0.330 x0.332=0.110

Probability of negative (A1): $850/2550 = 0.333$

Probability of negative (A2): $845/2550 = 0.331$

The expected agreement for negative: Pe (negative) = 0.333 x0.331=0.110

Probability of neutral (A1): $858/2550 = 0.336$

Probability of neutral (A2): $856/2550 = 0.336$

The expected agreement for neutral: Pe (neutral) = 0.336 x0.336=0.113

The total expected agreement Pe: = 0.110 +0.110+0.3=0.333

Cohen's Kappa is calculated using the formula: $k = \frac{po-pe}{1-pe}$ *(3)*

$$\Rightarrow \frac{0.998-0.333}{1-0.333} = \frac{0.665}{0.667} \approx 0.996$$

Thus, Cohen's Kappa value is approximately 0.996, which indicates almost perfect agreement between the two annotators.

As a result we trained model according to both annotator to see effects on their agreement, as the result is almost the same.

3. 5. Data preprocessing

Data preprocessing is the process of cleaning and prepare data for analysis[64]. Preprocessing is important to improve performance and accuracy by reducing computational process and feature space. This stage performed to remove any non- meaningful and irrelevant words and symbols such as URLs, special characters (@, #, $, & and other symbols, HTML scripts. Preprocessing step such as text cleaning, tokenization, normalization and stop word removing are applied for our study.

Text Cleaning: Remove any irrelevant information such as HTML tags, special characters, stop words and normalize text by converting it to lowercase.

Stop word removal: Stop words are frequently occurring words in a language that are typically removed during natural language processing (NLP) tasks because they contribute minimal meaningful information. While these words are essential for grammatical structure, they usually do not add significant value to tasks. But some stopwords in Afaan Oromoo language have a great role to determine the negativity of sentiment to particular subject. For example, “hin” and “miti” are Afaan Oromoo stopwords used to indicate the negativity of words/phrases. Also, the stopwords “gaariidha”, “gaarii miti,” “faayidaa qaba,” and “faayidaa hin qabu” have remained in the reviewed text since they reveal important information about a particular subject/service. Without these two stopwords, the rest are removed from the reviewed text technically. We used stop words from [65]

Tokenization: tokenization is one of the important part of preprocessing. It is the process of splitting a sequence of character in to token. Tokenization used in natural language processing paragraph and sentences into small units. For this study, the researcher use NLTK library and BERT tokenizer to split Afaan Oromoo sentences to smaller piece of words.

Normalization: Normalization is the process of converting text to standard format to reduce variability and noise, improving the quality of data for further analysis. The word which have the same meaning and different writing system should is normalized to the same format. All capital letters are normalized to lowercase.

Padding and Truncation: When preparing data for model input, sentences may need to be padded to a fixed length or truncated if they exceed a certain length. This preprocessing ensured that the dataset was clean, consistent, and ready for input into the aspect extraction and sentiment classification models.

For pre-trained BERT model, the dataset is preprocessed in particular way. For preprocessing, tokenization and addition of special token, two step should be followed. The BERT model was trained to convert sentences into tokens using WordPiece tokenization. Additionally, special tokens such as [CLS] and [SEP] were incorporated into the dataset. Subsequently, the main dataset was divided into two subsets: a training dataset and a test dataset. In this study the data splitting was performed using the Train-Test Split function from the scikit-learn library.

3.6. Aspect term extraction

Feature extraction in natural language processing involves transforming raw text into a structured and meaningful format that can be utilized by machine learning models. These features capture important information from the text, enhancing the performance of various tasks. The researcher used word2vec word embedding technique to convert input text into integer. The series of convolutions available in convolutional neural network used to extract relevant local feature from input text. Convolutional Neural Networks (CNNs) are traditionally used for image processing, but they can also be applied to text to capture local features like n-grams (e.g., short sequences of words) [66].

CNN layer is designed to capture local patterns in the text, such as phrases or n-grams (e.g., "Qulqullinna bunaa"). These local patterns are important for identifying potential aspect terms. The CNN applies a set of filters that slide over the sequence of word embedding. Each filter extracts features by performing a convolution operation, which involves multiplying the filter with the local region of the input sequence and summing the results. After convolution, an activation function typically relu is applied to introduce non-linearity, allowing the model to learn complex patterns. This step retains the most significant feature from each feature map, helping the model to be robust to small variations in the input.

Aspect term extraction is the task of identifying specific entities or aspects mentioned in a text. The goal is to find out an important term in a given sentence. For example, in a sentence like "The coffee quality is excellent," the aspect term "coffee quality" needs to be extracted. The researcher use BIO annotation scheme to annotate aspect terms. Then aspect terms are extracted by using CNN-BiLSTM model. The aspect terms are extracted from three domain coffee, gold and flower. CNN-BiLSTM is adept at recognizing and labeling aspect terms in the text through a combination of local feature extraction via CNN and contextual understanding via BiLSTM. The first step in the model is to convert each word in the input sentence into a dense vector representation, known as word embedding. Each word in the sentence is converted into a word embedding, resulting in a sequence of vectors.

The BiLSTM layer captures long-range dependencies and contextual information from both directions (forward and backward). This is important because a word's meaning and role in a sentence often depend on the surrounding words. LSTM units are designed to retain essential information over long sequences and discard irrelevant data using input, forget, and output gates to regulate the flow of information. In a BiLSTM, two LSTM networks are used: one processes the input from left to right (forward) and the other from right to left (backward). The outputs from both directions are combined to create a comprehensive representation of each word in context. This produces a contextualized representation for each word, considering the words before and after it. A fully connected layer then converts these representations into scores for each possible label (B-Aspect, I-Aspect, O).

In Aspect-Based Sentiment Analysis (ABSA)**,** aspect identification and sentiment classification are interdependent tasks. The quality of aspect identification directly impacts the performance of sentiment classification. Aspect Identification task identifies the targets or aspects of the sentiment in a given text. Without correctly identifying aspects, it becomes ambiguous what sentiment is being expressed. Sentiment classification task determines the sentiment polarity (positive, negative, or neutral) associated with each identified aspect.

Misidentified or missed aspects can lead to misclassification or incomplete sentiment analysis. Aspect should identified correctly unless, its sentiment cannot be classified. For instance in the sentences "The coffee aroma is excellent". If "aroma" is not identified as an aspect, the positive sentiment related to it will be lost. Misidentified aspects also leads to errors in sentiment classification, affecting overall metrics like accuracy, precision, recall, and F1-score. For this reason, aspect should be correctly identified in order to improve aspect sentiment classification performance. Considering this, after aspect is labeled we apply BIO annotation for correct identification of aspects.

Here is a sample how aspect terms are annotated with BIO labels. In the below table the sentences that have more than one aspect is given the same id.

1	id	review_text	aspect	sentiment	BIO_tags
2	100	hojiin albuuda baasuu oromiyaa kessat	albuuda baasuu	neutral	O B-Aspect I-Aspect O O O O
3	100	hojiin albuuda baasuu oromiyaa kessat	diinagdee	positive	O O O O O B-Aspect O O
4	101	Itoophiyaan biyya albuuda heddun qat	albuuda	neutral	O O B-Aspect O O
5	102	worqeen halluu nama hawwatu qaba a	halluu	positive	O B-Aspect O O O O O O O
6	102	worqeen halluu nama hawwatu qaba a	gatii	negative	O O O O O O B-Aspect O O
7	103	worqee halluu bareedaa qabu argine g	halluu	positive	O B-Aspect O O O O O O O O O O O O O O O
8	103	worqee halluu bareedaa qabu argine g	gatii	negative	O O O O O B-Aspect O O O O O O O O O O O

3.7. Aspect sentiment classification (ASC)

Aspect Sentiment Classification (ASC) is an interesting and challenging research task to identify the sentiment polarities of aspect words in sentences[15]. Aspect sentiment classification classify a sentiment with the corresponding aspects. The goal is to determine sentiment associated with the aspect in the sentences. The sentiment is classified as positive, negative and neutral for extracted aspect from sentences. For example "qulqullinni bunaa baayyee gaariidha garuu gatii qaalii kanaan bituun ni ulfaata. In this sentencise, it is positive sentiment for qulqullinni bunaa/coffee quality and negative sentiment for gatii/price aspect. For this reason we understand the aspect based sentiment analysis improve the feature of the product or service that their customer like or dislike.

For aspect sentiment classification task, the researcher fine tune pre-trained BERT model. Once BIO annotation is completed, aspect term are extracted using CNN-BiLSTM, BERT model used to predict and classify sentiment for each aspect. The sentiment classification process with BERT involves using the [CLS] token to predict the sentiment of the entire input sequence. During fine-tuning, the [CLS] token's representation, after passing through the transformer layers, is used as the aggregate summary of the entire input sequence. This representation captures the overall context and meaning of the text. The softmax function converts the raw output scores from the classification head into probabilities, where each probability represents the likelihood of the input text belonging to a particular sentiment class. The sentiment class with the highest probability is selected as the final prediction.

3.8. Evaluation Metrics

Various known evaluation metrics are employed to evaluate the model's performance on both task. They are assessed using True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) equations for accuracy, precision, recall, and F1-score [67].

Accuracy represents the proportion of correct predictions made by the model compared to the total number of predictions. It reflects how well the model's predicted sentiments align with the actual sentiments in the test data.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

Precision is the proportion of true positive predictions to the total number of positive predictions made by the model. It evaluates how accurately the model identified positive, neutral, or negative sentiments, highlighting the model's effectiveness in correctly predicting specific sentiments.

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

Recall is the ratio of true positive predictions to the total number of actual positive instances. It measures the model's ability to identify all relevant positive cases. By evaluating how well the model detects all actual positive, neutral, or negative sentiments in the dataset, recall provides insight into the model's sensitivity.

$$recall = \frac{TP}{TP+FN} \tag{6}$$

The F1-score is the harmonic mean of precision and recall, offering a balanced evaluation of the model's performance. It is particularly useful when dealing with imbalanced classes, as it combines both precision and recall into a single metric.

$$F1 - Score = \frac{2*(Precision \times Recall)}{Precision + Recall} \tag{7}$$

Chapter 4

Experimental result and discussion

4.1. Overview

The overall design of proposed model and its architecture are discussed in previous chapter. In this chapter the implementation and experimental evaluation of the proposed model for Aspect level sentiment analysis of Afaan Oromoo text is discussed in detail. Furthermore, the experimental details such as dataset used for training and testing the proposed model, implementation tools used, evaluation and test results will be presented in detail.

4.2. Dataset description

In this study, our main aim is determining the sentiment polarity towards each aspect. The dataset consists of 2,550 Afaan Oromoo texts. These texts were collected from FBC Afaan Oromoo Facebook, BBC Afan Oromo and another relevant social media, domains such as coffee, gold, and flowers. . For training and testing purpose the dataset is divided into 80%, 20%. 80% of dataset is used for training the model, while the remaining 20% is reserved for testing. These dataset consists four column “id”, “review_text”, “aspect” and “sentiment column. The dataset are stored excel with csv file format. Different annotation method and pre-processing steps applied to the dataset in order to improve accuracy of dataset. From 2550 sentences, 2040 sentences used for training purpose and 510 sentence is used for training purpose. From 2550, unique sentence is 1610.

4.3. Experimental environment setup

In this study we utilized the following Software and Hardware tools.

Software: The models were implemented using libraries such as Keras and Tensor flow. Data preprocessing and analysis have been done using Python libraries like Pandas, NumPy, and NLTK.

Hardware: The experiments were conducted on a system with a specific configuration, such as using a GPU for training the models to speed up the process. To develop the model, we have used

Intel® Core i5 HP computer with processor speed of 1.7GHz and 8GB RAM with windows 10 operating system. Google collaborator used for code edition. The detail is described in table 9.

Tool used	Specification
Computer	Intel® Core i5 HP, processor speed of 1.7GHz 8GB RAM with windows 10 os.
GPU	for training the models to speed up the process
Programming language used	Python
Libraries	Keras and Tensor flow for implementation , Pandas, NumPy, and NLTK for preprocessing

Table 9: Experimental environment setup

4.4. Hyperparameters 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 [60]. It controls training process of machine learning model and improve its performance. Hyperparameters are specified before training and influence how the learning process functions.

We have carried out total of 10 experiments for both experiment where four experiment is carried out for aspect term extraction and 6 experiments for aspect sentiment classification by applying different hyper parameter.

Experiment	Embedding Dim	CNN Filters	Kernel Size	LSTM Units	Dropout Rate	Batch Size	Learning Rate	Optimizer	Epochs	Accuracy
1	100	128	5	128	0.3	32	0.001	Adam	5	94.8%
2	100	64	3	50	0.1	32	0.001	Adam	10	96.55%
3	150	64	3	128	0.5	16	0.0005	Adam	5	95.22%
4	200	128	7	256	0.4	16	0.001	Adam	10	93.78%

Table 10: Experimental result using different hyperparameter for aspect extraction

Experimental result using different hyperparameter for aspect sentiment classification.

Experiment	Hyperparameter	Values Tested	Best Result	Observations
Learning rate	lr	1e-5,2e-5,3e-5,4e 5, 5e-5	2e-5	2e-5 yields high accuracy
Batch Size	batch_size	8,16,32	16	16 show good result
Sequence length	Max_length	64,128,256	128	128 is appropriate for our case
Number of epochs	epochs	3,5,10	5	Our model is saturated at 5 epoch
optimizer	AdamW			As it is Alternative for BERT model, it shows good result

Table 11: Hyperparameters experiment for aspect sentiment classification.

4.5. Experiment Results

To effectively capture both local and sequential features in text for aspect extraction, we employed a CNN-BiLSTM model. The CNN captures local dependencies and patterns, while the BiLSTM captures long-term dependencies in both forward and backward directions.

For this study the total of 2550 are collected from three domain. The available data is divided into training and testing sets using an 80/20 split for the experiment. Specifically, 80% of the data is which is 2040 data set is used to train and develop the model, while the remaining 20% which is 510 is reserved for evaluating how well the model performs on unseen data. To assess the performance of the proposed model, accuracy, precision, recall, and loss are used as key performance metrics.

The experiment for aspect sentiment analysis (ABSA) model is carried out in five stage. Initially raw text are collected and preprocessed. Next, the cleaned text transformed into integer based vector. Feature extraction method is applied for text to capture local feature, then aspect terms are extracted. Finally the task of aspect sentiment classification task is take place. For purpose of aspect extraction, the researcher utilized CNN-BiLSTM. This model integrates Convolutional Neural Networks (CNN) to detect local patterns in the text and Bidirectional Long Short-Term Memory (BiLSTM) networks to capture contextual information from both directions. Bidirectional Encoder Representations from Transformers (BERT) was utilized for sentiment classification. BERT is particularly useful for this study because it can handle the nature of Afaan Oromoo and learn from the limited data available.

4.5.1. CNN-BiLSTM experiment for Aspect Term extraction

CNN-BiLSTM is adept at recognizing and labeling aspect terms in the text through a combination of local feature extraction via CNN and contextual understanding via BiLSTM. A Softmax function is applied to these scores to produce a probability distribution over the labels for each word. The label with the highest probability is chosen as the final prediction for each word. The final output of the model is a sequence of BIO tags, where each word in the input sentence is assigned a label indicating whether it is part of an aspect term. For example, in the sentence "Qulqullinni bunaa baayyee gaariidha," the model output is: "Qulqullinni B-Aspect bunaa I-Aspect baayyee O Gaariidha O." correctly identifying "qulqullinni bunaa" as the aspect term. Embedding Layer converts tokens into dense vectors representing their semantic meaning. CNN Layer captures local dependencies and features in the text by applying convolutional filters over the word embeddings. BiLSTM layer Processes the sequence of embeddings in both forward and backward directions to capture long-term dependencies and context from both sides of the token. A fully connected layer is applied to each time step of the BiLSTM output to predict aspect labels for each token. The output Layer generates BIO tags (Begin, Inside, Outside) for aspect extraction.

During experiment we carried out four experiment for aspect extraction using different hyperparameter.

Number of Filters in Conv1D Layer: we carried out experiment with different numbers of filters to find the optimal balance between capturing features and computational cost. We tried different value of filters 64 and 128 from experiment and we get that 64 is optimal one.

Kernel Size in Conv1D Layer: kernel sizes is used to evaluate how much local context the model should focus on. For our experiment, we tried three value of kernel size; 3, 5 and 7.in our case 3 shows more effects than the other.

LSTM Units in Bidirectional Layer: Adjust the number of LSTM units to control the capacity for sequential feature learning. We tried different value of LSTM units 50,128 and 256. As a result we get that 50 appropriate as more unit may risk overfitting.

Batch size: during the experiment, we checked 16 and 32 batch size, 32 batch size show good result in our case.

Epochs: we tried different number of epochs but our model yields good result at 10 epoch.

In addition 0.1, 0.3, 0.4, and 0.5 dropout rate is used during experiment, 0.1 dropout rate with Adam optimization yields good result. As a result, with the appropriate hyperparameter which is 64 number of filter with 3 kernel size, 50 LSTM units 32, batch size, 10 epochs with Adam optimization yields good result on aspect term extraction. For training and testing purpose the dataset is divided into 80%, 20%. As a result model achieved 96.55% of training accuracy and 92.81% of validation accuracy. The detail is explained as follow:

```
Epoch 1/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 15ms/step - accuracy: 0.9628 - loss: 0.0828 - val_accuracy: 0.9308 - val_loss: 0.2183
Epoch 2/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 13ms/step - accuracy: 0.9619 - loss: 0.0796 - val_accuracy: 0.9290 - val_loss: 0.2407
Epoch 3/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 13ms/step - accuracy: 0.9638 - loss: 0.0768 - val_accuracy: 0.9276 - val_loss: 0.2460
Epoch 4/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 13ms/step - accuracy: 0.9640 - loss: 0.0761 - val_accuracy: 0.9260 - val_loss: 0.2525
Epoch 5/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 12ms/step - accuracy: 0.9650 - loss: 0.0732 - val_accuracy: 0.9271 - val_loss: 0.2604
Epoch 6/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 2s 18ms/step - accuracy: 0.9655 - loss: 0.0698 - val_accuracy: 0.9260 - val_loss: 0.2769
Epoch 7/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 19ms/step - accuracy: 0.9652 - loss: 0.0714 - val_accuracy: 0.9277 - val_loss: 0.2800
Epoch 8/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 16ms/step - accuracy: 0.9657 - loss: 0.0698 - val_accuracy: 0.9256 - val_loss: 0.3007
Epoch 9/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 13ms/step - accuracy: 0.9655 - loss: 0.0694 - val_accuracy: 0.9210 - val_loss: 0.2995
Epoch 10/10
65/65 ━━━━━━━━━━━━━━━━━━━━ 1s 13ms/step - accuracy: 0.9655 - loss: 0.0683 - val_accuracy: 0.9281 - val_loss: 0.3151
```

Figure 12: training process of CNN-BiLSTM

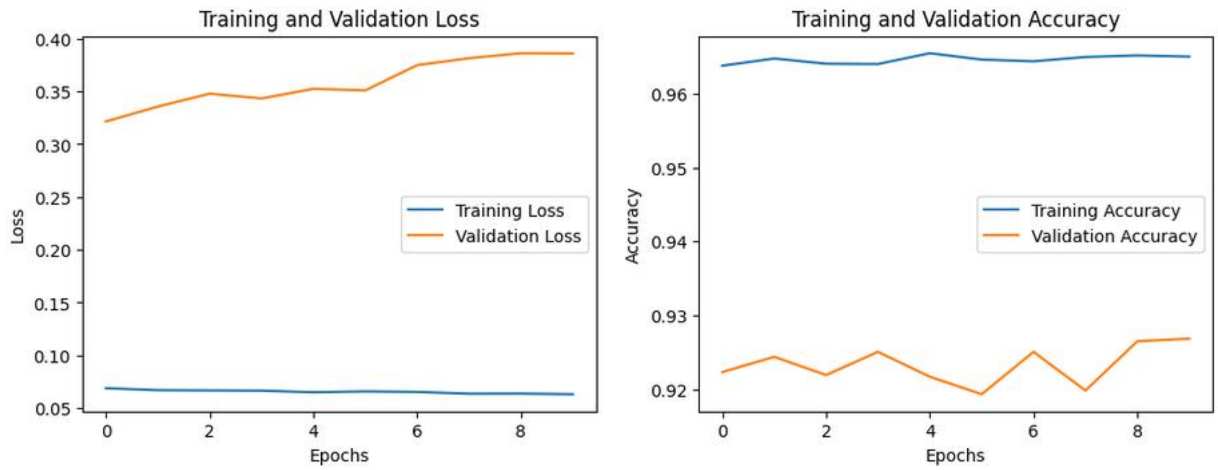

Figure 13: Training & validation accuracy and loss curve of CNN-BiLSTM Model

4.5.2. BERT experiment for Aspect Sentiment Classification

For aspect sentiment classification task Bert-base uncased model is fine-tuned on our dataset. Fine-tuning is the way of adopting pre-trained model to our model. To fine tune BERT for our task, first the text data is tokenized using BERT tokenizer, and prepared in the way that BERT can understand. Once text is breaks down to tokens, special tokens like Classification token [CLS] and separator token [SEP] are added. Then BertForSequenceClassification is fine-tuned for classification task. During fine-tuning, the last layer of BERT is replaced with classification head usually a fully connected layer with softmax to map the sentiment classes namely positive, negative and neutral. Finally the model that include BERT's layer is trained on our data set.

For aspect sentiment classification using BERT we carried out only 6 experiments due to limit of memory of google colab by changing different hyper parameter to see the effect.

Learning Rate: during experiment we test the effect of different learning rates in AdamW. We tried over 1e-5, 2e-5, 3e-5, 4e-5 and 5e-5. As a result standard BERT learning rate which is 2e-5 yields good result.

Batch Size: we tried different batch size; 8, 16 and 32. After we test these batch sizes we get that; the Smaller batches may improve generalization, while larger batches requires more memory. For our case 16 batch size shows good result.

Number of Epochs: during experiment we evaluate between under fitting (too few epochs) and overfitting (too many epochs). We tried three number of epochs 3, 5 and 10. As a result our model shows good result at 5 epochs.

In addition we used alternative optimize which is AdamW. As a result BERT-base uncased with 5 number of epochs,2e-5 learning rate, 16 batch size,128 sequence length and AdamW optimization yields good performance as showed in table below.

The dataset is divided 80%, 20% for experiment. The extracted aspect terms are used as input for aspect sentiment classification. For aspect sentiment classification task, BERT Achieve accuracy of 87%, 86.67% of precision, 86.68% of recall and 86.66 % of f1 score. Generally classification result is explained as follow;

```
Some weights of BertForSequenceClassification were not initiali
You should probably TRAIN this model on a down-stream task to b
/usr/local/lib/python3.10/dist-packages/transformers/optimizati
  warnings.warn(
Epoch 1, Average Training Loss: 1.0313, Accuracy: 0.4416
Epoch 2, Average Training Loss: 0.7607, Accuracy: 0.6705
Epoch 3, Average Training Loss: 0.5431, Accuracy: 0.7864
Epoch 4, Average Training Loss: 0.3856, Accuracy: 0.8581
Epoch 5, Average Training Loss: 0.2842, Accuracy: 0.8923
```

Figure 14: BERT Training process for sentiment classification

The evaluation metrics show accuracy, precision, recall, and F1-score for each sentiment category (positive, negative, and neutral) as follow;

sentiment	precision	recall	F1-Score
positive	84%	85%	84%
negative	87%	84%	85%
neutral	90%	91%	91%
Overall	86.67%	86.68%	86.66%

Table 12: Evaluation metrics for aspect sentiment classification

BERT-base-uncased successfully classified sentiment with context. For the sentence *"Buna dhandhama mi'aawaa qabuudha garuu gatiin isaa qaalii dha,"* BERT accurately classified the sentiment for *"taste"* as positive and *"price"* as negative.

The overall model summary

model	Accuracy	Task
CNN-BiLSTM	92.81%	Aspect term extraction
BERT	87%	Aspect sentiment classification

Table 13: experiment summary

4.6. Discussion

This study introduces an ABSA (aspect based sentiment analysis) model tailored for Afaan Oromoo and assesses its performance using test data. The system comprises several core components, each handling specific tasks necessary for the final outcome. These components include a preprocessing stage for eliminating extraneous symbols, numbers, and URLs, and a word embedding stage that transforms sentence sequences into integer sequences and then into fixed-length vectors. Once the input sentences are converted into integer vectors, the padded sequences are fed into a Convolutional Neural Network (CNN), which serves as the feature extraction component to identify relevant and distinguishing features from the sentences. Two distinct experiments were conducted to develop and evaluate CNN-BiLSTM and BERT models for Afaan Oromoo ABSA. The results from these experiments, detailed in the preceding section, reveal that the CNN-BiLSTM model for aspect extraction achieved an accuracy of 92.8%. In contrast, the BERT-based model achieved 87% accuracy in Aspect sentiment classification. BIO annotation tag show good result on data set. CNN-BiLSTM achieved good performance on aspect extraction. BERT also achieved good performance even if dataset prepared for this study is too small compared to BERT Pre-trained dataset.

Finally the research question is answered as follow:

1. How effective are pre-trained BERT model in performing aspect-based sentiment analysis on Afaan Oromoo text, considering the language's low-resource status?

Bert model is pre-trained model on large dataset. For the task of aspect based sentiment analysis of Afaan Oromoo text, we fine-tuned BERT model on our dataset. First dataset is preprocessed and prepared in the way that BERT model can understand. Once training model is completed we apply standard metrics such as accuracy, precision, recall, and F1-score for sentiment classification across different categories (positive, negative, and neutral). Since Afan Oromo is a low-resource language, there is no standard annotated dataset for ABSA Task. However, BERT achieved good performance which is accuracy of 87%, 86.67% of precision, 86.68% of recall and 86.66 % of f1 score on our dataset.

2. What is the effectiveness of CNN-BiLSTM in extracting aspects from Afaan Oromoo text?

First we consider the known model that perform well on small dataset. As we reviewed from different source combination of model such as CNN-BiLSTM, BiLSTM-CRF and BERT-CRF are showed good performance on small dataset. As we can't use BERT for both task because of limited memory on google colab, we select CNN-BiLSTM for aspect term extraction. CNN-BiLSTM is a good for aspect extraction because it can capture both local patterns (via CNN) and contextual information (via BiLSTM), making it robust on small datasets. After CNN-BiLSTM model trained on our dataset, the 96.55% training accuracy and 92.8% validation accuracy is achieved. Therefore, the combination of CNN-BiLSTM achieved good performance on our dataset.

Chapter 5

Conclusion and recommendation

5.1 Conclusion

Aspect-Based Sentiment Analysis (ABSA) offers an enhanced method for understanding sentiment by breaking down text into aspects and evaluating the sentiment associated with each aspect. This approach surpasses traditional sentiment analysis by providing deeper insights into the detailed opinions expressed within the text. ABSA is particularly valuable in applications where understanding the sentiment toward individual features or aspects is crucial, such as customer feedback, product reviews, and social media analysis. The implementation of ABSA in this study demonstrates its effectiveness in capturing detailed sentiments related to specific aspects within the text. By utilizing advanced techniques for aspect extraction and sentiment classification, ABSA facilitates a more precise analysis.

Our work is generalized as follow; First the data is trained diversely which means all the data from thee domain is trained at the same time. The dataset is well-annotated that cover all three domains ensure that the model learns to generalize across topics. We used proportional amount of data for each domain during training and labeled each aspect with corresponding sentiment. Beside this, basically we select these three domain as they are all exportable items, thinking that they can have common aspect terms. We used ten most frequent aspect terms namely "qulqullinna", "gatii", "galii", "diinagdee", "sanyii", "dhandhama", "foolii", "halluu", "oomisha bunaa" and "worqee baasuu". From three domain have common aspect terms such as qulqullinna, gatii, galii and diinagdee. In addition, Coffee and flower have also two common aspects namely," sanyii" and "foolii". Coffee and gold have also common aspect term which is "halluu". Therefore variability of aspect terms is handled as we concentrated on aspect terms which is common.

This study presents a significant step forward in the application of aspect-based sentiment analysis (ABSA) for Afaan Oromoo text, a language with limited resources in the field of natural language processing. In this study, we developed an aspect-based sentiment analysis (ABSA) system for Afaan Oromoo text. Utilizing the BIO annotation scheme for identifying aspect terms, CNN-BiLSTM for aspect extraction, and BERT for aspect sentiment classification. Our approach successfully handled the unique challenges posed by Afaan Oromoo ensuring that both aspect terms and their sentiments were accurately captured. The integration of deep learning models with advanced natural language processing (NLP) techniques proved to be effective in processing Afaan Oromoo text, making a significant contribution to the field of sentiment analysis, particularly for low-resource languages.

5.2. Recommendation

This study attempts to develop model for Aspect Based Sentiment Analysis (ABSA) using BERT for Afaan Oromoo review text. However, we cannot cover all due to limit of time and working area memory.in addition, more efficient and effective analysis is still recommended for future work. Future work involves conducting a more detail analysis of specific mechanisms, applying the approach to various languages with the help of translators, and exploring new methods through alternative proposals. Therefore the following future research areas are recommended**.**

- ✓ In this research 2550 Afaan Oromoo dataset are collected, collect a larger and more diverse dataset of Afaan Oromoo text to further enhance the model's performance and generalizability. Incorporating additional domains beyond coffee, gold and flower could provide a more comprehensive understanding of sentiment in different contexts.
- ✓ This study integrate BIO tagging with CNN-BiLSTM for Aspect extraction and BERT for sentiment classification, Explore more advanced models like Transformer-based architectures for aspect extraction, which may offer improvements over the CNN-BiLSTM approach, especially in capturing long-range dependencies in text.
- ✓ Feature researches can also focus on implicit aspect expressions, opinion holder as well as aspect category detection which are not covered in this study.

5.3. Contribution

- ✓ This study presents a pioneering effort in applying aspect-based sentiment analysis to Afaan Oromoo text, contributing valuable insights and methodologies for analyzing sentiment in this under resourced language.
- ✓ The integration of BIO tagging with CNN-BiLSTM for aspect extraction and BERT for sentiment classification represents a comprehensive approach that can be adapted for other

low-resource languages. This contribution is significant for advancing NLP in languages with limited annotated data and tools.

- ✓ The methodologies and code developed in this research could be made available to the community as an open-source resource, encouraging further research and development in sentiment analysis for Afaan Oromoo and other similar languages.

References

[1] J. R. Kaur, J., & Saini, “Emotion Detection and Sentiment Analysis in Text Corpus,” A Differ. Study with Informal Form. Writ. Styles. Int. J. Comput. Appl. – 8887, vol. 101, p. 9, 2014.

[2] A. K.Joshi, “Natural Language Processing,” doi 10.1126/science.253.5025.1242, vol. 253, no. 5025, pp. 1242–1249.

[3] B. white erik cambaria, “A review of natural language processing,” jumping NLP curves.

[4] henry lieberman Reda mihalcea, hugo liu, “NLP (Natural Language Processing) for NLP (Natural Language Programming),” in Computational Linguistics and Intelligent Text Processing, p. pp 319–330.

[5] M. E. Mowlaei, M. Saniee Abadeh, and H. Keshavarz, “Aspect-based sentiment analysis using adaptive aspect-based lexicons,” Expert Syst. Appl., vol. 148, 2020, doi: 10.1016/j.eswa.2020.113234.

[6] Cai Z. and Z. He, “Trading private range counting over big IoT data, 2019 IEEE”

[7] X. W. F. H. et al. Lin Y., “Dynamic control of fraud information spreading in mobile social networks”.

[8] B. Huang, Y. Ou, and K. M. Carley, “Aspect level sentiment classification with attention-over-attention neural networks,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10899 LNCS, pp. 197–206, 2018, doi: 10.1007/978-3-319-93372-6_22.

[9] L. Z. Liu, Bing, “Sentiment analysis and opinion mining.,” Synth. Lect. Hum. Lang. Technol., vol. 5(1), pp. 1–23, 2016.

[10] A. Guangyao Pang , Keda Lu, Xiaoying Zhu, Jie He, Zhiyi Mo, Zizhen Peng and B. Pu, “Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model,” Priv. Prot. Incent. AI-Driven IoT, 2021.

[11] B. P. and L. Lee, “Opinion mining and sentiment analysis,” "Opinion Min. Sentim. Anal. Found. Trends Inf. Retr., vol. 2, nos. 1–2, pp. 1–135, 2008.

[12] L. Z. Liu, Bing, “survey of opinion mining and sentiment analysis,” "A Surv. Opin. Min. Sentim. Anal. Min. text data, no. . Springer, Boston, MA, pp. 415-463.., 2012.

[13] J. Smailović, “Sentiment Analysis in Streams of Microblogging Posts,” Kt.Ijs.Si, p. 140, 2014, [Online]. Available: http://kt.ijs.si/theses/phd_jasmina_smailovic.pdf

[14] V. Gupta and G. S. Lehal, “A survey of text mining techniques and applications,” J. Emerg. Technol. Web Intell., vol. 1, no. 1, pp. 60–76, 2009, doi: 10.4304/jetwi.1.1.60-76.

[15] P. S. Xu, H., Liu, B., Shu, L. and Yu, “A failure of aspect sentiment classifiers and an

adaptive reweighting solution," arXiv Prepr. arXiv1911.01460, pp. 1–12, 2019.

[16] B. Sangamithra, P. Neelima, and and M. S. Kumar, "A Memetic Algorithm for Multi Objective Vehicle Routing Problem with Time Windows," in Proceedings - 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering, ICEICE 2017, 2017, pp. 1–8. doi: 10.1109/ICEICE.2017.8191931.

[17] L. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit and and I. P. Jones, Aidan N. Gomez, Lukasz Kaiser, "Attention is all you need.," CoRR, abs/1706.03762, 2017.

[18] O. L. Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill and and S. Bowman, "GLUE: A multi-task benchmark and analysis platform for natural language understanding.," in In Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, 2018, pp. 353–355.

[19] K. Devlin, J., Chang, M. W., Lee, K., & Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv Prepr. arXiv1810.04805, 2019.

[20] F. D. la P. 3 Nhan Cach Dang, María N. Moreno-García, "Sentiment Analysis Based on Deep Learning: A Comparative Study," Electronics, vol. 9, p. 483.

[21] W. TARIKU, "Sentiment Mining and Aspect Based Summarization of Opinionated Afaan Oromoo News Text," 2017.

[22] M. Oljira, K. Kaliyaperumal, B. Natea Sima, and K. Sori, "Sentiment Analysis for Afaan Oromoo Using Combined Convolutional Neural Network and Bidirectional Long Short-Term Memory," Int. J. Adv. Res. Eng. Technol., vol. 11, no. 11, pp. 101–112, 2020, [Online]. Available: http://iaeme.com/Home/journal/IJARET101itor@iaeme.comhttp://iaeme.comhttp://iaeme.com/Home/journal/IJARET102editor@iaeme.comhttp://iaeme.com

[23] M. M. and A. A. J. Abate, "Unsupervised Opinion Mining Approach for Afaan Oromoo Sentiments".

[24] M. O. Rase, "Sentiment Analysis of Afaan Oromoo Facebook Media Using Deep Learning Approach," New Media Mass Commun, vol. 90, no. do, pp. 7–220.

[25] Z. Qin, "Framework and practical implementation for sentiment analysis and aspect expl," 2016.

[26] A. H. and H. K. W. Medhat, "Sentiment analysis algorithms and applications," A Surv. Ain Shams Eng. J, vol. 5, pp. 1093–1113, 2014.

[27] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," Data Min. Knowl. Discov., vol. 24, no. 3, pp. 478–514, 2012, doi: 10.1007/s10618-011-0238-6.

[28] I. Liu, B., and Lane, "Attention-based recurrent neural network models for joint intent detection and slot filling," in in Proceedings of the 17 th Conference on International Speech Communication Association (San Francisco, CA), 2016, pp. l685–689.

[29] M. H. and B. Liu, “Mining and summarizing customer reviews,” in in Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Seattle, Washington D. C., USA, August 2004.

[30] J. Smailović, “Sentiment Analysis in Streams of Microblogging Posts,” Int. Postgrad. Sch. Ljubljana, 2014.

[31] K. Patel, V., Prabhu, G., & Bhowmick, “A Survey of Opinion Mining and Sentiment Analysis.,” Int. J. Comput. Appl., vol. 131(1), no. Https://doi.org/10.5120/ijca2015907218, pp. 24–27, 2015.

[32] T. D’Andrea, A., Ferri, F., Grifoni, P., & Guzzo, “Approaches, Tools and Applications for Sentiment Analysis Implementation,” Int. J. Comput. Appl., vol. 125, no. 3, pp. 26–33, 2015.

[33] M. A. . Salas-Zárate, M.P.; Medina-Moreira, J.; Lagos-Ortiz, K.; Luna-Aveiga, H.; Rodriguez-Garcia and R. J. C. Valencia-García, “Sentiment analysis on tweets about diabetes,” An Asp. approach. Comput. Math. Methods Med. 2017, 2017.

[34] A. Huq, M.R.; Ali, A.; Rahman, “Sentiment analysis on Twitter data using KNN and SVM,” IJACSA Int. J. Adv. Comput. Sci. Appl., vol. 8, pp. 19–25, 2017.

[35] W. B. Pinto, D.; McCallum, A.; Wei, X.; Croft, “Table extraction using conditional random fields,” in In Proceedings Informaion, of the 26th Annual International ACM SIGIR Conference on Research and Development in Retrieval, Toronto, ON, Canada, pp. 235–242.

[36] M. Pandey, A.C.; Rajpoot, D.S.; Saraswat, “Twitter sentiment analysis using hybrid cuckoo search method,” .Inf. Process. Manag., vol. 53, pp. 764–779, 2017.

[37] A. A. Workineh, “sentiment analysis for Afaan Oromoo Socio-politics Text,” 2019.

[38] X. W. Tao Chen a, Ruifeng Xu a b, Yulan He c, “Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,” Expert Syst. Appl., vol. 72, pp. 221–230.

[39] N. L. and B. Shen, “Aspect-based sentiment analysis with gated alternate neuralnetwork,” Knowledge-based Syst., vol. 188, no. a, p. 2020, 2020.

[40] C. C. Aggarwal, “Neural Networks and Deep Learning,” in Springer, 2018.

[41] B. hang, L.; Wang, S.; Liu, “Deep learning for sentiment analysis,” A Surv. Wires Data Min. Knowl. Discov., no. e1253, p. 8, 2018.

[42] D. Britz, “Recurrent Neural Networks Tutorial, Part 1–Introduction to Rnns.,” Available online: http://www.wildml.com/2015/09/recurrent-neural-networkstutorial-part-1-introduction-to-rnns/(accessed on 12 March 2020).

[43] Mazurowski, “Training neural network classifiers for medical decision making,” Eff. imbalanced datasets Classif. Perform., vol. 21, pp. 427–436, 2008.

[44] B. Muller, “BERT 101 State Of The Art NLP Model Explained,” Hugging Face.

[45] B. Skwarecki, “language,” https://en.wikipedia.org/wiki/Language. p. 01, 2019.

[46] M. Bulcha, “The politics of linguistic homogenization in Ethiopia and the conflict over the status of afaan Oromoo,” Afr. Aff. (Lond)., vol. 96, no. 384, pp. 325–352, 1997, doi: 10.1093/oxfordjournals.afraf.a007852.

[47] “Ethiopia,” Ethnologue. Retrieved 15 July 2021.

[48]. Komoshinii Aadaafi Tuurizimii Oromiyaa, Gumii Qormaata Afaan Oromo. Caasluga Afaan Oromo. 1996.

[49] G. D., “Afaan Oromo News Text Summarizer," Faculty of Informatics, Addis Ababa University, Ethiopia, Msc Thesis,” 2013.

[50] M. L. Kejela, “Named Entity Recognition for Afan Oromo," Department of computer Science, Addis Ababa University, Ethiopia, Msc Thesis,” 2010.

[51] D. Tesfaye, “A rule-based Afan Oromo Grammar Checker,” Int. J. Adv. Comput. Sci. Appl., vol. 2, no. 8, pp. 126–130, 2011, doi: 10.14569/ijacsa.2011.020823.

[52] W. Mekonnen, “Development of Stemming Algorithm for Afaan Oromo Text. M. Sc. Theses, Addis abeba University.,” 2000.

[53] Q. Tan, X., Cai, Y., Xu, J., Leung, H.F., Chen, W. and Li, “Improving aspect-based sentiment analysis via aligning aspect embedding,” in Neurocomputing, vol. 383, 2020, pp. 336-347.

[54] W. W. Zhiqiang Toh, “DLIREC: Aspect Term Extraction and Term Polarity Classification System,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014),, Dublin, Ireland, August 23-24, 2014, pp. 235–240.

[55] and A. I. T. . M. Abdelgwad, T. H. A. Soliman, “Arabic aspect sentiment polarity classifcation using BERT,” J. Big Data, vol. 9, no. 1, p. 115, 2022.

[56] N. L. and J. Zhao, “BERT -Based Aspect-Level Sentiment Anal ysis Algorithm for Cross-Domain Text,” Comput Intell Neurosci, no. doi: 10.1 155/2022/8726621, 2022.

[57] A. El-halees, “Arabic opinion mining using combined classification approach,” Proceeding Int. Arab Conf. Inf. Technol. Azrqa, Jordan, pp. 264–271, 2011.

[58] A. Hunegnaw, “sentiment analysis model for Afan Oromo short message service Text.machine learning approach,” TURCOMAT, vol. 12, no. 13, 2021.

[59] N. Wayessa and A. Sadik, “Multi-Class Sentiment Analysis from Afaan Oromo Text Based on Supervised Machine Learning Approaches,” Int. J. Res. Stud. Sci. Eng. Technol., vol. 7, no. 7, pp. 10–18, 2020.

[60] O. G. H. and K. K. Tune, “aspect based sentiment analysis for Afaan Oromoo movie reviews using machine learning techniques,” Appl. Comput. Intell. Soft Comput., no. https://doi.org/10.1155/2023/3462691, p. 12, 2023.

[61] S. Gebremeskel, “Sentiment Mining Model For Opinionated amharic Texts,” 2002.

[62] Tulu.T, “Opinion Mining from Amharic Blog,M.Sc. Thesis, Addis Ababa University,

Addis Ababa, Ethiopia," 2013.

[63] M. Zemenu, "Aspect Level Sentiment Analysis Using Hybrid Deep Learning Approach for Amharic News Comments," 2022.

[64] . Vanaja, "Aspect-level sentiment analysis on E-commerce Data," in in Proceedings of the 2018 International Conference on Inventive Research in Computing Applications, Coimbatore, India, pp. 1275–1279.

[65] G. G. Eggi, "Afaan Oromo Text Retrieval System,MSc Thesis, School of Information Science," Addis Ababa University, 2012.

[66] Y. Kim, "Convolutional Neural Networks for Sentence Classification," https://arxiv.org/abs/1408.5882).

[67] Z. Vujovic, "Classification Model Evaluation Metrics.," Inter- Appl. Natl. J. Adv. Comput. Sci., vol. 12, no. 6, p. 6, 2021.