Jimma University

Jimma Institute of Technology

Faculty of Computing and Informatics

Department Of Information Technology

Automatic Detection of Extremist Affiliations Using Deep Learning-Based Sentiment Analysis on Social Media Posts for Afaan Oromo.

By: Amanuel Solomon

A Thesis Submitted to Faculty of Computing and Informatics to Jimma University for Partial Fulfillment for the Degree of Master of Science in Information Technology.

Jimma, Ethiopia

November 2021

JIMMA UNIVERSITY
JIMMA INSTITUTE OF TECHNOLOGY
FACULTY OF COMPUTING AND INFORMATICS
DEPARTMENT OF INFORMATION TECHNOLOGY

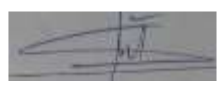
Automatic Detection of Extremist Affiliations Using Deep Learning-Based Sentiment Analysis on Social Media Posts for Afaan Oromo.

BY: -Amanuel Solomon

Name And Signature of Members of Advisors

Main Advisor: Dr. Teklu Urgessa		**26/11/2021**
	Signature	**Date**
Co-Advisor: Mr. Abaynew Guadie		**26/11/2021**
	Signature	**Date**

Signed by the Examining Committee: -

1. Worku Jifara Sori(Ph.D.)		**20/12/2021**
External Examiner	**Signature**	**Date**
2. Mr.Worku Birhanie(Msc)		**20/12/2021**
Internal Examiner	**Signature**	**Date**
3. Mr. Hambisa Mitiku(Ass.profesor)		**20/12/2021**
Chair Person	**Signature**	**Date**

Declaration

I, the undersigned, declare that this is my original work, that it has not been submitted as part of a degree requirement at any other university, and that all sources of materials included in the thesis have been properly acknowledged.

Research Thesis Submitted By:

Amanuel Solomon	______________________	20/12/2021
	Signature	Date

Dedication

Dedicated to my father and mother, whom I adore and for their endless love, encouragement and strength throughout my study from the early beginning!!!!

Acknowledgement

First and foremost, I would like to thank my almighty God for giving me the strength, knowledge, ability and opportunity to undertake this research study and to persevere and complete it outstandingly. Without his blessings, this achievement would not have been possible.

Next, my deepest gratitude goes to my advisors **Dr. Teklu Urgessa** and **Mr. Abaynew Guadie** for their continuous follow-up, guidance, suggestions and constructive comments were very essential to complete my thesis work.

Many friends have helped me stay rational through these two years. I greatly value their friendship and deeply appreciate their belief in me. I have to give a special mention for my best friends Lalisa Tadesse, Tadese Fikadu, Ashenafi Hailu, Diriba Girma and all my class mates whose name are not mentioned here. I have no word to express my feeling for their endless help to pull off my study for the past two years. They are unrevealed person for my success.

Most importantly, none of these would have been possible without the love and patience of my family. Over the years, my father and mother, to whom this thesis is dedicated, have been a continual source of love, concern, support, and strength. In addition, I would like to express my heart-felt gratitude to my brother: Abreham Solomon who have aided and encouraged me throughout this study.

Finally, I share the credit of my work for the rest of all my brothers, sisters, friends and staff members of Mettu University, Bedelle campus that I have not mentioned their name here.

Abstract

In recent years, the Internet has become a global platform for communicating and disseminating information. Today, popular social media sites have a huge global reach and audience with many daily active users who interact and communicate through those social networks far apart from each other physically. Because there is no limitation that guides the users of the platform such as Facebook, twitter and YouTube, all users freely express their feelings, attitudes, beliefs and opinions towards various issues such as politics, social, religious, ethnic, business and etc. As a result, large volume of opinionated texts is produced on a regular basis; the comments might be positive or negative. It's impractical and tedious to manually classify each and every review with human power into different class such as extreme, anti-extreme and neutral. Classifying the sentiments is used for the individuals, company as well as the government to make decision and action on the posts and comments to be removed that contains radical contents that may cause offence and conflict among the people. In the proposed work we designed and implemented deep learning-based sentiment analysis model for Afaan Oromo review. Dataset from three domains which is politics, religious, and ethnic is collected for the model. Totally 2410 reviews which is labeled with three classes extreme, anti-extreme and neutral labeled with 0,1, and 2 respectively prepared for the mode. Preprocessing techniques such as stop-word removal, tokenization, lower case conversion and other preprocessing techniques are used for cleaning the dataset. In the proposed model sentence level sentiment analysis is used. We implemented convolutional neural network (CNN) and long short memory (LSTM) with word2vec that classify the reviews into extreme, anti-extreme and neutral. The proposed model convolutional neural network achieve accuracy of 80% and long short-term memory achieved performance of 70%. According to our experiment convolutional neural network is good in classifying the review to the target class than the long short-term memory.

Keywords: Convolutional Neural Network, Sentiment Analysis, Word2Vec, Deep Learning Classification Mode, Afaan Oromo.

Table of contents Page

List of tables

List of figures

List of acronyms

NLP: Natural Language Processing

NLTK: Natural Language Toolkit

RNN: Recurrent Neural Network

ML: Machine Learning

NB: Naïve Bayes

RNN: Recurrent Neural Network

SA: Sentiment Analysis

SVM: Support Vector Machine

OP: Opinion Mining

ISD: Institute for Strategic Dialogue

OBN: Oromia Broadcasting Network

FBC: Fana Broadcast Corporation

VOA: Voice of America

CNN: Convolutional Neural Network

LSTM: Long Short-Term Memory

R: Radical

NR: Non-Radical

I: Irrelevant

HTML: Hyper Text Markup Language

ASBA: Aspect Based Sentiment Analysis

CHAPTER ONE: INTRODUCTION

This chapter discusses about basic definition of sentiment analysis, the objective, problem statement, methodology and contributions of this research work.

1.1. Background

In recent years, the internet has become a global platform for communication and dissemination of information. Today, popular social media sites have a huge global reach and audience, with Facebook having more than 2.80 billion monthly active users as of December 31, 2020, an increase of 12% year-over-year[1], while YouTube currently has 2 billion users. Up from 800 million users in 2012. 42.9% of all global internet users access YouTube monthly. Daily Active Users (DAUs): More than a billion hours of content is consumed on YouTube every single day[2]. Similarly, Twitter currently has 396.5 million users Twitter brought in $3.72 billion in revenue last year(2020) and it increased with 14% per day over the last few years [3]. There are 4.48 billion people currently use social media worldwide, up more than double from 2.07 billion in 2015. The average social media user uses 6.6 different social media networks on a daily basis. Since 2015, the average year-over-year growth rate for social media has been 12.5%. However, with a 9.2 percent growth rate in 2019-2020, growth is slowing [4]. Some other popular social media platforms include Instagram, LinkedIn, Tumblr, WeChat and WhatsApp. The registered number of total users on these social sites is in billions.

According to worldwide social media reports, 4.55 billion people used social media in October 2021, accounting for 57.6% of the global population[5]. Nowadays, due to the extensive use of social media, a large volume of data is generated on a daily basis. With numerous microblogging and social networking services creating vast amounts of data, social media has been rapidly developing[6].

The data generated from these sites is in an unstructured form that creates challenges for processing and analysis to make a useful understanding of the underlying hidden patterns which will be useful for decision making. Data gathered from the social media sites are used for various purposes in different contexts of individuals, business organizations, public offices as well as societies. The purposes may include;

- Individual decision making for interaction and collaboration

- Marketing & Promotions
- Recommendation system
- Sentiment analysis
- Review system
- Fraud detection
- Crime monitoring
- Terrorism/Extremism detection

The high rate of growth of the Internet and the wide users of social media, such as Facebook, Twitter, Instagram, and YouTube, has brought many new opportunities for people to express their attitudes and feelings towards any individual, services, organizations, political parties, government policy, religious beliefs ,ethnic issues by using their own native languages[7]. As the numbers of users on social media are increasing day by day, the cases of extremisms on social media are also increasing. Because social media provides a convenient way to people where they can share their ideas and beliefs with others through posting and commenting various topics such as social, politics, economic religious and etc. People are using these websites to promote their businesses and spreading news and useful information and also use for disseminating radical contents. Twitter tweets, Facebook posts and YouTube comments are also used for opinion mining or sentiment analysis. Extremists are using social media to propagate their beliefs, recruiting new people in their groups, raising funds and planning operations. Extremist's groups are misguiding and brain washing people by propagating their ideologies through extreme or radical contents on social media. Social media provides an easy way to them to interact with large number of people without physically present there[8].

The users of social media usually write on numerous topics that reflects their opinions and thoughts towards those various issues posted on social media either their thoughts are positive or negative. As well many users are abusing the capability of this social media to spread distorted beliefs and negative influence to other users. These include terrorism, politics, religions, fraudsters, ideology and others[9].

Because there is large volume of data generated daily, it's impractical to classify sentiments with human powers. However, great efforts are needed to develop algorithms and machines that can imitate the natural ability of human beings to understand emotions of human beings, analyze situations and understand the sentiments associated with the context. The sentiment analysis is an effective mechanism to explore the socio-economic or demographic influence in human reciprocation [10].

Social Media has also affected the way people socially interconnected, communicate, interact and opinionize. The advancement in technology has increase communication and dissemination of information. Unconditionally , many terror group communities and individuals have started consolidating a virtual community online for various purposes such as recruitment, fund raising , online donations, targeting youth online and spread of extremist ideologies which may cause instability among the peoples of the country [11].

In our country context, the rise of intolerant groups and individuals using social media to publish written texts, images and stories about vulnerable groups creates a special problem for media ethics. Extremism leads to serious media harm, including unjustifiable profound offence. The more that extreme and radical messages are circulated, the greater likelihood that citizens, frustrated by slow-moving moderate politics, may adopt more extreme “solutions” to complex problems[12].

However, electronic communication has become an important part of our lives and at the same time has changed our lives so dramatically that it would be strange if the communication behavior of extremists had not also changed[13]. And most of the time the word extremism and hate speech are taken as similar terms but both the words highly different. According to[13] the ISD defines extremism as the advocacy of a system of belief that posits the superiority and dominance of one ‘in-group’ over all ‘out-groups’, propagating a dehumanizing ‘othering’ mindset that is antithetical to the universal application of human rights. Extremist groups advocate, through explicit and subtler means, a systemic change in society that reflects their worldview. Extremists try to be present wherever people obtain information and communicate. Because of this, over the past few years, the new media ecosystems which resulted from social media played an increasingly important role in spreading extremist propaganda.

Extremists increasingly concentrated on optimally interlinking their online communication methods with their offline activities in order to maximize their circulation and effectiveness.

In [13]the Council of Europe defines hate speech as: “All forms of expression which disseminate, incite, promote or justify racism, xenophobia, anti-Semitism or other forms of intolerance based on hate, including intolerance which is expressed in the form of aggressive nationalism and ethnocentricity, discrimination and hostility to minorities, migrants and people with a migrant background."

Due to the law enforcement from the start, tech companies' commitment to free expression admitted some exceptions. Child pornography, spam, phishing, fraud, impersonation, and copyright breaches were all prohibited under the terms of service and community guidelines. After extensive conversations with advocacy groups, threats, online stalking, non-consensual pornography, and hate speech were made illegal. The goal was to establish a good balance between free speech and abuse protection while maintaining platform market share. More recently, social media companies have revised their speech policies concerning extremist and hateful expression[14].

As stated in [14]on May 31, 2016, Facebook, Microsoft, Twitter, and YouTube entered into an agreement with the European Commission to remove "hateful" speech within twenty-four hours if appropriate under terms of service. The same firms announced plans for a shared database of banned extremist content for assessment and removal elsewhere six months later.

In today's world there is hot research area in the field of artificial intelligence specially working on natural language processing (NLP). Natural language processing is a linguistics and artificial intelligence sub-field that refers to any attempt to process natural language using computers. It investigates the issues of automated language generation and comprehension in natural human languages. The major fields of interest in NLPs are speech recognition, speech synthesis, machine translation, text categorization, natural language understanding, spell checker, Sentiment Analysis (SA) or Opinion Mining (OM) and the like[15]. Sentiment analysis is computational research in the field of text mining that learns about a text's idea or viewpoint, sentiment, emotion, and even attitude. The fundamental goal of sentiment analysis is to determine the document's polarity through sentiment classification[16]. With the widespread usage of social networking sites such as Twitter and Facebook, the online community is exchanging information in the form of ideas, thoughts, emotions, and intents, all of which indicate their connections and attitude for a particular entity, event, or policy. Sentiment analysis (also known as opinion mining) is a natural language processing technique for determining the positive, negative, or neutral nature of the documents or texts using various machine and deep learning algorithms. Sentiment analysis is commonly used on textual data to help businesses analyze brand and product sentiment in customer feedback and better understand client expectations. Opinion mining or sentiment analysis is a technique to detect and extract subjective information in text documents.

The techniques used for sentiment analysis can be used for extremist text post detection. Sentiment analysis is popularly also known as opinion mining. The general purpose of sentiment analysis is to automatically classify a text as neutral, positive or negative. In sentiment analysis sentiments, opinions and attitude are analyzed through the texts. Sentiment analysis is usually done to know the sentiments and opinion of people towards a particular entity. Therefore, sentiment analysis is a way to extract hidden patterns from opinion of people towards the policies of government that may be helpful for a political party to design its strategies and opinion on a particular product (positive or negative) may be used by manufacturing companies for making crucial decisions on their product and goods. In this section approaches used for sentiment analysis is discussed as the problem extremist post detection is the problem of classifying a text document as extremist ,anti-extremist and neutral which is quite similar to sentiment analysis in which a text document is classified as positive, negative or neutral on the basis of its meaning[8]. Machine learning algorithms which are used for general text classification can be employed for sentiment analysis and extremist post detection.

According to various reports on social media large volume of unstructured data is generated on daily basis which is in practical, time consuming and tedious to classify this large volume of data with human power manually. The growing use of social media to communicate opinions necessitates a focus on natural language processing and machine learning research to develop approaches that can assist government authorities in automatically identifying people with extreme or radical views. Such measures will undoubtedly aid law enforcement authorities in their attempts to combat crime, as the large prevalence of such content on social media is one of the most pressing worries today. The difficulties raised above highlight the importance of looking into online extreme content as a research topic.

In general, sentiment analysis tries to determine the sentiment of a writer about some aspect or the overall contextual polarity of a document. The sentiment may be his or her judgment, mood or evaluation. A key problem in this area is sentiment classification, where a document is classified as an extremist, anti-extremist and neutral. In this research work we are going to develop a deep learning model to detect extremist-based text from the social media posts and comments for Afaan Oromo.

1.2. Motivation

A social networking sites allows users to create, share, and influence others through the creation, sharing information, and ideas. With the proliferation of these social platforms, more emphasis is being placed on the extraction of information from these communications. As a result of the large number of Afaan Oromo users on social media, it is now necessary to comprehend people's emotions, particularly in socio-politics, religious, and ethnic reviews.

Several considerations led to the decision to work with the Afaan Oromo language. First, people are born with the ability to understand and interpret emotions, as well as the sentiments that go along with them. However, from the standpoint of the Afaan Oromo language, how effectively can we educate a machine to display the same behavior becomes an important and vital subject to be explored. Second, Afaan Oromo sentiment analysis is important because of its already vast audience and large amount of data available on the internet. Finally, while most of the tools are available for English and other languages, it is critical to develop a model for Afaan Oromo to classify sentiment of text in order to reduce the work of analyzing, suggesting, and providing crucial information for the community and motivated us to take this problem as our research work.

1.3. Statement of the problem

World Wide Web allows Internet users to collaborate and share information online, and therefore create large virtual societies. Afaan Oromo users of social network sites Facebook, Twitter and YouTube generate daily a large volume of Afaan Oromo textual reviews related to different social, political, religious, ethnic and scientific subjects.

Nowadays social medias like Facebook have a mechanism to detect and remove hate speech from their network but those posts and comments that have hate content should be reported by the network users and based on the report given by the users they remove post but it is possible to develop automatic detection of extremist content post by using sentiment analysis specially for local languages like Afaan Oromo to detect and remove the content from their network since it’s a critical issue in current social media ecosystem where extremist based content causing conflicts and profound offence are increasing.

The number of social media users are increasing rapidly and produce large volume of reviews on different issues. This large volume of different Afaan Oromo textual reviews cannot be analyzed manually. Due to its large in volume the data produced its impractical to classify the reviews manually to drive decision. The algorithmic recognition of emotional expression has been proven using sentiment analysis to classify the textual documents to the correct class by using machine learning and deep learning algorithms. Manual classification is impractical due to the large volume of extremist information available on the internet. However, terrorist, extremist groups and individuals also adopt using social media for different functions including dissemination of information which can cause a conflict between peoples. The post may include religious-extremism, political-extremism and ethnic-extremism which leads different parts of the community in to conflicts. Incase Internet becomes a national security hazard in such situations which is extremist content is radically increasing daily. Countering the use of the Internet for extremism or terrorism is a priority for many laws enforcement and intelligence agencies. Classification and identification of social media text with radical contents is one technique to accomplish this objective. However, because the vast amount of information on the Internet makes it difficult for humans to classify all of it, machines are needed to assist with automatic classification. As we reviewed different research done on sentiment analysis with Afaan Oromo and other local languages use machine learning for sentiment analysis with various approaches but nowadays deep learning shows promising result in sentiment analysis so that deep learning approach will used as method for the above state of art. Sentiment analysis for the mentioned problem above is not deal with this topic as per the researchers knowledge and social media extremist post are radically increasing from day to day which cause conflicts among the society therefore it is good to design a deep learning approach to detect extremist post and comments from the social media to make the social media environment for a peaceful communication only. Automated methods for overall review and classification of Web documents are necessary to give insights and aid law-enforcement agencies in decision-making as well as for the social media company to improve their user guidelines on the users posts and comments specially for Afaan Oromo language. In the proposed approach sentence-level sentiment analysis will be used because mostly network user posts their ideas, opinions and attitudes in sentences as well as we can get data set for our model easily. Deep learning convolutional neural network developed with Word2Vec embedding approach is used for the state-of-the-art problem.

1.4. Research questions

At the end of this study the following research question will be answered by studying and uses different methodologies to realize the research work.

Research question answered with the proposed work are:

- What are the challenges of developing Afaan Oromo opinion mining (sentiment analysis) model for detecting extremist-based content using deep learning approach?
- What are the preprocessing steps used for data cleaning for Afaan Oromo extremist review sentiment analysis?
- Which deep learning Algorithm is appropriate for extremist sentiments analysis for Afaan Oromo?

1.5. Objective of the study

1.5.1. General objective

The general objective of this research is to investigate and develop a deep learning-based sentiment analysis for automatic detection of extremist affiliation for Afaan Oromo on social media posts.

1.5.2. Specific objectives

In order to achieve the general objective of the study, specific objective are formulated as follows:

- To review literature so as to have conceptual understanding identify related works done in the area of sentiment analysis.
- To analyze the general structure of Afaan Oromo statements related to sentiments such as identifying extremist, anti-extremist and neutral statements.
- To analyze challenges in developing Afaan Oromo reviews.
- To collect representative sample datasets taken from social as well as determine the opinions polarity classification of the review for the model.
- To develop sentiment analysis model for Afaan Oromo.
- To evaluate the performance of the model constructed using validation and test metrics.
- To forward the results obtained and show directions for future studies on Afaan Oromo sentiment analysis.

1.6. Scope and Limitation

The main aim of this study is to develop sentiment mining and sentence-level opinion mining review for Afaan Oromo language. This research will be conducted based on acquired data from different social media post and comments which contains extremist content.

Opinion holders like individuals or groups can express their opinion on a given posts in different mechanisms like text, voice, images, videos from these mechanisms we process text-based user generated content only. As well as different emojis which can express different feeling is not consider we only use text data. The research work is limited to sentiment polarity mining (only extremist, anti-extremist and neutral classification) from social media reviews and opinions such as Facebook, YouTube and twitter provided in Afaan Oromo language.

1.7. Research design

In this research work we followed design science research methodology. Design Science Research (DSR) is a problem-solving paradigm that aims to improve human knowledge through the creation of novel artifacts. Simply put, DSR aims to improve technology and science knowledge bases by creating innovative artifacts that solve problems while also improving the environment in which they are implemented[17].

Research methodology is the path through which researchers need to conduct their research. It demonstrates how these researchers construct their problem and purpose, as well as how they present their findings based on the data collected during the study period. This chapter on research design and technique also explains how the final research result will be obtained in accordance with the study's goal. The research approach can supports the researcher on how to come across the research result findings[18]. So, in the proposed research work we follow the design science research process to reach the final result of our work.

We follow all the steps like problem identification, setting the objective of the work, design and development of our model, demonstration, and finally we evaluate the proposed work with different parameters to check the accuracy of the model.

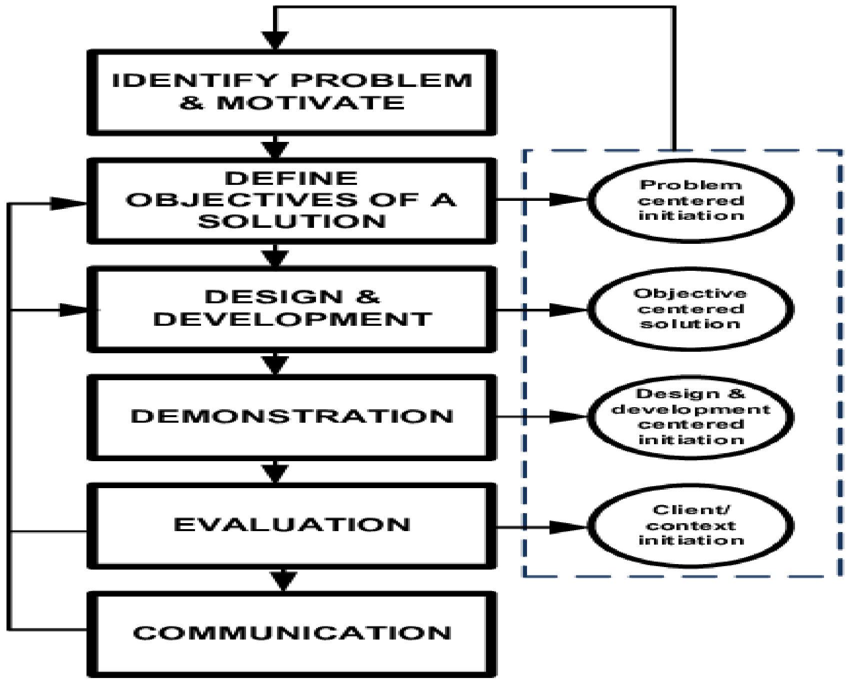


Figure 1:Design science research methodology.

1.8. Problem Identification

We have conducted extensive literature review prior of any other activity in order to identify the problems, knowledge gaps, and state of the art of sentiment analysis, then after reviewed related work we identified the gaps and setting the objective of solution. Hence, we decided to build convolutional neural network architecture to classify the sentiment expressed in a given Afaan Oromo text which passed through many phases of development task. In the first phase, collect or extract the opinion from official Facebook pages, twitter and YouTube platform. Next, we understand the nature of extremist or radical review in context with an Afaan Oromo language structures and labelling the text. Natural language processing techniques, like pre-processing to reduce data noise or irrelevant information from training is also employed. Since it is from Facebook, twitter and YouTube social media we gave more attention on data cleaning. Then vocabulary was constructed from the data set prepared for training and testing the model. Moreover, designing an algorithm, implementing with programming language and training was done. In the final phase implementation and evaluation of the algorithm was done to finalize the research objective.

1.9. Design Approach

To model sentiment classification of Afaan Oromo text, we used deep neural network with word embedding. To realize this, data preparation like separation of data into training and test sets, loading and cleaning the data to remove punctuation and numbers, and defining a vocabulary of preferred words are play vital role for this work. Then, we train the model using the Keras deep learning library with Convolutional Neural Network (CNN) as it confirmed to be successful at classification problems.

1.10. Implementation Tools

In order to meet the objective of the study, we used a number of environments and tools. The main tools that we used in accompanying this research are excel and notepad to store data collected by manually; online tools export comment website and Face pager API. We collected the data set from official pages which are known under the terms and services of social media service providers those are Voice of America Afaan Oromo (VOA), Oromia Broadcasting Network (OBN) And Britain Broadcasting Corporation Afaan Oromo (BBC). For implementing our proposed model, we used Google colab online tool for experimentation. Some of the library used are also Keras, pandas and NLTK for preprocessing and cleaning.

1.11. Significance of The Study

Sentiment analysis is nowadays an increasingly and growing interest from the natural language processing researchers, which is particularly motivated by the widespread need for an opinion-based applications, such as social media post and comment review, political reviews, product reviews, market research, and analysis and opinion summarization. In addition, being an academic exercise to fulfill the requirement for the program, this research is believed to produce a prototype system on the opinionated Afaan Oromo texts model developed based on deep learning techniques that can help the government of the country and social media companies to improve their user guideline standard to remove extremist posts and comments from the online platform. The research combines contemporary theories and methods in linguistics to develop sentiment mining model for opinionated Afaan Oromo texts an automated research tool for rich text-mining. The research gives new insights into the complexity of language use, the linguistic modeling of subjectivity and the representation of this knowledge in a lexicon. It will also shed new light on the complex dimensionality of opinionated Afaan Oromo texts or reviews. In the current business and political situations, knowing what other people think is a determinant factor in decision making.

Extremist speech detection implemented at feature level would create a monitored environment that no extremist speech is tolerated and exercised; creating a safe and tolerable web environment where everyone respects and interacts peacefully. Extremist speeches that are tagged would be investigated and people exercising it would become accountable for their actions which may create catastrophic results. In the long run, extremist-free speech would build societal unity and peace for all people worldwide. It is also meant to create awareness about the seriousness of the action and its dangerous outcome. And motivate researchers on the area to further investigate sentiment analysis with various approaches using different local languages.

1.12. Thesis Structure

Relevant background theory and fundamental concepts involved in the sentiment analysis/classification, in addition to fundamental concepts of sentiment analysis, Afaan Oromo language and literature review is discussed in chapter 2. Chapter 3 describes the related work done on sentiment area with various approach as well for international language and local languages are briefly discussed. Chapter 4 describes the detailed implementation of the model, what the architecture of the system looks like and different approach used for implementing the state of art. Related technologies like tokenization, word embedding, and Convolutional Neural network, pooling, softmax, evaluation parameters will be presented. Experiments conducted on these systems and their results are presented in Chapter 5. Finally, Chapter 6 conclude the study and suggested future work.

CHAPTER TWO: FUNDAMENTAL CONCEPTS AND LITERATURE REVIEW

2.1. Introduction

Thanks to the tremendous growth and popularity of social media and social networks, people have never had greater opportunity to express and share their thoughts, views, opinions, and feelings about nearly everything. They can do so through personal webpages and blogs, as well as social networking sites such as Facebook, Twitter, YouTube, and Blogger. An opinion is an individual's private state; it denotes the personality's assessments, evaluations, emotions, beliefs, judgments and ideas concerning a particular item/subject/topic.

Sentiment analysis is the field of data mining and natural language processing that study about analyzes the opinion data from social media like Facebook, Twitter, YouTube, organization sites, online news report, user reviews and etc. It is the process of classifying opinions under defined polarities classes. Sentiment analysis is a hot research area in the field of data mining and natural language processing[19].

Sentiment analysis is a rapidly emerging topic at the crossroads of linguistics and computer science that aims to automatically assess the sentiment, or positive or negative opinion, present in text. Positive or negative appraisal expressed through words is referred to as sentiment. Automatically determining if a review made online (political, religious, ethnic, movie, book, or consumer product) is favorable or negative towards the item being evaluated is one of the most common applications of sentiment analysis. Sentiment analysis is now a standard tool in the arsenal of firms, marketers, and political analysts who conduct social media analysis. Sentiment analysis research derives information from the context of positive and negative words in text, as well as the text's linguistic structure[20].

2.2. Sentiment analysis approaches

Sentiment analysis can be performed at several levels or approaches, including fine-grained sentiment analysis ,emotion detection, document level, phrase level, and aspect/feature level. Sentiment is derived from the full review in this procedure, and an entire opinion is categorized based on the sentiment of the opinion holder as a whole. Sentiment analysis models examine polarity (positive, negative, and neutral), as well as feelings and emotions (angry, joyful, sad, etc.), urgency (urgent, not urgent), extremism (extremist vs anti-extremist), and even intents (interested vs not interested).

we can construct and customize our categories to match our sentiment analysis needs based on how we wish to interpret client comments and enquiries. Meanwhile, here are some of the most commonly used types of sentiment analysis:

2.2.1. Fine-Grained Sentiment Analysis Approaches

The majority of previous sentiment analysis (SA) work for customer feedback analysis has focused on binary categorization of customer evaluations (i.e., positive or negative), with less attention paid to fine-grained sentiment classification. Reviews are divided into many sentiment groups using fine-grained sentiment classifications, such as weak, moderate, strong, or extremely strong[21].

If polarity precision is important to the business review, it is possible to expand polarity categories to include:

- Very positive
- Positive
- Neutral
- Negative
- Very negative

Fine-grained sentiment analysis can be used to analyze 5-star ratings in a review, for example: 5 stars for Very Positive and 1 star for being extremely negative.

2.2.2. Emotion Detection Approach

Emotion detection (ED) is a type of sentiment analysis that involves extracting and analyzing emotions. The rise of Web 2.0 has pushed text mining and analysis to the forefront of business success. It enables service providers to give customers with custom-tailored services. Many research are being carried out in the field of text mining and analysis due to the simplicity with which data can be sourced and the numerous benefits that its deliverables to provide[22].

Sentiment analysis of this sort seeks to detect emotions such as happiness, frustration, rage, sadness, and so on. Many emotion recognition systems rely on lexicons (lists of words and the feelings they evoke) or sophisticated machine learning techniques. People communicate emotions in a variety of ways, which is one of the drawbacks of employing lexicons. Some words that are commonly used to communicate anger, such as bad or kill (for example, your product is so horrible or your customer service is killing me), can also be used to express satisfaction (e.g., this is bad ass or you are killing it).

2.2.3. Aspect-Based Sentiment Analysis Aprroach

Aspect-based sentiment analysis (ABSA) is a more difficult process than standard sentiment analysis at the text level. It is concerned with identifying the traits or aspects of an entity mentioned in a text, as well as the sentiment expressed toward each of these aspects[23].

When evaluating text sentiments, such as product evaluations, we typically want to discover which specific parts or features people are citing in a good, neutral, or negative light. An aspect-based classifier would be able to determine that the sentence indicates a negative judgment about the feature battery life in this text:

Example:

"The battery life of this camera is too short."

The above given opinion is given about the quality of the camera. Since it is given towards a single entity this sentiment analysis type is aspect based sentiment analysis. Aspect based sentiment analysis always used to describe the negative or positive thoughts towards single entity.

2.2.4. Document Level Classification Approach

The sentiment of the entire review is extracted in this procedure, and a full opinion is categorized depending on the overall sentiment of the opinion bearer. The objective is to determine whether a review is positive, negative, or neutral[24]. Document level sentiment analysis used for extracting the overall document piolarity whether it is positive, negative or neutral based on the contant of the document towards different entity. And it is good if the document have a single polarity than mixed polarity. The goal is to determine if an entire opinion document reflects a favorable or negative attitude. Document level sentiment analysis evaluates if a product review indicates an overall good or negative judgment about the product[25].

Example:

“I bought an iPhone a few days ago. It is such a nice phone, although a little large. The touch screen is cool. The voice quality is clear too. I simply love it!”

“Bilbilla ayfoonni haraan bittadhe guyya murasa dura. Bilbilichi bayye garidha, akkasuma ille guddadha. Fodaan do’i issas garidha. Qulqulinni sagalee issas akkasuma. Kanafu bayyen jaladhe!”

Document level classification works best when the document is written by a single person and expresses an opinion/sentiment on a single entity rather than mixed polarity.

2.2.5. Sentence Level Sentiment Analysis Approach

Sentence-level sentiment analysis is one of the main directions in sentiment analysis area. The existing work on the task concentrated on recognizing the polarity of a sentence (e.g., positive, neutral, negative), according to semantic information learned from the textual content of sentences. Earning high quality word embedding is still challenging. Matthew E proposed a deep bidirectional language model (biLM) to encode various types of syntactic and semantic information about words in-context[26]. Sentence level analysis is a task that examines individual sentences to assess whether they communicated a favorable, negative, or neutral viewpoint. In most cases, neutral means that the review don't have an opinion. This level of analysis is linked to the subjectivity categorization, which distinguishes between sentences that represent factual information (called objective sentences) and sentences that express subjective ideas and opinions (called subjective sentences)[25]. This task is commonly defined as classifying a given text (usually a sentence) into one of two classes: objective or subjective. A sentence is said to be subjective if it contains non-factual information such as personal opinions, predictions and judgements. So usually sentence level sentiment analysis uses non-factual review because it’s a subjective one that everyones feelings ,atitudes, and opinion towards different issues.

Example:

- “COVID-19 vaccines are dangerous and it’s risky to get it at early stages of development. The side effects are deadly.”

"Talallin dhibee COVID-19 bayye hamaa fi talalli yalii irra jiru fudhachun midha gudaa fida."

This above example is not factual truth rather it is a personal opinion towards the COVID-19 vaccine so that we can say this is a subjective sentence which used for sentiment analysis.

A sentence is objective if it contains facts rather than opinions, attitudes feelings, means those sentences are facts which cannot changed ever. Objective sentence always stands for general truth that never changed and do not used for sentiment analysis.

Example:

- The sun rises in the east.
  "Aduun karaa baha batti."

The above given example is a general fact than personal ideas so this kind of objective sentences are not used for sentiment analysis as a whole. We clearly reviewed that objective sentence are not preferred for sentiment analysis.

Because most polarity identification techniques are specialized for discriminating between positive and negative text, subjectivity detection is an important subtask of sentiment analysis. As a result, subjectivity detection ensures that factual data is filtered away and only opinionated data is sent to the polarity classifier. In addition, subjective extracts account for only 60% of the evaluation and provide the same polarity findings as complete text categorization[27]. As stated on the state of art in this research work, we are going to use sentence level sentiment analysis and subjective based sentences are used due to the availability large volume of sentence-based data which is daily produced by the users of social media that shows their opinion, attitude, or feelings towards different issues. Therefore, in the proposed model we implemented sentence level sentiment analysis for Afaan Oromo opinionated sentences that have radical content related to religious, politics and ethnic topics to classify the review polarity as extreme, anti-extreme and neutral.

2.2.6. Feature Level Analysis Sentiment Analysis Approach

Both the document and phrase level analyses fail to reveal exactly what people liked and disliked. Finer-grained analysis is performed at the aspect level. Previously, aspect level was referred to as feature level (feature-based opinion mining and summarization). Aspect level examines the opinion itself rather than language constructs (documents, paragraphs, sentences, clauses, or phrases). It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion)[25].

Sentiment analysis is a field of study that examines one's thoughts, feelings, sentiments, evaluations, attitudes, and subjectivity towards entities and characters described in review. With the advancement of the Internet, an increasing number of network users are likely to publish their views on the internet, attracting significant attention not just in academia but also in business and society. For example, businesses desire to acquire real-time feedback on their products or services by assessing client comments. And most of the time users of social media write their thoughts, opinions, feelings and attitudes in the form of sentence. So that we proposed sentence level sentiment analysis for Afaan Oromo extremely affiliated reviews.

2.3. Deep Learning Techniques For Sentiment Analysis

Deep learning is gradually becoming the leader in the field of machine learning, which is entering its golden age. To develop computer models, deep learning employs numerous layers to represent data abstractions. Deep learning methods like generative adversarial networks, convolutional neural networks, and model transfers have totally transformed our understanding of information processing[28]. Deep learning (also known as deep structured learning) is a type of machine learning technology that uses artificial neural networks to learn representations. There are three types of learning: supervised, semi-supervised, and unsupervised.

2.3.1. Supervised Deep Learning Approach

Supervised learning is one of the deep learning techniquse used for sentiment analaysis based on a data sample from data source with corrct classification already assigned .Such techniques are utilized in feedforward or Multilayer Perceptron (MLP) models. A supervised learning algorithm learns from labeled training data, helps to predict outcomes for unforeseen data. In a supervised artificial neural network model, learning is accomplished by training, which is also known as the error back-propagation algorithm[29]. The error correction-learning algorithm trains the network using input-output samples, finds the error signal, which is the difference between the calculated and desired output, and adjusts the synaptic weights of the neurons, which are proportional to the product of the error signal and the input instance of the synaptic weight. In our proposed model we used supervised leaning approach for the developed model to classify the review polarity into extreme ,anti-extreme and neutral classes.

2.3.2. Unsupervised Deep learning Approach

Unsupervised learning has no labeled data set and consequently no learning based on feedback, whereas supervised learning has labeled data that is used to train the network. Unsupervised learning involves pre-training neural networks with generating models such as RBMs, which can then be fine-tuned using normal supervised learning techniques. It's then applied to a test batch of data to find patterns or classifications. With its sheer volume and variety of data, big data has pushed the envelope for deep learning even farther. Contrary to popular belief, there is no strong consensus on whether supervised learning or unsupervised learning is superior[30]. To find latent patterns in unlabeled input data, self-organizing neural networks use an unsupervised learning technique. The ability to learn and organize information without providing an error signal to evaluate the proposed answer is referred to as unsupervised learning[29]. The lack of direction for the learning algorithm in unsupervised learning can sometime be advantageous, since it lets the algorithm to look back for patterns that have not been previously considered.

Deep learning has grown ingrained in our daily lives, from search engines to self-driving automobiles that require a lot of processing power. It has already had a significant impact in fields including cancer detection, precision medicine, self-driving cars, predictive forecasting, and speech recognition, sentiment analysis ,image processing ,text classification among others[30]. Some of the algorithms such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks, long short term memorys and convolutional neural networks, have been used in fields such as computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, material inspection, and board game programs, with results that are comparable to, if not better than, human expert performance[31].

Deep learning is just a subset of machine learning. Deep learning is, in reality, a type of machine learning that works similarly to traditional machine learning (hence why the terms are sometimes loosely interchanged). However, its capabilities are different. Deep learning performs better in large data set corpus but in case of machine learning with small data we can achieve higher performance as shown in the below graph.

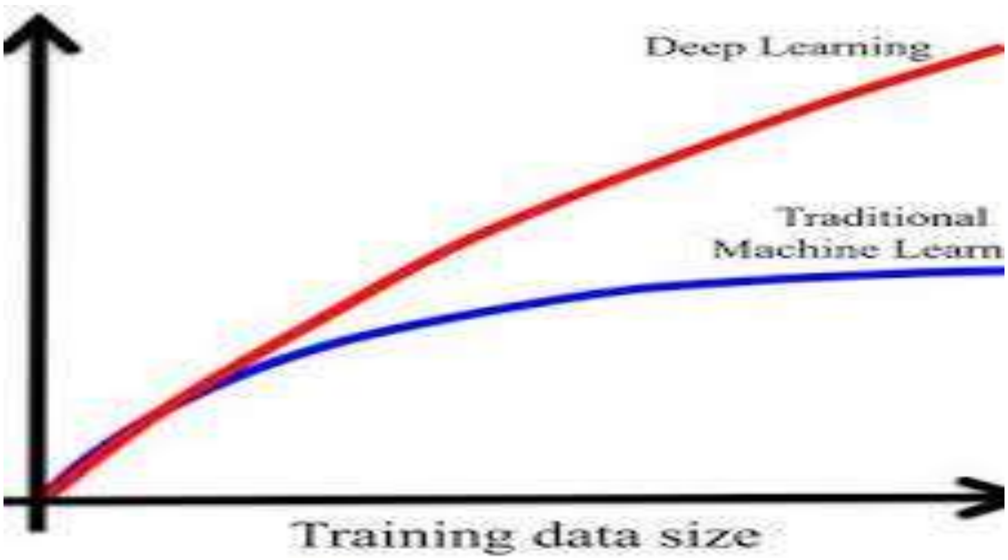


Figure 2:Deep learning and machine learning accuracy on data size

The above graph is the representational form of deep learning and machine learning algorithms accuracy based on the volume of data used for training purpose. In deep learning case if the training dataset is large in volume the accuracy become increase while machine learning achieves best accuracy with small volume of dataset.

2.3.3. Deep Neural Network(DNN) and Its Architecture

A deep neural network is made up of multiple layers of nodes. To solve challenges in various sectors or use-cases, many designs have been devised. For example, CNN is frequently utilized in computer vision and image recognition, while RNN is frequently used in time series problems and forecasting. On the other hand, there is no obvious winner for generic issues such as categorization because the design chosen may be influenced by a variety of circumstances[30].

An artificial neural network (ANN) having numerous layers between the input and output layers is known as a deep neural network (DNN). Neural networks come in a variety of shapes and sizes, but they all include the same basic components: neurons, synapses, weights, biases, and functions. These components work in the same way as human brains and can be trained just like any other machine learning algorithm[31].

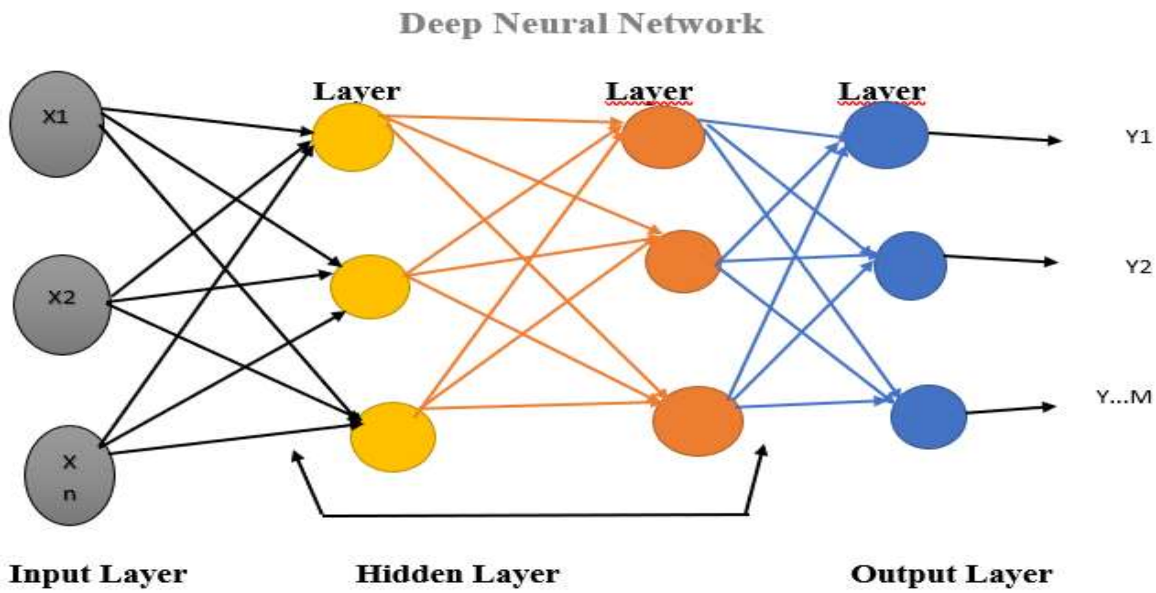


Figure 3: Architecture of DNN

Many variations of a few basic approaches make up deep architectures. In various domains, each design has proven to be successful. Unless multiple architectures have been assessed on the same data sets, it is not always viable to compare their performance.

DNNs are feedforward networks that transfer data from the input layer to the output layer without looping back. The DNN starts by creating a map of virtual neurons and assigning random integer values, or "weights," to their connections. The inputs and weights are multiplied, and the result is a value between 0 and 1. An algorithm would alter the weights if the network didn't recognize a pattern correctly. As a result, the algorithm might increase the influence of specific factors until it finds the optimal mathematical manipulation to fully process the data.

2.3.4. Recurrent Neural Network (RNN)

A recurrent neural network (RNN) is a type of artificial neural network in which nodes are connected in a directed graph that follows a temporal sequence. This enables it to behave in a temporally dynamic manner. RNNs, which are derived from feedforward neural networks, can process variable length sequences of inputs by using their internal state (memory). As a result, tasks such as unsegmented, linked handwriting recognition and speech recognition are feasible. In principle, recurrent neural networks are turing complete, which means they can run any program to handle any set of inputs[32]. The recurrent neural network (RNN) is neural network that uses connections between units to build a directed cycle, allowing the model to have dynamic temporal behavior. An RNN contains three layers: one input layer, a variable number of hidden levels, and one output layer. Basic RNNs consist of a network of neuron-like nodes, each having a directed (one-way) connection to every other node and a configurable real-valued weight on each link (synapse). Through repeated iterations of the neural network, these weights are constantly modified. Handwriting ,text classification ,sentiment analysis and speech recognition are some of the applications where RNNs are commonly employed currently[33]. Nowadays it is widely used and popular algorithm in deep learning, especially in NLP and speech processing. Unlike traditional neural networks, RNN utilizes the sequential information in the network. This property is essential in many applications where the embedded structure in the data sequence conveys useful knowledge[28]. For example, to understand a word in a sentence, it is necessary to know the context. Therefore, an RNN can be seen as short-term memory units that include the input layer *x*, hidden (state) layer *s*, and output layer *y* to understand the input and finally produce the desired output from the network.

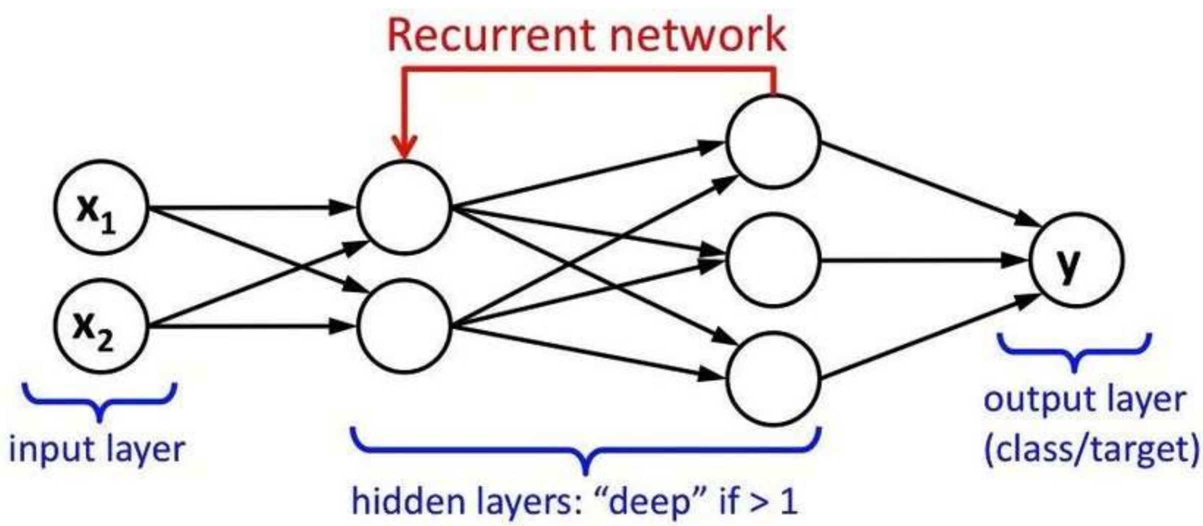


Figure 4:Recurrent Neural Network

2.3.5. Convolutional Neural Network(CNN) Model

An artificial neural network that is most commonly employed to assess visual imagery can also be used to classify sentiment and text. Artificial neural networks that are shift invariant or space invariant are built using a shared-weight architecture of convolution kernels or filters that slide along input features and produce translation equivariant outputs known as feature maps.

Surprisingly, rather than being invariant under translation, most convolutional neural networks are merely equivariant[34]. They can be utilized in image and video recognition, recommender systems, image classification, picture segmentation, medical image analysis, natural language processing, sentiment analysis, text classification, brain-computer interfaces, and financial time series, to name a few area where we apply convolutional neural network. CNN-based architectures have grown so common in the field of computer vision that few people today would develop a commercial product or compete in a competition involving picture recognition, object detection, or semantic segmentation without using them[35]. Convolutional neural networks are used for spolving a variety of problem in different areas currently.

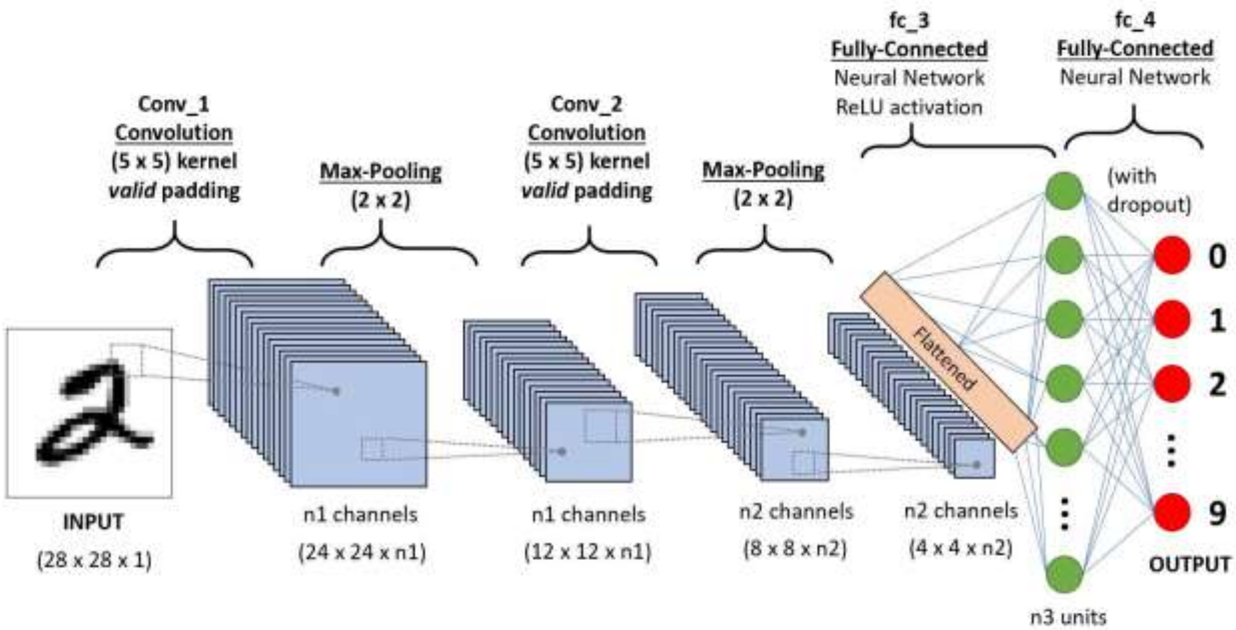


Figure 5: Convolutional Neural network Architecture

The architecture of convolutional neural network is shown in figure 5 above contains three layers which are the input layer for the purpose of receiving input for further analysis using the network, hidden layer for various mathematical computation to produce the desired output from the network anad final the output layer to display expected output from the network. In the proposed model convolutional neural network is used to classify the reviews into extreme, anti-extreme and neutral classes.

2.3.6. Long Short-Term Memory (LSTM) Model

Among various deep learning algorithm long short-term memory is the one used for classification of texts, sentiments and etc. Long Short-Term Memory (LSTM) is a sort of recurrent neural network (RNN) architecture that was developed as an improvement over simple RNNs for more correctly representing temporal sequences and their long-range dependencies[33]. The long short-term memory was developed to overcome the drawbacks of recurrent neural network. The memory of original RNN will reduce the performance of the model due to the number of complex layers after multiple recursions. As a result, the concept of a long short-term memory (LSTM) neural network is introduced. It's a unique and significant RNN model that can memorize long-term or short-term values while allowing the neural network to retain only the information it needs[37]. So, LSTM is one among deep learning models which is used for different classification purpose. We implemented LSTM model for the proposed work to analysis performance to choose the more accurate model for extremist review anlysis for Afaan Oromo social media post and comments.

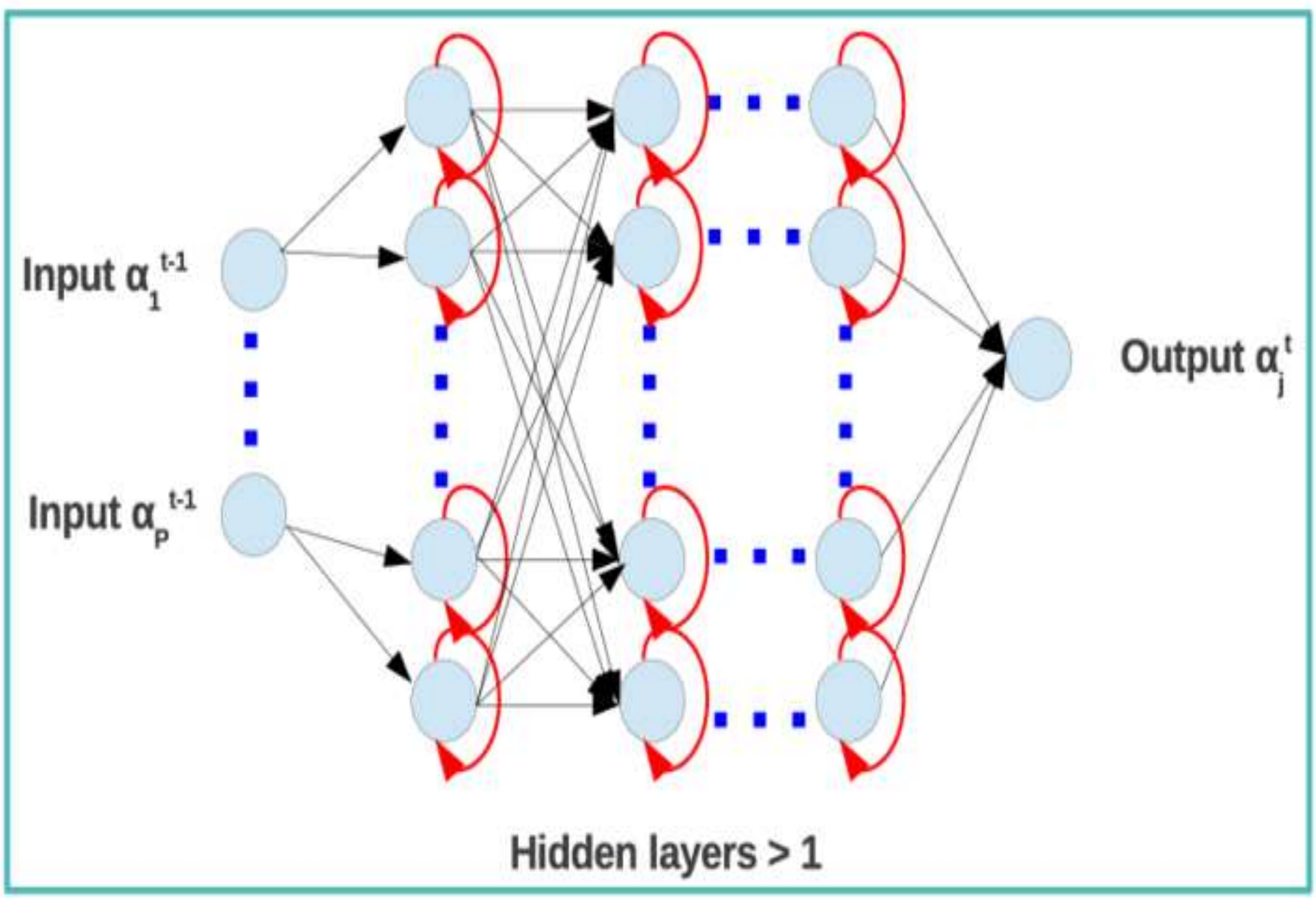


Figure 6:Architecture of long SHORT-TERM memory (LSTM)

2.4. Feature Extraction Approach

A feature is a function of fundamental measurement variables or characteristics that identifies some quantifiable property of an object and is useful for pattern identification and categorization. Obtaining a decent data representation is a domain-specific effort that is dependent on the measurements available.

The process of representing raw data in a reduced form to aid decision making such as pattern identification, classification, or recognition is known as feature extraction. Finding and extracting accurate and discriminating features is always a necessary stage in the sentiment categorization process[38]. Feature extraction plays a golden role in sentiment analysis to analyze the opinion, feelings, attitudes of the people and extract the hidden patterns from texts used for decision making. Therefore, we implemented feature extraction in the proposed model to identify the polarity of the reviews as well as to enhance the performance of the proposed model.

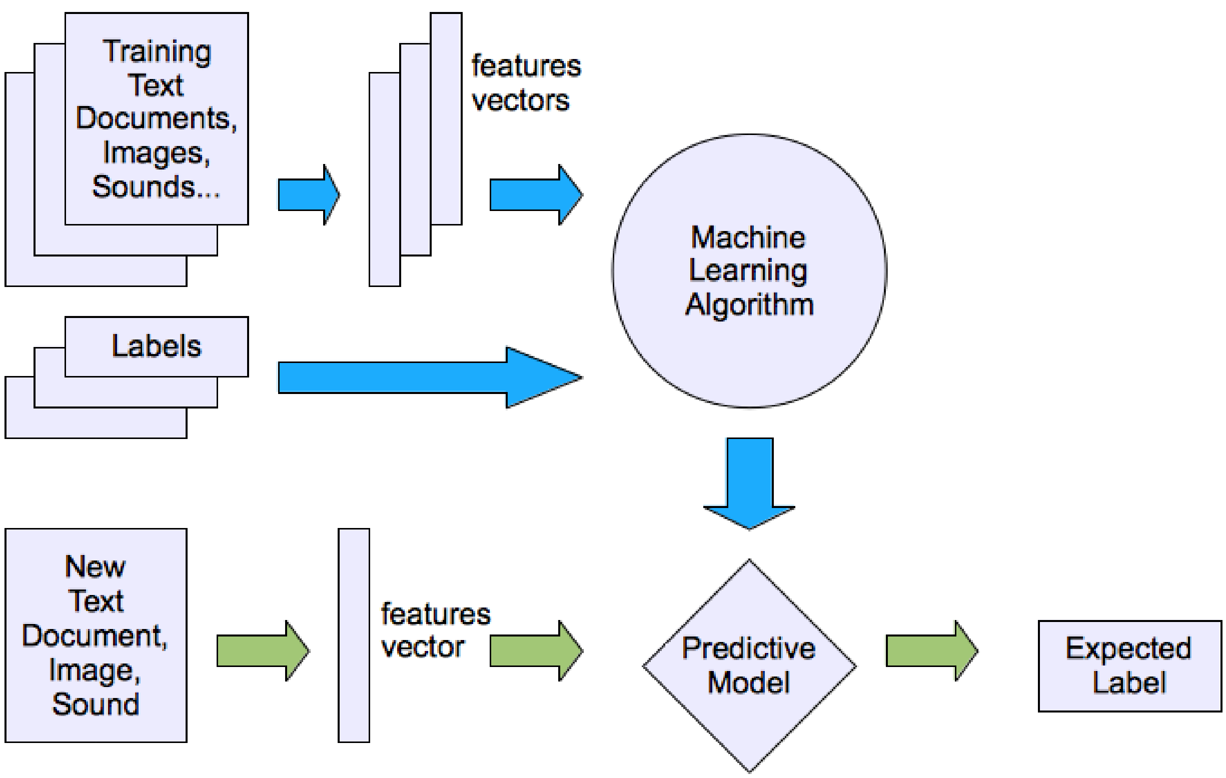


Figure 7:Feature extraction model

2.4.1. Word Embedding

Humans have an unrivaled ability to comprehend linguistic nuances. In a single sentence, the perceptive human brain can easily understand humor, sarcasm, negative sentiment, and much more. The only criteria are that we must understand the language in which the phrase is written. For example, if someone commented on a post or news item in a language other than English, the author would almost certainly not understand what the person was trying to say.

We must interact with the listener in the language he or she understands best in order to communicate effectively. It is critical to represent any type of text in a language that a machine can understand, which is in the form of numbers or digits "0" and "1." Regardless of the data we give it: video, audio, image, or text, a machine can only work with numbers. As a result, one of the most actively researched topics is representing text as numbers, or embedding text, as it is known. The initial embedding techniques dealt with only words. Given a set of words, you would generate an embedding for each word in the set. The easiest way was to one-hot encode the sequence of words provided, assigning a 1 to each word and a 0 to the remaining words. While this worked well for expressing words and other simple text-processing tasks, it didn't work so well for more sophisticated tasks like locating terms that were similar.

For example, if we search for a query*:* “Best Ethiopian restaurant in USA”, we would like to get search results corresponding to Ethiopian food, restaurants in USA and best. However, if we get a result saying: Top Ethiopian food in USA, our simple method would fail to detect the similarity between ‘Best’ and ‘Top’ or between ‘food’ and ‘restaurant’.

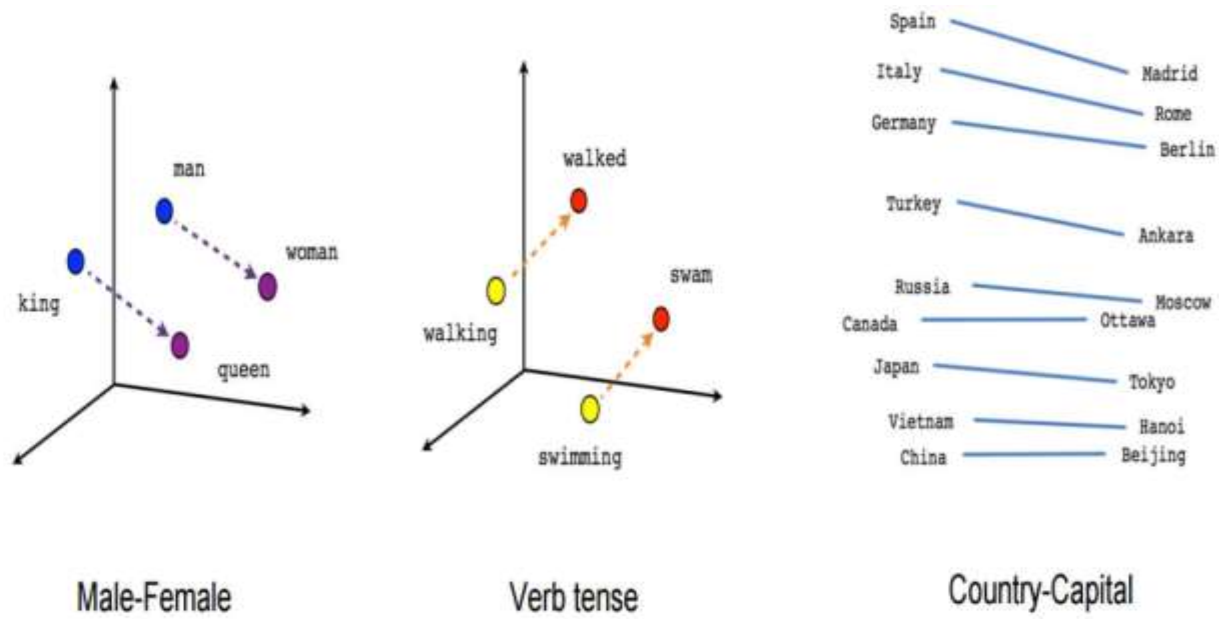


Figure 8:Word embedding

As a result of this problem, we now have word embeddings. In essence, a word embedding not only converts the word, but also recognizes the semantics and syntaxes of the word in order to create a vector representation of the data. Word2Vec, GloVe, ELMo, FastText, and other prominent word embedding approaches are examples.

2.4.2. Word2Vec Approach

Big data is a large category of data that has been applied to a variety of applications. Processing a large data set takes time, not only because of the bulk of the data, but also because the data type and structure can be varied and complex. Many data mining and machine learning techniques are currently being used to solve big data problems; some of them can build a decent learning algorithm using a large number of training examples. However, if the learning algorithm is capable of picking useful features or minimizing the feature dimension, it will be more efficient[39]. Word2Vec is a collection of two learning models: Continuous Bag of Words (CBOW) and Skip-gram, which was suggested and backed by Google. In the CBOW model, the distributed representations of context (or surrounding words) are combined to predict the word in the middle**.** While in the Skip-gram model, the distributed representation of the input word is used to predict the context**.**

Word2Vec produces word vectors that can be represented as a significant chunk of text or even the complete article by putting text data into one of the learning models.

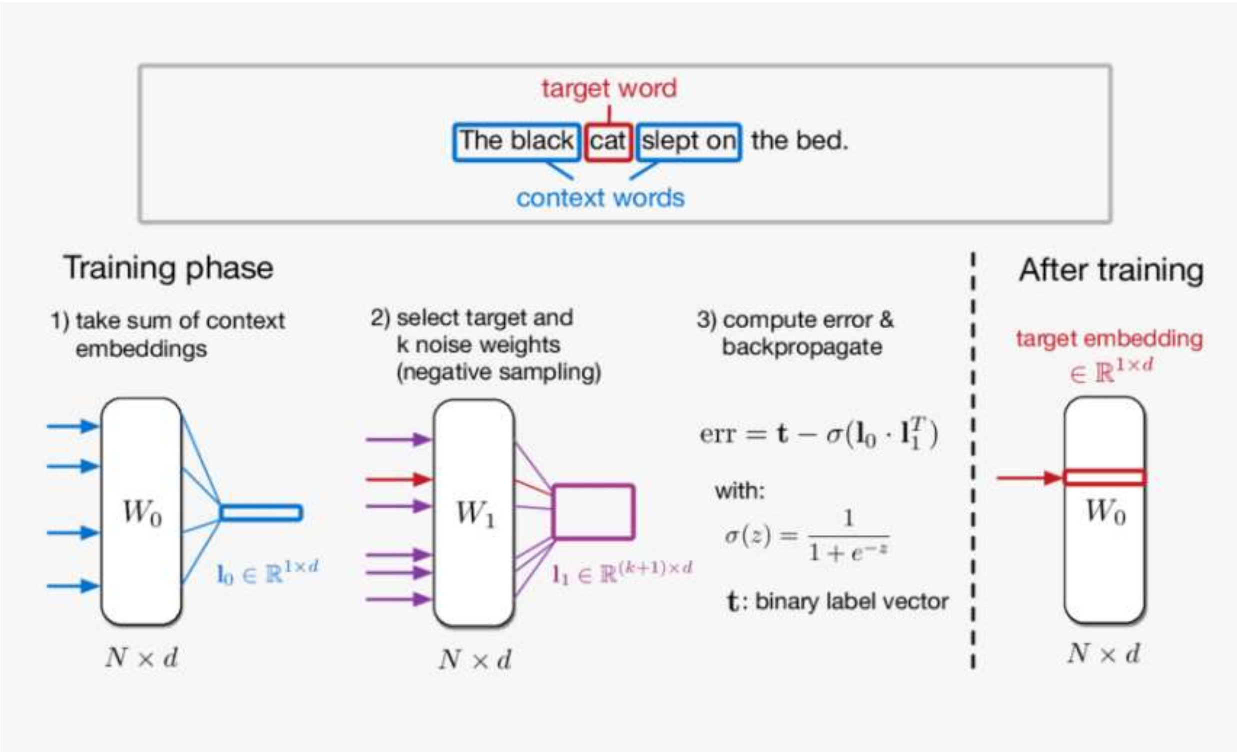


Figure 9:Word2Vec model

2.5. Literature review

Recently, some sentiment analysis researches were studied regarding Afaan Oromo text. As far as researchers know different researches have been proposed within various feature and techniques on a variety of domains using data set of Afaan Oromo. In work[19] they proposed multi-class sentiment analysis for Afaan Oromo text based on supervised machine learning approaches. They developed sentence level sentiment analysis which is classified into five multiple classes- very negative, negative, neutral, positive and very positive. The researchers proposed two methods of supervised machine learning approaches-Support Vector Machine and Random Forest algorithms to classify sentiment polarity from Oromia broad casting network (OBN) Twitter by Ethiopian language Afaan Oromo. They prepared 1810 dataset from the Oromia Broadcasting Network for training and testing the model. Finally, they achieved for both support vector machine and random forest algorithms accuracy of 90% and 89% respectively. While their work achieves better accuracy for the proposed state of art, but they implemented old machine learning models than implementing current models which can perform classification of sentiment analysis than the approach the authors used in their proposed work as well as they only analysis short texts for their model instead of using sentence, document or aspect-based sentiment analysis levels. The other recent paper[7] used combined Convolutional Neural Network and Bidirectional Long Short-Term Memory (CNN-Bi-LSTM) to analyze sentiment for Afaan Oromo. For the experiment, they used data from two domains: Facebook and Twitter. The cleaned data was manually annotated into five categories by four separate annotators: 2, 1, -2, -1, and 0, which signify very positive, positive, very negative, negative, and neutral, respectively. After they prepared and annotated the dataset, experiments using Convolutional Neural Networks, Bidirectional Long Short-Term Memory, and combined Convolutional Neural Network-Bidirectional Long Short-Term Memory with character level word embedding were performed on the prepared corpus from Facebook and Twitter. They reached a promising performance accuracy of 93.3 percent, 91.4 percent, and 94.1 percent for CNN, Bi-LSTM, and CNN-Bi-LSTM, respectively, based on the deployed Facebook dataset. Also, for CNN, Bi-LSTM, and CNN-Bi-LSTM, 92.6 percent, 90.3 percent, and 93.8 percent were attained, respectively, for CNN, Bi-LSTM, and CNN-Bi-LSTM. There is also paper which is done on the area of extremist based content detection in foreign languages English among them[11] proposed work as stated in the paper everyday a large number of articles, tweets, posts, posters, blogs, comments, views and news are posted online without a check which in turn imposes a threat to the security of any nation. However, different agencies are working on getting down this radical content from various online social media platforms. To overcome the state of the art they designed deep learning algorithm in detection of radicalization contrary to the existing works based on machine learning algorithms. Researchers used LSTM based feed forward neural network to detect radical content. They gathered a total of 61601 records from various internet news, article, and blog sources. The Domain experts classify these data into three categories: radical (R), non-radical (NR), and irrelevant (I), which are then used to classify radical content using an LSTM-based network. Finally, using the proposed method, they were able to attain a model precision of 85.9%.

On another[40] proposed classification of extremist text on the web using sentiment analysis approach for English language. The high volume of extremist content available on the internet makes manual classification impractical at this time.

However, automatic classification approaches are required. They used J48 decision tree to classify the data set to pro-extremist, anti- extremist and neutral. They used dataset comprises of 7500 web pages manually classified into "pro-extremist," "neutral" and "anti-extremist." the web pages were classified based on the contents they exhibited. According to the results of the J48 decision tree algorithm, 93.8 percent of the 7500 web pages processed were accurately sorted into their appropriate classes. With 98.7% of pro-extremist cases and 94.2 percent of anti-extremist cases correctly identified, the pro-extremist and anti-extremist classes had the most correctly identified pages. However, performance on the neutral class was low at the rate of 88.7%. Another work done on the area of automatic detection of extremist affiliation is done on English language[41]this work aims at proposing a terrorism-related content analysis framework with the focus on classifying tweets into extremist and non-extremist classes. They develop a Twitter tweet categorization system based on user-generated social media posts, using deep learning-based sentiment analysis algorithms to identify tweets as extremist or non-extremist.

However, some researchers are investigated on Afaan Oromo currently by different researchers on sentiment analysis using various approach like machine learning and deep learning models. In the proposed work, we review those previously done researches we are going to implement sentiment analysis using deep learning approach called Convolutional Neural Network (CNN) we choose this by reviewing different papers and we reached on decision to use it because it achieves higher performance in sentiment classification and texts. To decide the polarity of the statements, word embedding also used. Among various word embedding used in other previously done works we chose word2vec embedding for our work as approach and also not used in other previous researches done in Afaan Oromo on sentence level sentiment analysis. We also used subjective data for our proposed model to classify the non-factual data to its correct category. As per the knowledge of the researcher there no work is done on sentence level sentiment analysis for Afaan Oromo reviews with deep learning. Finally, we implement sentence level sentiment analysis using convolutional neural network and long short term memory for Afaan Oromo reviews.

2.6. Related Work

This part briefly discusses some of the sentimental analysis related works for some languages. In addition, mostly employed approaches for sentiment analysis i.e., machine learning and deep learning are reviewed.

In the area of sentiment analysis, most of the research has been done in English. However, there were some attempts to tackle the sentiment analysis task in other languages such as, Afaan Oromo, Amharic, Arabic, English, French, Turkish and etc.

2.6.1. Sentiment Analysis Using Machine Learning Based Approach for Afaan Oromo

A lot studies done on sentiment analysis for local language and international language using lexicon and rule-based approach. We also review some of the research works and article that has been done on both local and international language using different approach as follows.

In [42] proposed a Sentiment Analysis of Afaan Oromo using Machine learning Approach. Because automatically recognizing and classifying opinions from social media posts might give major economic and societal benefits, the author study focuses on sentiment analysis of Afaan Oromo social media content. They employed Naïve Bayes machine learning algorithm and different n-grams such as unigram, bigram, trigram and their combinations as features. The author of this study looked at document-level sentiment analysis, which implies that each document represents a single sentiment. This paper's target website is the social networking site Facebook. They stated that because Facebook has many members and vast user-generated data is available. Thus, the research focuses on sentiment analysis of Afaan Oromo texts on Facebook. In this study, they developed Naïve Bayes machine learning approach. They collected reviews from Facebook, since online data is mostly in the text format and unstructured in nature. Thus, they remove stop words and other unwanted characters from the review for further analysis. These reviews go through a vectorization process, which converts text data into a matrix of numbers. Machine learning techniques are then used to classify the comments using these matrices. Different parameters are then used to evaluate the performance of the machine learning algorithm. For this study, the authors primary data source was Oromo Democratic Party /ODP official Facebook page is extracted by using face graph API. They chose the page because there is a lot of users created content. The authors concentrated on sociopolitical issues, government policy, and other relevant topics. There are 1452 reviews total, with 726 good and 726 negative ones. They saved the extracted data in comma delimiter (CSV) format using excel. According to their findings, the accuracy of unigram, bigram, trigram, unigram-bigram, unigram-trigram, and bigram-trigram is 90.7 percent, 71.1 percent, 54.6 percent, 92.7 percent, 92.4 percent, and 75 percent, respectively.

Authors	Language	Level	Review	Approach	Model	Technique	Accuracy
Megersa oljira,2020.	Afaan Oromo	Document level sentiment analysis.	Positive and negative	Machine learning	Naïve bayes	Unigram	90.7
						Bigram	71.1
						Trigram	54.6
						Unigram-bigram	92.7
						Unigram-trigram	92.4
						Bigram-trigram	75

Table 1: Experimental result achieved by Megersa Oljira.

The other study done by[43] propose that sentiment analysis model for Afaan Oromo short message service text: a machine learning approach. This study attempts to design a two-step approach for Afaan Oromo text sentiment classification model, clustering followed by classification algorithms.

To conduct the experiment, the author used a total of 1597 data from Oromia Broadcasting Corporate's (OBN) "8331SMS database" from three domains (news, entertainment, and general service domain). To begin, text preprocessing is performed to clean the data and prepare it for subsequent processing. The unlabeled Afaan Oromo text opinion reviews are then clustered using a clustering technique to establish natural grouping. The clustering methods K-means and Gaussian Mixture (GMM) were put to the test. GMM outperforms the competition and is chosen to retrieve a specific set of Afaan Oromo documents. The clustering algorithm's output is directly exported to a CSV file, which is then prepared for classification jobs. In each domain, three supervised learning techniques are employed to classify the sentiment of short Afaan Oromo text: Nave Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). SVM outperforms NB and KNN in the news, entertainment, and general service domains, with accuracy of 91.66 percent, 93.76 percent, and 92.87 percent, respectively.

Authors	Language	Approach	Techniques	Accuracy
Ashebir Hunegnaw et.al.2021	Afaan Oromo	Machine learning	Naïve bayes	91.66
			Support Vector Machine	93.76
			K-Nearest Neighbor	92.87

Table 2: Experimental result achieved by Ashebir Hunegnaw.

The authors [19] proposed that multi-class sentiment analysis from Afaan Oromo text based on supervised machine learning approaches. In this work the authors used sentence level sentiment analysis is done into five multiple classes- very negative, negative, neutral, positive and very positive. Here, they proposed two methods of supervised machine learning approaches-Support Vector Machine and Random Forest algorithms to classify sentiment polarity from Oromia Broadcasting Network (OBN) Twitter by Ethiopian language Afaan Oromo. These approaches have been proposed for classifying opinions from Afaan Oromo text at sentence level into five multiple sentiment classes. They used tokenization, stop word removal, normalization and stemming as preprocessing and tf-idf used as feature extraction. The performance of proposed approaches Support Vector Machine and Random Forest achieved accuracy 90% and 89% on OBN Twitter dataset with 1810 corpus size respectively.

In[7] this also another study done on Afaan Oromo sentiment analysis using deep learning based sentiment analysis titled with sentiment analysis for Afaan Oromo using combined convolutional neural network and bidirectional long short-term memory. The authors used character level multiscale sentiment analysis for Afaan Oromo in the study. Sequential correlations can be learned by CNNs. In contrast to CNN, RNN are specialized for sequential modeling but unable to extract features in parallel. So, they proposed to CNN-BiLSTM to get the advantage of the two algorithms. The researchers collected short text from Facebook and Twitter for their model. They also examined the performance of CNN and Bi-LSTM separately with CNN-LSTM together. In addition, the authors compared the performance of CNN, Bi-LSTM, and CNN-Bi-LSTM on small and large datasets. Because there is currently no standardized and appropriate corpus for Afaan Oromo natural language processing (NLP) tasks including sentiment analysis, the authors gathered data from two domains: Facebook and Twitter. They removed user names, links, non-Afaan

Oromo texts, and any unnecessary characters after collecting data. The cleaned data was manually annotated into five categories by four different annotators: 2, 1, -2, -1, and 0, which represent very positive, positive, very negative, negative, and neutral, respectively. This multi-scale sentiment analysis provides a more refined analysis, which is important for prioritizing and comparing various points of view. After that, they applied Convolutional Neural Network, Bidirectional Long Short-Term Memory, and coupled Convolutional Neural Network Bidirectional Long Short-Term Memory with character level word embedding to the prepared corpus from Facebook and Twitter for the puppose of evaluating the performance of the model in classifying the review.

In both the Facebook and Twitter datasets, the experimental results show that the proposed model outperforms than both CNN and Bi-LSTM. Based on the implemented Facebook dataset the authors achieved a promising performance accuracy of 93.3 percent, 91.4 percent, and 94.1 percent for CNN, Bi-LSTM and CNN-Bi-LSTM respectively. Consequently, they executed twitter dataset and achieved 92.6 percent, 90.3 percent, 93.8 percent for CNN, Bi-LSTM and CNN-Bi-LSTM respectively.

Authors	Language	Level	Approach	Algorithm	Data set	Accuracy
Megersa Oljira 2020.	Afaan Oromo	Character level	Deep learning	CNN	Fb, twitter	93.3,92.6
				Bi-LSTM	Fb, twitter	91.4,90.3
				CNN-Bi-LSTM	Fb, twitter	94.1,93.8

Table 3: Summary of experimental result achieved by Megersa Oljira

2.6.2. Sentiment Analysis Using Machine Learning Based Approach for Amharic

Research work done under[44] presents the study of sentiment analysis for Amharic social media texts. For the sentiment classification task, the authors followed the document classification approach. The data collected from tweets mainly constitute of one or two sentences, which are limited to the maximum length of texts allowed by Twitter. They explored both classical supervised machine learning and deep learning approaches for the classification task. Instead of manually crafted features, the authors have used automatic text representation techniques such as Term-Frequency Inverse Document Frequency (TF-IDF) and word representations (embedding). The TF-IDF document representation is produced using the scikit-learn, CountVectorizer and TF-

IDF transformer built-in methods in the proposed work. For the TF-IDF computation, each tweet is considered as a document. To build word2vec-based word representations, they have collected around 15 million sentences from different sources, such as a News dataset using the Scrapy Python API5, YouTube comments using the YouTube Data API6, and a Twitter dataset using the Twitter API7. They collected the datasets every day and store relevant metadata such as the date, title, and language of the dataset. As they have discussed in their work each tweet is annotated with 'Positive', 'Negative', 'Neutral', and 'Mixed' sentiment classes. They collected 9.4k annotated tweets were further split into training, development, and test instances using an 80:10:10 split. They have used the development dataset to optimize the learning algorithms. All the results reported in the remaining sections are based on the test dataset instances. After error analysis, they found out that tweets annotated with the 'Mixed' class are noises, which can be regarded as 'Positive', 'Negative', or 'Neutral'. The proposed model achieved accuracy of 60.51 for machine learning approach.

The other study under taken in Amharic sentiment analysis investigated sentiment analysis for Amharic using deep learning-based approach[45]. The authors have analyzed some Facebook posts to understand socio-political sentiments. They applied the state of the art in sentiment analysis on Amharic language using a deep learning approach in the socio-political sphere as part of this broad scope. Further preprocessing of the dataset on other domains improves the teview quality for furthere analysis, so their research aims to show data extracted from Fana broadcasting corporation's official Facebook page using Facebook's Graph Application interface on immigration, war, and public relations issues, and prepare the data for further preprocessing such as tokenization, stop word removal, and sentence stemming are all performed after collecting data from the FBC using post id as stated in the paper.

The extracted data, which includes both the text file and Emoji, is manually annotated by linguistic experts into seven categories: positive, very positive, extremely positive, neutral, negative, very negative, and extremely negative, while taking into account the effect of the most common Emoji. The researcher creates a feature extraction approach for their custom language using the Scikit-learn feature extraction classes Count Vectorizer and TF-IDF Vectorizer. They compare the two feature extractors and discover that their model outperforms Count Victories because it provides a more straightforward description of the data.

The total amount of review collected is1600 from the immigration, war, and public relations areas to assess the systems' performance. Then they use three trials to evaluate their system's training and validation accuracy, modifying the training and testing split 90 percent, 10%:80 percent, 20%:70%, 30%, the size of the dataset, the number of epochs, and network layers. As a result, the first technique obtained 90.1 percent average training accuracy and 90.1 percent average validation accuracy results in the first trial. The second strategy results in an average training accuracy of 82.4 percent and an average validation accuracy of 40 percent. The third experiment conducted by increasing the number of data set 1600 and five network layers the model achieves 70.1 training accuracy and 40.1 validation accuracy.

Author	Language	Approach	Accuracy
Yeshiwa Getachew et.al.2019.	Amharic	Deep learning	70.0

Table 4: Experimental result achieved by Yeshiwa Getachew.

Proposed work[46] done under sentiment analysis for Tigrinya, a Low-Resource language. The researchers suggested a unique transfer learning approach for adapting a strong source language model to an unknown low-resource language in this paper. On the Cross-lingual Sentiment (CLS) dataset and a novel sentiment analysis dataset for the low-resource language Tigrinya, they exhibit competitive performance with mBERT and a pre-trained target language model using the XLNet language model. They created a sentiment analysis dataset for the Tigrinya language, which is divided into two categories: positive and negative. The authors gathered information from Eritrean and Ethiopian music videos and short film channels on YouTube.

They collected 30k automatically labeled training set and a test set of 4k size labeled by two native-professionals independently. The labeled comment is considered for the test set, only when they gave the same label for it. The final 4k test examples are balanced, having 2k positive and 2k negative comments. Additionally, they have used another dataset for testing the proposed method on languages like German, French, and Japanese. CLS dataset is used widely in the context of cross-lingual evaluations. It consists of English, German, French and Japanese languages collected from Amazon reviews on three different domains (Music, Books, and DVD).

English XLNet achieved 78.88 percent F1-Score with only 10k samples of the given Tigrinya sentiment analysis dataset, surpassing BERT and mBERT by 10% and 7%, respectively.

Furthermore, fine-tuning the (English) XLNet model on the CLS dataset produces encouraging results when compared to mBERT, even outperforming mBERT for one Japanese language dataset.

2.6.3. Sentiment Analysis Using Machine Learning Based Approach for International Languages

In the proposed work[40] a sentiment analysis classification of extremist speech in English language. The classification of such extreme websites was subsequently automated using a sentiment-based classification algorithm. The classification model assesses how effectively the pages could be matched to their relevant classifications automatically. The approach also finds specific data items that are classified different manually than they are automatically. The authors described the data set used, the evaluation metrics and the classification model using the J48 decision tree classifier. The dataset comprises 7500 web pages manually classified into "pro-extremist," "neutral" and "anti-extremist." the web pages were classified based on the contents they exhibited according to the work done by the researchers.

The data, comprising each Web page with their associated sentiment score and manual classification were deployed into WEKA, where the J48 algorithm was applied with 10-fold cross validation. That is, the dataset was split in such a way that 90% of the dataset was used for training and the remaining 10% was used for testing, this process was repeated 10 times and the mean accuracywas taken. Precision, Recall, F-measure, and Accuracy were employed as the metrics used for performance evaluation of the system.According to the experiment J48 decision tree algorithm achieves 93.8 percent of the 7500 web pages processed were accurately sorted into their appropriate classes; with 98.7% of pro-extremist cases and 94.2 percent of anti-extremist cases correctly identified, the pro-extremist and anti-extremist classes had the most accurately identified pages. However, performance on the neutral class was low at the rate of 88.7%. Overall, their system accurately matched online pages to manual classification with a 93 percent success rate. In addition, a feature selection method was able to cut the original 26-feature set by one feature, resulting in a 94 percent overall classification accuracy.

Another proposed work[11] propose sentiment analysis for English language under the title of detecting radical text over online media using deep learning. In contrast to prior research based on machine learning algorithms, the goal of this paper is to use a deep learning method to predict radicalization. The radical content is detected using an LSTM-based feed forward neural network. They collected total 61601 records from various online sources constituting news, articles and blogs, they analysis the collected review by domain expert to identify data sources that may contain radical content. They only chose records from articles and blogs since they were mostly made up of radical text. For the constructed model, 1274 data were used for training and validation, which were labeled into three categories: R (radical), NR (non-radical), and I (Irrelevant). Domain experts classify these data into three categories: radical (R), non-radical (NR), and irrelevant (I), which are then used to classify radical content using an LSTM-based network.

Using a Levenberg training process, it appears that the word2vec framework was used, and vectors were neutrally trained independently for radical and non-radical texts to yield word embeddings. Propogative iterations and gradient satisfaction were utilized as validation criterion during training their proposed model. Because it is assumed that texts are connected, a feed forward neural network supported by LSTM is employed for weight training. Many models employ word embeddings as direct features, while others employ the avg task, which involves calculating the weighted average of word embeddings in similar contexts as text representations. As a result, the fully connected layer compares thresholds created using the mean of radical and non-radical weights to forecast labels for test datasets.

The authors also compare the proposed model with few existing techniques for text analytics in order to establish the credibility of model in terms of different performance measures observed. At first, bag of-words technique is employed to build a term document matrix. Then three machine learning algorithms namely SVM, Random Forest and MaxEnt are implemented to achieve classification of texts being radical or not radical in their work. The algorithms applied have different ways of operating. Support Vector Machine focuses on finding a hyperplane in N-dimensional space that classifies the data points. Random forest takes into account many decision trees predicting classes individually, then the votes from all the trees are aggregated to decide final prediction of the model. Maximum Entropy Classifier selects the model with largest entropy among all the models fitting training data for classification. They used an LSTM-based technique to find radical content on the internet. This methodology can aid in the removal of inappropriate content from online social networking sites. In terms of precision, the proposed strategy has been able to outperform most state-of-the-art procedures by achieving accuracy of 85.9 percent.

Author	Language	Approach	Accuracy
Armaan kaura	English	Deep learning, LSTM	85.9%

Table 5: Experimental result achieved by Arman kaura.

The other study[41] conducted under the sentiment analysis for English language. This work attempts to develop a framework for analyzing terrorism-related content, with a focus on categorizing tweets into extremist and non-extremist categories. The researchers used deep learning-based sentiment analysis algorithms to identify tweets as extremist or non-extremist, based on user-generated social media posts on Twitter. The authors suggested using the LSTM-CNN model, which functions as follows: For capturing the global dependencies of a sentence in the document with respect to tweet classification into extremist and non-extremist, the CNN model is used for feature extraction, and (ii) LSTM model receives input from the CNN model's output and retains the sequential core-relation by taking into account the previous data.

They treat the job of detecting extremist affiliation as a binary classification problem. Also, consider the training set Tr = t1, t2, t3, tn, with Extrimist-affliation, yes, no for class tags (labels). A tag is assigned to every tweet. The goal is to create a model that can learn from the training data set and categorize fresh tweets as extremist or non-extremist. Individuals and organizations, especially radicals and extremist groups, use Twitter-based messaging to communicate. Future terrorist acts could be traced if this type of communication is used. They proposed a method for detecting tweets with radical content. They also categorize user feelings in terms of emotional attachments shown toward persons and groups who hold radical views. In their work, they investigated an experiment with multiple Machine Learning (ML) classifiers such as Random Forest, Support Vector Machine, KN-Neighbors, Naïve Bayes Classifiers, and deep learning (DL) classifiers. Task-driven embedding trained over different classifiers: CNN, LSTM, and CNN + LSTM encodes the feature set for these classifiers.

They gathered over 25,000 tweets by converting non-English tweets to English using Twitter's Python-based Google Translate API. Each review is compared to the seed words in the extremist's vocabulary lexicon, which was carefully compiled using BiSAL, a bilingual sentiment lexicon for researching dark web forums. All posts containing one or more terms from the manually produced lexicon are gathered in this way. They utilized a python-based beautiful-soup script for this. The data is saved in a ".CSV" file that may be read by machines. In this way, they acquired manually tagged training datasets for conducting experiments. Finaly their model achieves accuracy of 92 percent based on the experiment investigated.

Study	Techniques	Features	P%	R%	F%	A%
Wei et al.	Machine Learning (KNN)	Classical features (tf, tf-idf, tf-idf) BOW	73	71	71	74
Azizan and Aziz	Machine Learning (NB)	Classical features (tf, tfidf, tfxidf)	70	68	69	72
Proposed	Deep learning (LSTM+CNN)	Word embedding	90	86	84	92
Chalothorm and Ellman	Lexical-Based Sent word net	Sent score and polarity class	69	68	69	73
Others models experimented by the researchers	Machine learning (SVM)	Tfxidf	79	78	79	79
	Machine learning (RF)	Tfxidf	83	81	82	84

Table 6: Authors comparison of their proposed work with others.

2.7. Afaan Oromo language

Ethiopia is a country having over 80 languages and nearly 200 dialects. Amharic, Oromifaa, and Tigrinya are the three primary languages. The official government language is Amharic. During the 1970s, the government outlawed Oromifaa and a number of other languages.[47]. Along with Amharic, Oromo is Ethiopia's most important national language, and it is spoken in trade, local government, and the media. The Oromo ethnic group is the largest ethnic group in the country, accounting for around 40% of the population. There are approximately eight different Oromo groups in Kenya. The original homeland of the Oromo people was in Ethiopia and Northern Kenya[48].

Afaan Oromo is a language based entirely on nature. Each Afaan Oromo root was produced from corresponding Sounds or accessible roots, and so converges to sounds close by. Verbs, nouns, and new roots were generated by repeatedly combining similar sound imitating roots. Thus, Oromo followed up the idea and elaborated language[49].

Afaan Oromo is an Afroasiatic language that belongs to the Cushitic branch. It is native to the Ethiopian state of Oromia and spoken predominantly by the Oromo people and neighboring ethnic groups in the Horn of Africa. It is mostly spoken as a lingua franca in Ethiopia and northern Kenya. Oromo is Ethiopia's most widely spoken language, with native speakers accounting for 33.8 percent of the country's total population. It is the second most widely spoken language in Ethiopia, behind Amharic, in terms of overall number of speakers (including second-language speakers). In portions of northern and eastern Kenya, Oromo is spoken as a first language by an extra half-million people. Smaller groups of emigrants speak it in other African nations such South Africa, Libya, Egypt, and Sudan. Oromo is the most widely spoken Cushitic language and among the five languages of Africa with the largest mother-tongue populations[50].

Afaan Oromo 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 Dawa regional states and of the Oromia Zone in the Amhara Region. It is a language of primary education in Oromia, Harari, Dire Dawa, Benishangul Gumuz and Addis Ababa and of the Oromia Zone in the Amhara Region. It is used as an internet language for federal websites along with Tigrinya. Oromo was forbidden in school, discussion, and governmental concerns under Haile Selassie's reign[50].

2.7.1. Afaan Oromo speakers

The Oromo, Ethiopia's largest ethnic group, account for 50 percent to 60 percent of the country's population, or around 25 million people. They are "an ancient race, likely the indigenous stock on which most other peoples in this section of Eastern Africa (the Horn of Africa) were grafted". Their fertile country, Oromia, is 600,000 square kilometers in size and is located between 2 and 12 north latitude and 34 and 44 east latitudes[51].

About 85 percent of Oromo speakers live in Ethiopia, mainly in the Oromia Region. In addition, there are some Somalis who speak the language. Borana and Orma, two languages closely related to Ethiopian Oromo, have 722,000 speakers in Kenya, according to the Ethnologue. Oromo is Ethiopia's most widely spoken language, with over a million native speakers. After Arabic (assuming one counts the mutually unintelligible spoken variants of Arabic as a single language and assumes the same for the varieties of Oromo), Swahili, and Hausa, Oromo has the fourth most speakers in Africa. A number of members of other ethnicities that interact with the Oromo speak it as a second language, in addition to first language speakers. In northwestern Oromia, for example, the Omotic-speaking Bambassi and the Nilo-Saharan-speaking Kwama.

2.7.2. Writing Systems

Oromo is largely thought to have been written only in the early 1840s, and the first written evidence for the language is a treatise on the elements of the language produced by Krapf in 1840[52]. The Oromo language is written using the Qubee Latin script, which was formally accepted in 1991[51]. Various forms of the Latin-based spelling had previously been employed, mostly by Oromos outside of Ethiopia and, by the late 1970s, by the OLF (Heine 1986). Between 1991 and 1997, more works in the Oromo language are thought to have been created than in the previous 100 years, thanks to the adoption of Qubee. The Borana and Waata peoples of Kenya utilize Roman letters as well, although their systems are different.

Sheikh Bakri Sapalo (1895-1980; also known by his birth name, Abubaker Usman Odaa) created an indigenous Oromo script in the late 1950s that was later utilized underground.

It is an aesthetically separate innovation intended particularly for Oromo phonology, despite structural and organizational influences from Ge'ez and the Arabic script.

It is largely alpha syllabic in nature, but lacks the inherent vowel found in many such systems; in practice, all consonant characters must be marked either with vowel signs (creating CV syllables) or with separate marks to denote long consonants and consonants not followed by a vowel (e.g., in word-final environments or as part of consonant clusters). The Arabic script has also been used intermittently in areas with Muslim populations.

2.7.3. Consonant And Vowel Phonemes

Oromo has a set of ejective consonants, which are voiceless stops or affricates that are accompanied by glottalization and an explosive blast of air, just like most other Ethiopian languages, whether Semitic, Cushitic, or Omotic. Another uncommon glottalized phone in Oromo is the implosive retroflex stop, or "dh" in Oromo spelling, which is a sound similar to an English "d" made with the tongue curled back slightly and the air sucked in so that a glottal stop is heard before the next vowel begins. In most dialects, it is retroflex, but it is not substantially implosive and can be reduced to a flap between vowels. It's been described as "voiceless" by one source.

The five short and five long vowels in Oromo are indicated in the orthography by doubling the five vowel letters, as in Eastern Cushitic. For example, hara 'lake' and haaraa 'new' have a contrastive length difference. In Oromo, gemination is also important. That is, consonant length can distinguish between words, such as *badaa* 'bad' and *baddaa* 'highland.' The digraphs *ch, dh,* *ny*, *ph*, and *sh* appear in the Qubee alphabet. Digraphs are not required to be denoted with gemination, though some writers do so by doubling the initial element: *qopphaa'uu* 'be ready.'

In the charts below, the International Phonetic Alphabet symbol for a phoneme is shown in brackets where it differs from the Oromo letter. The phonemes /p v z/ appear in parentheses because they are only found in recently adopted words. Note that there have been minor changes in the orthography since it was first adopted: ⟨x⟩ ([t’]) was originally rendered ⟨th⟩, and there has been some confusion among authors in the use of ⟨c⟩ and ⟨ch⟩ in representing the phonemes /tʃʼ/ and /tʃ/, with some early works using ⟨c⟩ for /tʃ/ and ⟨ch⟩ for /tʃʼ/ and even ⟨c⟩ for different phonemes depending on where it appears in a word. This article uses ⟨c⟩ consistently for /tʃʼ/ and ⟨ch⟩ for /tʃ/.

Consonants

		Labial	**Alveolar/ Retroflex**	**Palato-alveolar**	**Velar**	**Glottal**
Plosives and Affricates	**voiceless**	**(p)**	**t**	**tʃ ⟨ch⟩**	**k**	**ʔ ⟨'⟩**
	voiced	**b**	**d**	**dʒ ⟨j⟩**	**ɡ ⟨g⟩**	
	ejective	**p’ ⟨ph⟩**	**t’ ⟨x⟩**	**tʃʼ ⟨c⟩**	**k’ ⟨q⟩**	
	implosive		**ɗ ⟨dh⟩**			
Fricatives	**voiceless**	**f**	**s**	**ʃ ⟨sh⟩**		**h**
	voiced	**(v)**	**(z)**			

Nasals	**m**	**n**	**ɲ ⟨ny⟩**		
Approximants	**w**	**l**	**j ⟨y⟩**		
Rhotic		**r**			

Table 7:Afaan Oromo consonants

Vowels			
	Front	**Central**	**Back**
Close	ɪ ⟨i⟩, iː ⟨ii⟩		ʊ ⟨u⟩, uː ⟨uu⟩
Mid	ɛ ⟨e⟩, eː ⟨ee⟩		ɔ ⟨o⟩, oː ⟨oo⟩
Open		ɐ ⟨a⟩	ɑː ⟨aa⟩

Table 8:Afaan Oromo Vowels

Short Vowel	Long Vowels in words	Gloss
a	ā **gāri** [ga:ri:]	good
i	ī **nīti** [ni:ti:]	wife
o	ō **dulōma** [dulo:ma:]	old
u	ū **gūtu** [gu:tu:]	full

Table 9:Afaan Oromo long and short vowels

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY

3.1. Introduction

In this chapter, we proposed sentiment mining model for opinionated Afaan Oromo text on extreme related review is described in detail. The proposed model has the following components: pre-processing, sentiment word detection, word similarity and polarity classification. Each component is composed of sub components which are the building blocks of the system. Pre-processing is responsible for normalization of reviews and dataset cleaning. In the sentiment words detection component, all possible sentiment words and contextual valence shifter terms are checked for existence in the sentiment lexicon. The weight manipulation component contains subsystems: weight assignment and polarity propagation we used word2vec embedding system for our model. After the weight manipulation is completed, the next step is the polarity classification of the reviews. The strength of the polarity (whether it is Extreme, anti-extreme and neutral) is rated in the post-classification analysis step. The sentiment word detection and weight manipulation activities are fully dependent on the lexicon of Afaan Oromo opinion terms that contains opinion terms tagged with a readily interpret able values. In addition, methodology ,tools used for implementing the prototype and the proposed algorithms are also presented. The model will have a great contribution for improving the service provided by the social media, and can enable our country to get the expected return from the government civil service and in politics, as well as crucial to know how extremist speech impact on people's opinions towards a particular party , religious issues , ethnic and government as a whole.

3.2. Methodology

In order to solve a research problem in a systematically way it is required to follow an appropriate research methodology. This research was conducted to construct sentimental analysis model for Afaan Oromo social media post on extremist-based related to religious, political and ethnic text from social medial Facebook, twitter, and YouTube comments. The research was adopted the following methodologies as the main focus of achieving its objectives.

Reviewing literatures on the study area from different sources like journal articles, books, internets, questionaries' to deeply understand about the sentiment analysis and identify approaches which is better perform in sentiment analysis. Data collection; we used different social media text Facebook, Twitter, and YouTube streamed containing some extremism related contents. The information gathered is saved in a format that can understandable by machines. We were able to obtain manually tagged training datasets in this manner, which we used to conduct the experiments. Applying different preprocessing techniques, such as tokenization, stop word removal, tokenization and special symbol removal. And also split the dataset for training and testing we used 80% for training and 20% used to test the designed model accuracy at predicting the reviews to its correct class. We plan to construct an automatic detection method for Afaan Oromo posts based on user-generated social media posts and comments, utilizing deep learning-based sentiment analysis techniques to classify the posts and comments as extremist, anti-extremist, or neutral. Word embedding which is word2vec is used to analysis the similarity between the collected data so as to feed the model for classification of the data set into its correct class. We develop convolutional neural network and long short term memory for classifying the review as extremist, anti-extremist and neutral review.

3.3. Implementation Tools

In order to meet the objective of the study, we used a number of environments and tools. The main tools that we used in accompanying this research are excel and notepad to store data collected by manually, online tools export comment website and Face pager API. We collected the data set from official pages which are known under the terms and services of social media service providers those are Voice of America Afaan Oromo (VOA), Oromia Broadcasting Network (OBN) And Britain Broadcasting Corporation Afaan Oromo (BBC). For implementing our proposed model, we used Google colab online tool for experimentation. We also used different python library to build our model such as Keras, pandas and NLTK for preprocessing and cleaning.

3.4. General System Architecture

The general architecture of the proposed sentiment analysis to extract valuable hidden patterns from extremist related review and opinions for Afaan Oromo text from social media posts and comments and finally to predict the reviews into the proposed classes. The proposed model's general architecture (sentiment mining model for opinionated Afaan Oromo texts) is depicted in figure 10. As shown in the figure, the system contains different components based on the processes required. These components are: pre-processing, feature extraction, word similarity and polarity classification.

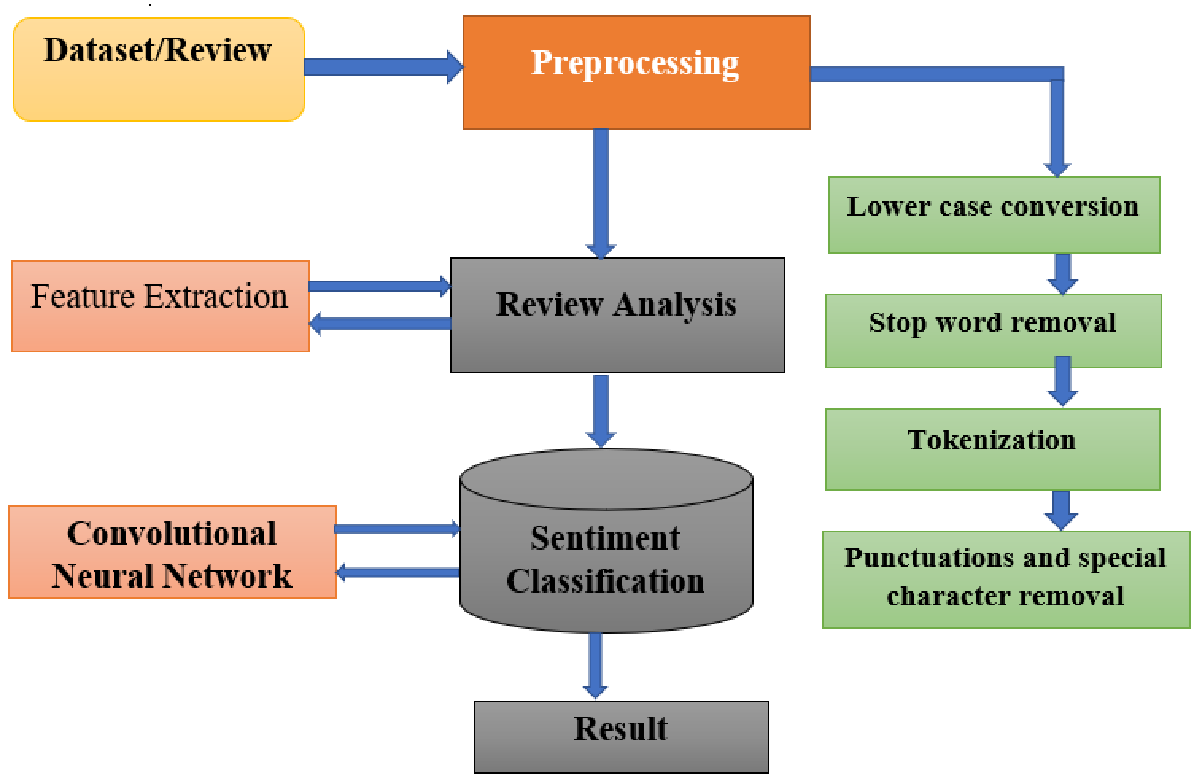


Figure 10: System Architecture

3.4.1. An Overview of How the System Works

As depicted in figure 10, the proposed system first takes a collection of reviews as input, and preprocesses them using three core steps. Data preparation, review analysis and sentiment classification. The result produced by opinion mining is sentiment classification of the reviews based on opinions expressed in the reviews. Each component is composed of sub components which are the building blocks of the system. Preprocessing is responsible for tokenization, normalization, and stop word removal of reviews.

The data preparation step performs necessary data preprocessing and cleaning on the dataset for the subsequent analysis. Preprocessing steps often utilized include deleting non-textual content and markup tags (for HTML sites) as well as information about the reviews that isn't required for sentiment analysis, such as review dates and reviewer names. The review analysis step analyzes the linguistic features of reviews so that interesting information, including opinions and/or sentiment, can be identified. Two commonly adopted tasks for review analysis are word embedding which is word2vec is used for converting the dataset to numerical form to be understandable by machine and for feature extraction. After this phase which means preprocessing is undertaken the model will classify the dataset to its correct category, sentiment classification is performed to get the results.

Data preparation also involves the sampling of reviews for building a classifier. We split the data set for training and testing the model so as we used 80% of the data set to train the model and 20% of the dataset is used for testing the model. Positive reviews often predominate in review datasets as reported in a number of studies. Some researchers therefore use review datasets with balance class distributions when training classifiers to help demonstrate the performance of their algorithms.

3.5. Identifying The Semantic Orientation of Words

The challenge of determining the semantic orientation (polarity) of a word is one of the most basic tasks in sentiment analysis. Variety of techniques have been used.

3.5.1. Opinion Words Extraction

People use opinion words to express a positive or negative opinion. Opinion words and product features are not independent of each other the opinion words always locate around the feature in the sentence. For example, let us look at the following sentence: "Oromoon yoo Biyya Itoophiyaa jedhamtu irraa jijjirama eege dogongore jira" In this sentence, "Oromoon yoo", the feature, is near the opinion word "dogongore jira" Based on this observation; we can extract opinion words in the following way. If a sentence in the review has a frequent term or feature, use the word2vec model to extract the nearby word. If such an adjective is found, it is classified as a term of opinion. The noun phrase that is frequently used is modified by an adjacent adjective. In this method, we can compile a list of words that express different points of view. We proposed word2vec with dimension of one hundred (100) in the proposed work to classify the reviews as extreme, anti-extreme and neutral class.

3.6. Data Source and Data Collection Methods

To realize the objective of our study, we collected different reviews manually from Facebook, twitter and YouTube social media on selective extremely affiliated speech's in Afaan Oromo reviews. We use both the three social medias because there is huge volume of Afaan Oromo texts are produced every day which contains radical contents. That is why we choose as preferable data source for our data collection and preparation. Those selected posts and comments are mainly which is written as a news on different domains and the users' comments, flings, attitudes and opinions towards the posts. Our data source was from official Facebook, twitter, YouTube page those are VOA Afaan Oromo, BBC Afaan Oromo, and OBN Afaan Oromo written by Afaan Oromo and it is legal under the company terms and condition. Even though posts and comments on this page covers many topics, we pay special attention to extreme or radical related domain focusing on ethnicity, politics, and religious related speech. The reason behind choosing these domains is the availability of user generated content in Afaan Oromo language pretty good than in others domain.

3.6.1. Data Collection and Preparation for The Model

One of the most critical and crucial parts of the Sentiment Analysis application is data collection. Because machine learning models are so widely used, simply having huge datasets on a domain-specific job does not guarantee higher performance. The model's performance is determined on the dataset's quality and labeling/annotation. Automatic predictions are likely to replicate the human disagreement observed during annotation since machine learning models learn from the data they are trained with. So, we collected dataset for the proposed work through manually and using online tools used for data collection from online such as Face pager API from social media posts and comments from three domains politics, religious and ethnic which have extreme, anti-extreme and neutral opinions. After the data is collected manually labeling of the data is carried out to its correct category based on their context by experts.

Data preparation also involves the sampling of reviews for building a classifier. After the dataset is labeled, it goes through various preprocessing stages for cleaning the dataset to feed to the proposed model. Positive reviews often predominate in review datasets as reported in a number of studies.

3.6.1.1. Preprocessing

The preprocessing methods are used to prepare the data for feeding the model and achieving the best outcome. The preprocessing steps included removing Null values, label the review to lower cases, spelling corrections. This is manually done by linguistic experts, tokenization, stop word removal, removing punctuation, and lowercase conversion carried out using NLTK python library. Each review in the dataset is labelled based on the review content such as extreme, anti-extreme and neutral class. If the review has extremely affiliated content, it is labelled as extreme, if the review has terms which express positive terms, it is labelled as anti-extreme, and the data set that is neutral which doesn't express any opinion is annotated as a neutral one manually. Furthermore, the dataset was split into 80% for training and 20% for testing.

Preprocessing is a data mining technique which involves transforming the raw data into a clear and understandable format. Preprocessing is considered a crucial step that must be done before importing data to the machine learning algorithms or deep learning to maximize the performance and the accuracy of the analysis, especially when the dataset being analyzed has a textual nature. Several steps were taken during the processing. First, the empty rows form, and the "Reviews" column was dropped (null values). Then, all textual data was converted to the lower case. Furthermore, the natural language toolkit library was used (NLTK), which is a machine learning library within NLP domain. Figure 4.10 illustrates the sequence of preprocessing for extremist reviews. Spelling corrections are necessary to ensure that the analysis delivers good findings; therefore, spelling errors must be taken into account, as they can occasionally alter the meaning of a text in our case made the spelling correction manually by linguistic experts of Afaan Oromo. The spellchecker procedure was utilized to manually detect misspelled words and recommend the best solution. The text is transformed to the same case, ideally lower case, in one of the most common preprocessing stages. However, you do not need to perform this step every time you work on an NLP problem because lower casing can result in information loss in some cases. For example, if we are working on a project that involves a person's emotions, upper case phrases can indicate dissatisfaction or excitement but in case of sentiment analysis its mandatory to convert all the data set to lower cases.

When working with text data, tokenization is one of the most often utilized strategies. The process of transforming sensitive data into tokens is known as tokenization. Text data is tokenized and filtered for sentiment analysis to remove any extraneous tokens. Tokenization process is under taken to feed the data set to the model for the betterment of the model accuracy.

Stop words are words that are regarded meaningless in the context of sentiment analysis. These words do not aid in determining the true meaning of the sentence or review; thus, eliminating them will have no effect on the model's results or the analyses' precision or recall, with the exception of a few crucial stop words used to indicate negativity in Afaan Oromo languages. Keeping them would, however, increase the index's size, necessitating more computer resources for very big datasets.

Two methods were utilized to remove stop words. The first method, which is the most common, used the NITK library identify tokens containing stop words and removing them from the reviews such as (e.g., akka, kan, fi, immo etc). The second method is used to add a word to NLTK stop words collection that is not included in the library and needs to be removed.

Removing punctuation is to remove marks such as comma, full stop, and exclamation mark. In this step, all the punctuations from the text are removed. String library of Python contains some pre-defined list of punctuations such as '!" #$%&'() *+, -. /: @ [\] ^_` {|} ~' so all those should be removed from the data set to reach high performance of the model.

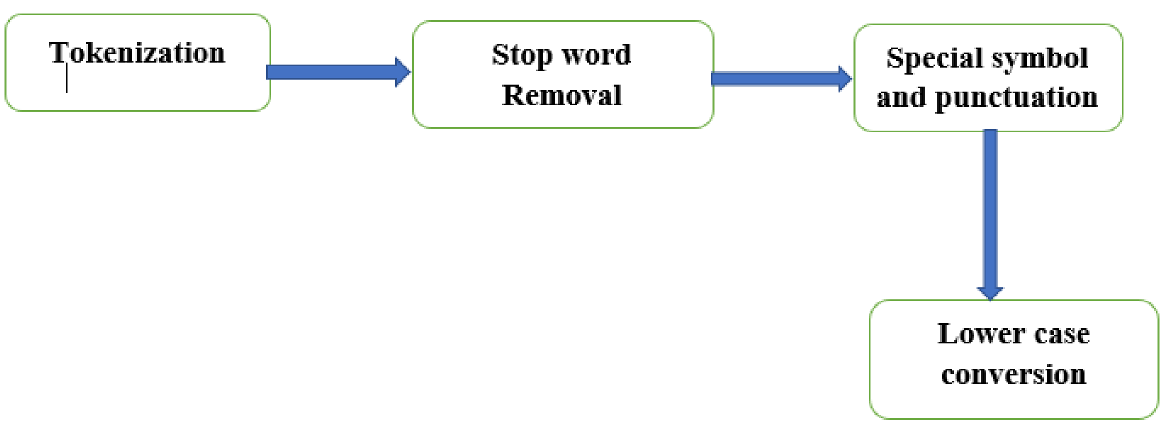


Figure 11.Preprocessing diagram

3.6.1.2. Tokenization

Tokenization is the process of breaking text document into words, phrases, symbols and other meaningful elements called token. In natural language strings were broken into words, digits and punctuations. Then tokens were generated from tokenized strings. It was performed by using white space.

Open file containing corpus
Split documents at word level
For a word in wordlist

If a word is in r"[^a-zA-Z’] +",' ',

Remove unwanted character

end if

end for

returnlist of cleaned text

close file

End algorithm

Algorithm 1:Algorithm to Remove Special Characters and Numbers

3.6.1.3. Stop Word Removal

There is no structured stop word list prepared for Afaan Oromo text document. Hence, researchers manually prepared Afaan Oromo text document stop words lists depending on corpus using Afaan Oromo dictionaries. These words are language specific words which carry no information like prepositions, conjunctions, articles, and particles but some stop words are used to indicate negative sentiments due to that they are not removed from the document. Some examples of Afaan Oromo stop words are sun, kun, ani, etc.

Open a file that contains stop word list

Read a sentence from the corpus

For each word in the corpus

If a word in list of stop words

Remove a word

else

Move to non-stop word list

end if

end for

return non-stop word listsEnd algorithm

Algorithm 2: Algorithm to remove stop words

3.6.1.4. Feature Extraction Word Embedding

In sentiment classification, there are two basic steps: first, finding a word embedding method to convert text into numerical representations, and then fitting the numerical representations of the text to machine learning algorithms or deep learning architectures for further analysis. However, in our paper, we used word2vec embedding as a technique. The technique of converting text into a vectorized numerical representation is called word embedding. Many machine learning algorithms and nearly all deep learning architectures are incapable of reading strings or plain text in its raw form, or of executing any type of task, including classification and regression. In general, they require numerical inputs. Furthermore, with such a large volume of data in text format, it is critical to extract information and make applications from it.

3.7. Evaluation Parameters

Evaluation metrics are used to display the classifier that performs well on this test dataset, so we need to be confident that it has the power to generalize well beyond the data from which it was trained. In this paper, different performance evaluation parameters including precision, F-measure, recall, and accuracy are calculated.

3.8. Proposed Models used for implementation

3.8.1. Convolutional neural network

Deep learning refers to the shining branch of machine learning that is based on learning levels of representations. Convolutional Neural Networks (CNN) is one kind of deep neural network. It can study concurrently[53]. Deep Learning is a type of neural network NN, but unlike traditional NN, Deep Learning contains additional hidden layers, allowing it to solve complicated issues. Convolutional Neural Networks, Recurrent Neural Networks, and Recursive Neural Networks are some of the different types of Deep Learning[54]. Deep learning techniques are based on neural networks, which are a subset of machine learning. They're made up of node levels, each of which has an input layer, one or more hidden layers, and an output layer. Each node is connected to the others and has a weight and threshold assigned to it. If a node's output exceeds a certain threshold value, the node is activated, and data is sent to the next layer of the network. Otherwise, no data is sent on to the network's next tier. Due to this in the proposed work we implemented a convolutional neural network for the opinionated Afaan Oromo reviews to analyze the sentiment of the review to its correct class into extreme, anti-extreme and neutral.

3.8.1.1. Convolutional Layers Work in Deep Learning Neural Networks

In convolutional neural networks, the major building elements are convolutional layers. A convolution is the basic process of applying a filter to an input to produce an activation. The location and strength of a detected feature in an input, like as an image, are shown by a map of activations termed a feature map, which is created by repeatedly applying the same filter to an input. The capacity of convolutional neural networks to learn a large number of filters in parallel particular to a training dataset under the restrictions of a certain predictive modeling problem, such as image classification, sentiment analysis with is its unique feature. As a result, extremely precise traits appear on input photographs and texts that can be identified everywhere. The higher performance of convolutional neural networks with picture, speech, audio signal, text classification, and sentiment analysis inputs set them apart from other neural networks. The proposed convolutional neural network for the state of the art has the following layers used for the classification of the sentiment. They have three main types of layers, which are input layer, hidden layer and also output layer basically. Some of the layers we used in our proposed models are: input layer ,convolutional layer, pooling layer, fully-connected (FC) layer, dense layer, dropout layer and output layer discussed in detail as bellow.

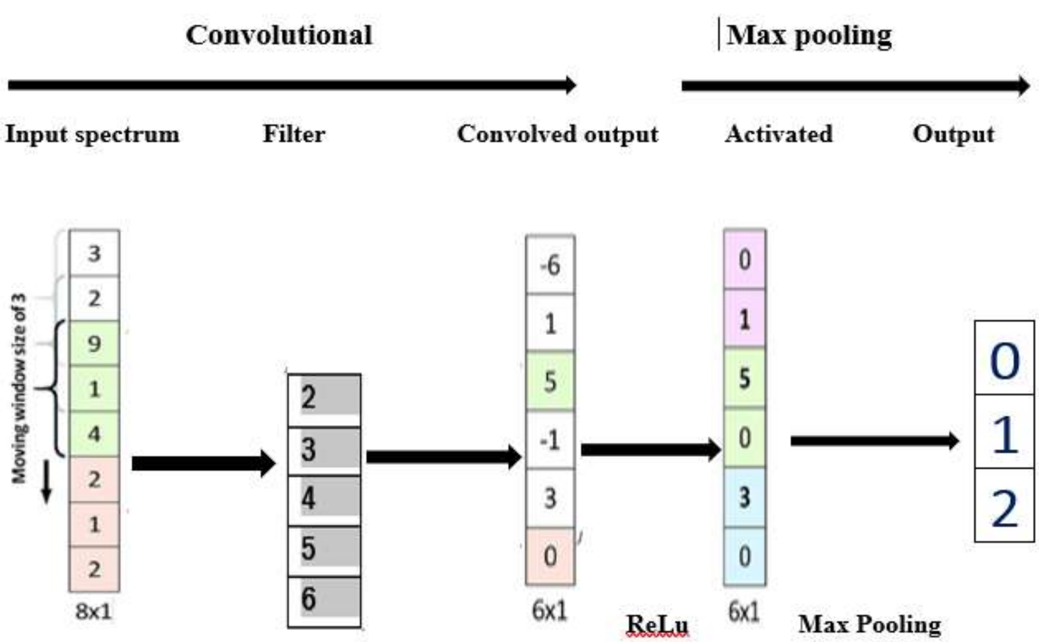


Figure 12: Architecture of convolutional neural network for the proposed work

A convolutional network's first layer is the convolutional layer. While further convolutional layers or pooling layers can be added after convolutional layers, the fully-connected layer is the last layer. The CNN becomes more complicated with each layer, recognizing text polarity from the provided input data set. The first layers concentrate on basic features like sentiment and vocabulary words. As the text input travels through the CNN layers, it begins to understand the polarity of the provided data set's emotion.

3.8.1.2. Convolutional Layer

The convolutional layer is the most important component of a CNN because it is where the majority of the computation takes place. It requires input data, a filter, and a feature map, among other things. Convolution, a specific sort of linear operation, is a major component of CNN, which is made up of a stack of mathematical operations, including convolution[55]. The convolution layer is an important part of the CNN design since it performs feature extraction, which is often a combination of linear and nonlinear processes, such as convolution and activation functions. Convolution is a form of linear operation used for feature extraction in which a tiny array of numbers called a kernel is applied over the input, which is a tensor of numbers. A feature map is created by computing an element-wise product between each element of the kernel and the input tensor at each position of the tensor and adding it to the output value in the corresponding place of the output tensor.

3.8.1.3. Pooling layer

A Pooling Layer is usually applied after a Convolutional Layer. This layer's major goal is to lower the size of the convolved feature map in order to reduce computational expenses. This is accomplished by reducing the connections between layers and operating independently on each feature map. There are several types of Pooling operations, depending on the method used. We employed Max Pooling in our proposed work to extract the largest element from a feature map. The average of the elements in a predefined sized text is calculated using Average Pooling. Sum Pooling calculates the total sum of the components in the predefined section. Between the Convolutional Layer and the Fully Connected (FC) Layer, the Pooling Layer commonly acts as a bridge. A pooling layer performs a conventional down sampling operation on the feature maps, reducing their in-plane dimensionality to introduce translation invariance to tiny shifts and distortions and reducing the number of learnable parameters. It is of note that there is no learnable parameter in any of the pooling layers, whereas filter size, stride, and padding are hyper parameters in pooling operations, similar to convolution operations[55].

3.8.1.4. Max pooling

Pooling to the limit Max pooling is the most common type of pooling procedure, which selects patches from input feature maps, outputs the largest value in each patch, and discards all other values. In practice, a max pooling with a 2 × 2 filter and a stride of 2 is commonly used. This down samples the in-plane dimension of feature maps by a factor of 2. Unlike height and width, the depth dimension of feature maps remains unchanged[55]. Maximum pooling is also useful for reducing the number of parameters and computations in the network.

3.8.1.5. Dropout layer

Machine learning systems with a large number of parameters, such as deep neural nets, are extremely powerful. Overfitting, on the other hand, is a serious issue in such networks. Large networks are very sluggish to utilize, making it difficult to avoid overfitting by merging predictions from multiple large neural networks at test time. Dropout is a method of dealing with this issue. During training, units (along with their connections) are dropped at random from the neural network. This inhibits units from over-co-adapting. Dropout samples from an exponential number of different "thinned" networks are taken during training. At test time, a single unthinned network with fewer weights can easily simulate the effect of averaging the predictions of all these thinned networks. This strategy drastically decreases overfitting and outperforms other regularization methods[56].

Usually, when all the features are connected to the fully connected layer, it can cause overfitting in the training dataset. Overfitting occurs when a particular model works so well on the training data causing a negative impact in the model’s performance when used on a new data. To overcome this problem, a dropout layer is utilized wherein a few neurons are dropped from the neural network during training process resulting in reduced size of the model. In the proposed model we used dropout to overcome the overfitting problem on passing a dropout of 0.2, 20% of the nodes are dropped out randomly from the neural network.

3.8.1.6. Dense layer

The dense layer is a deep-connected neural network layer, meaning that each neuron in the dense layer receives input from all neurons in the previous layer. In the designed model, the dense layer is used to revealed the most usually used layer. A matrix-vector multiplication is done in the background by the dense layer. Backpropagation can be used to train and update the values in the matrix, which are actually parameters. The dense layer produces an m-dimensional vector as its output. As a result, the dense layer is primarily used to modify the vector's dimensions. On the vector, dense layers apply operations such as rotation, scaling, and translation.

3.8.1.7. Fully connected layer

The weights and biases, as well as the neurons, make up the Fully Connected (FC) layer, which is used to connect the neurons between two layers. The last several layers of a CNN Architecture are usually positioned before the output layer. In this, the input review from the previous layers is flattened and fed to the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place.

3.9. Evaluation metrics

The performance of the model is determined by a variety of parameters. A model is regarded good if it achieves high accuracy in production or test data and can generalize effectively to unknown data. If it's simple to put into production and scalable. There are various techniques are used to evaluate our model like Confusion matrix, Accuracy, Precision and Recall.

3.9.1. Confusion matrix

A confusion matrix is a method of analyzing a classification algorithm's performance. If the number of observations in each class is unbalanced or the dataset has more than two classes, classification accuracy alone can be misleading. Calculating a confusion matrix can give a better idea of what the classification model is getting right and what types of errors it is making. Confusion is defined as 2*2 matrix that talks about the performance of the model. It’s just a representation of the above parameters in a matrix format.

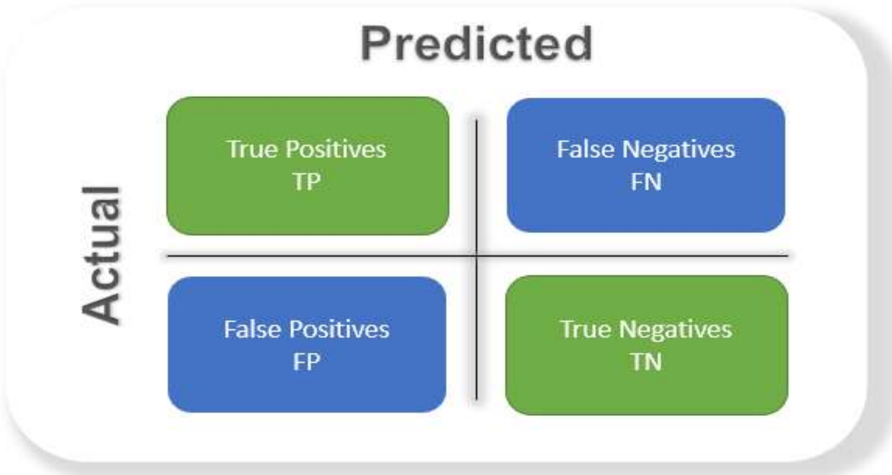


Figure 13:confusion matrix diagram

3.9.2. Accuracy

It is used to test the performance of the overall implemented model; how much the model is accurate in classifying the reviews into its correct category as per the labeled dataset feed to the network. The most widely used statistic for evaluating a model, however it is not a reliable indicator of its performance. When classes are unbalanced, the situation becomes even worse. It is the score that is generated when the class is generalized. The model's ability to generalize appropriately. In the proposed model we used accuracy as evaluation metrics to measure the performance of the developed model to measure the over all accuracy of the model.

$$\frac{TP + TN}{TP + FP + TN + FN}$$

Accuracy formula (Equation 1)

Note:

True positives (TP): Predicted positive and are actually positive.

False positives (FP): Predicted positive and are actually negative.

True negatives (TN): Predicted negative and are actually negative.

False negatives (FN): Predicted negative and are actually positive.

3.9.3. Precision

The precision is measured as the ratio of the number of correctly identified positive samples to the total number of positive samples (either correctly or incorrectly). The precision of the model in categorizing a sample as positive is measured. Positive instances as a percentage of total expected positive instances. The model prediction made as positive from the entire dataset is the denominator. Consider it a test to see 'how much the model is correct when it says it is correct. We also used precision evaluation metrics as evaluation parameter for the proposed model.

$$\frac{TP}{TP+FP}$$ Precision formula(Equation 2)

3.9.4. Recall

The recall is determined by dividing the total number of Positive samples by the number of Positive samples accurately categorized as Positive. The model's ability to recognize Positive samples is measured by the recall. The higher the recall, the greater the number of positive samples found. All the stated evaluation metrics are used to test the performance of our proposed model.

$$\frac{TP}{TP+FN}$$ Recall formula (Equation 3)

3.10. Summary

In this study, a deep learning approach is used to develop a sentiment mining model for Afaan Oromo extremely affiliated comments and posts which is collected from social network sites. The deep learning technique is based on the idea of teaching a machine to classify opinionated materials review data set into predetermined Extreme, anti-extreme and neutral categories. The use of a deep learning approach for sentiment classification entails data preparation and classification using convolutional neural network. The details of their application are implimented in the following chapters.

CHAPTER FOUR: EXPERIMENT AND IMPLEMENTATION

4.1. Introduction

We employed many methods and evaluation indicators to ensure that the model was operating well in order to meet the study's goal. The degree to which a system achieves its goal is referred to as effectiveness. A model that has been trained on a decent dataset with enough data are an approximation of the genuine model. As a result, we construct a dataset to test our suggested system. In this work, sentiment analysis or opinion mining are used to classify Afaan Oromo social media reviews and opinions on a variety of issues including as politics, ethnicity, and religion. Classification can be applied to many of the tasks in sentiment analysis. Many algorithms built for this purpose are available in deep learning, but the process of classifying text-based sentiment faces a number of distinct obstacles. The NLTK (natural language toolkit) text processing library in python is used to generate the cleaned dataset for the deep learning model. The model's performance is then assessed using a variety of metrics. Understanding the result, testing for interest, and interpreting the result are all part of the evaluation process is briefly discussed in this chapter.

4.2. Tools used

We use a laptop PC (with Windows 10 operating system, Intel (R)core (TM) i5-4300M CPU @2.60, 8 GB RAM, and 500GB hard disk), we used colab online platform for implementing our model and used deep learning frame work for implementation like Keras to construct our proposed model. We also used NumPy, NLTK, Pandas, and Genism are mostly used tools with other machine learning functions to implement our proposed model.

4.3. Dataset

Sentiment categorization is carried out on the Afaan Oromo corpus in this study. The data we used were collected from various official Facebook pages such as Voice of America Afaan Oromo (VOA), Britain Broadcast Corporation (BBC) Afaan Oromo, and Oromia Broadcasting Network (OBN) official Facebook pages which is written in Afaan Oromo only and known under the company terms of service. The purpose of the study is to classify the sentiment polarity of the reviews which is collected from three main domains religious, politics, ethnic posts and comments contains extremely affiliated contents in to three proposed classes called extreme, anti-extreme and neutral classes. There is no pre-made dataset with polarity tags for each review, which is one of the most difficult aspects of developing a model in Afaan Oromo sentiment analysis. Due to there is no standard corpus is for Afaan Oromo we collected the corpus for our model by manual and using online tools. As a result, we completed the task of categorizing the reviews into extreme, anti-extreme and neutral categories in order to satisfy the study's goal. We collected 2410 reviews from the above-mentioned social network sites which is manually tagged into extreme, anti-extreme and neutral reviews to feed to the proposed model convolutional neural network for training and testing model accuracy. We split the dataset into two for training and testing the proposed model. We applied 80% of the dataset for training purpose and 20% of the dataset is for testing purpose. From the toal of 2410 reviews 2928 of the data is used for training the model and 482 of the total review is used for testing purpose of the developed model in our proposed work.

4.3.1. Tokenization

In any NLP pipeline, tokenization is the first step. It has a significant impact on the remainder of our pipeline. Unstructured data and natural language text are broken down into chunks of information that can be regarded separate elements using a tokenizer. The token occurrences in a document can be utilized to create a vector that represents the document. An unstructured string (text document) is instantly transformed into a numerical data structure appropriate for machine learning. They can also be used to direct helpful activities and reactions by a computer. They could also be utilized as features in a machine learning pipeline to trigger more complicated choices or behavior. Due to tokenization is the first step in natural language processing we carried out tokenization by using NLTK preprocessing. Using the stated python library, we tokenize our data set for further analysis. Here below are sample dataset from the corpus before its is tokenized for further analysis and tokenization is the first step in preprocessing for sentiment analysis model.

Example :

Oromoo koo rabbii issin wallin haa jiratuu jabadhaa.

Nutti osoo jiruu Itoophiyaan hin diigamtu hin Yaalinaa.

Tokkumman Oromoo kan duraa calaatti ittitee dhadhaa ba’u qaba.

```python
from nltk import word_tokenize, WordNetLemmatizer
tokens = [word_tokenize(sen) for sen in data.Text_Clean]
```

Figure 14: Tokenization

After it is tokenized with the above python code it looks like :

[Oromoo, koo ,rabbi, issin, wollin, haa, jiratuu, jabadhaa]

[Nutti, osoo, jiru, Itoophiyaan, hin, diigamtu, hin, Yaalinaa]

[Tokkumman, Oromoo, kan, duraa, calaatti, ittitee, dhadhaa, ba'u, qaba]

4.3.2. Stop Word Removal

Stop word removal comes after tokenization and normalization in the preprocessing stage. Stop word lists are carefully inspected for the objectives of this study. This is done by removing any verbs and nouns from the stop word list that are more or less directly related to the primary subjects of the underlying collections and replacing them with non-content bearing words. But in sentiment analysis there are some stop words in Afaan Oromo which indicates negativity because of this we did not remove those stop words. Appendix A contains the general stop word lists which is collected for this study.

Figure 16: shows the iplimentation that reads a stop word list from a text file and compares it to a tokenized term. Then, if the word is similar to another, it is removed from the index words.

Example: **Sample input from the prepared corpus.**

Oromoo koo rabbii issin wallin haa jiratuu jabadhaa.

Nutti osoo jiruu Itoophiyaa diigu osoo yaalu batanni.

Tokkumman Oromoo kan duraa calaatti ittitee dhadhaa ba'u qaba.

Oromoon walqooduu dhiisee tokkummaa isaa cimsachuu qaba.

Wallin tane biyya kenyaa gudissu qabna wal qoqodu dhisne.

```
text_file=open('AfaanOromoStopwords.csv',encoding ='unicode_escape').read()
```

```
def remove_stop_words(tokens):
    return [word for word in tokens if word not in stoplist]
```

```
filtered_words = [remove_stop_words(sen) for sen in lower_tokens]
result = [' '.join(sen) for sen in filtered_words]
data['Text_Final'] = result
```

Figure 15: Stop word removal implimentation

The output of the above review corpus after the stop words are removed is as shown below. Some stopwords are used to indicate negative feelings in sentiment analysis because of that we don't remove those stop words like 'hin' as shown in the sample output.

Sample output after stopwords are removed.

Oromoo rabbii wollin jiratuu jabadhaa.

Nutti jiruu Itoophiyaa diigu yaalu batanni.

Tokkumman Oromoo duraa calaatti ittitee dhadhaa qaba.

Oromoon walqooduu dhiisee tokkummaa cimsachuu qaba.

Wallin tane biyya kenyaa gudissu qabna wal qoqodu dhisne.

4.3.3. Removal of Punctuations and Special Characters

Punctuation mark removal is an important NLP preprocessing step; these marks, which are used to divide text into sentences, paragraphs, and phrases, affect the results of any text processing approach, particularly those that rely on word and phrase occurrence frequencies, because punctuation marks are used frequently in text[57]. For the quality of our data to feed the model and get high accuracy we removed all the punctuation marks and special characters from the collected dataset such as !@#$^&*,"; are removed to increase the quality of the dataset and achieve better accuracy of the classification model.

```
def remove_punct(text):
    text_nopunct = ''
    text_nopunct = re.sub('['+string.punctuation+']', '', text)
    return text_nopunct

data['Text_Clean'] = data['Text'].apply(lambda x: remove_punct(x))
```

Figure 16:Removal of punctuations and special characters

4.3.4. Lower Case Conversion

This is also another preprocessing stage in dataset cleaning. Lowe case conversion is transforming all the dataset collected into uniform lower-case form. In our case we converted our dataset into lower case with the below shown experiment. This enhances the model's performance and training stability. We normalize our data set into lower case using the blow implimentaion.

Example: **Sample input from dataset for lowercsa conversion.**

BARREEFFAMA KAMUU DUBISTEE BIRA KUTUUN GAARIIDHA.

Jabaadha Ummata Kenya Gootota Kenyaa.

```
def lower_token(tokens):
    return [w.lower() for w in tokens]

lower_tokens = [lower_token(token) for token in tokens]
data['tokens']=lower_tokens
```

Figure 17:Lower case conversion implementation

Sample output after it is converted to lowercase.

barreeffama kamuu dubistee bira kutuun gaariidha.

jabaadha ummata kenya gootota kenyaa.

4.4. Proposed Model Implementation

We developed a CNN in our experiment to classify reviews into extreme, anti-extreme and neutral categories and to assess the effectiveness of our model method. Concatenated word vectors of the text are required as input for the convolutional neural network design. We also used word embedding with convolutional neural network for feature extraction of the dataset review to feed the model. The convolution neural network is implemented with the below python code on google colab platform to experimentally test our collected review classification in to its correct category.

4.4.1. Parameteres Used For Convolutional Neural Network

Parameters are configured for the convolutional neural network (CNN) to train and build the model. We implemented values of batch size, number of iteration, optimizers, dropout rate and activation functions. We train our model using the training data and optimizing parameters on test split. To train our parameter, we mainly based on the back-propagation algorithm which updates parameters on every training example, throughout the time. For parameter optimization and the loss function minimization, we employ relu activation function because relu function is used for multiclass sentiment classification and in our proposed model we have three classes because of that in the proposed model relu functionis used for determining the output of the model to one of the three classes and Adam optimizer algorithm with dropout rate of 0.2 or 20% percent of the input randomly dropout from the model to minimize the overfitting problem happens during the implementation process of the model and for each input to regularize during input data. All the parameters used in the convolutional neural network is generalized in the below table.

Parameters used in the CNN model	**Value**
Batch size	68
Dropout	0.2 or 20%
Number of iteration(epochs)	10 epochs
Optimization algorithm	Adam Optimizer
Activation function	Softmax function
Activation function	ReLu activation function
Embedding dimension	100 embedding dimension is used
Evaluation metrics	Accuracy, recall,precision and F1-score

Table 10:Convolutional neural network parameters used

```python
def ConvNet(embeddings, max_sequence_length, num_words, embedding_dim, labels_index):

    embedding_layer = Embedding(num_words,
                            embedding_dim,
                            weights=[embeddings],
                            input_length=max_sequence_length,
                            trainable=True)

    sequence_input = Input(shape=(max_sequence_length,), dtype='int32')
    embedded_sequences = embedding_layer(sequence_input)

    convs = []
    filter_sizes = [2,3,5,6]
```

```python
for filter_size in filter_sizes:
    l_conv = Conv1D(filters=200, kernel_size=filter_size, activation='relu')(embedded_sequences)
    l_pool = GlobalMaxPooling1D()(l_conv)
    convs.append(l_pool)


l_merge = concatenate(convs, axis=1)

x = Dropout(0.2)(l_merge)
x = Dense(128, activation='relu')(x)
x = Dropout(0.2)(x)
preds = Dense(labels_index, activation='softmax')(x)

model = Model(sequence_input, preds)
model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['acc'])
model.summary()
return model
```

Figure 18:Convolutional neural network model implimentation

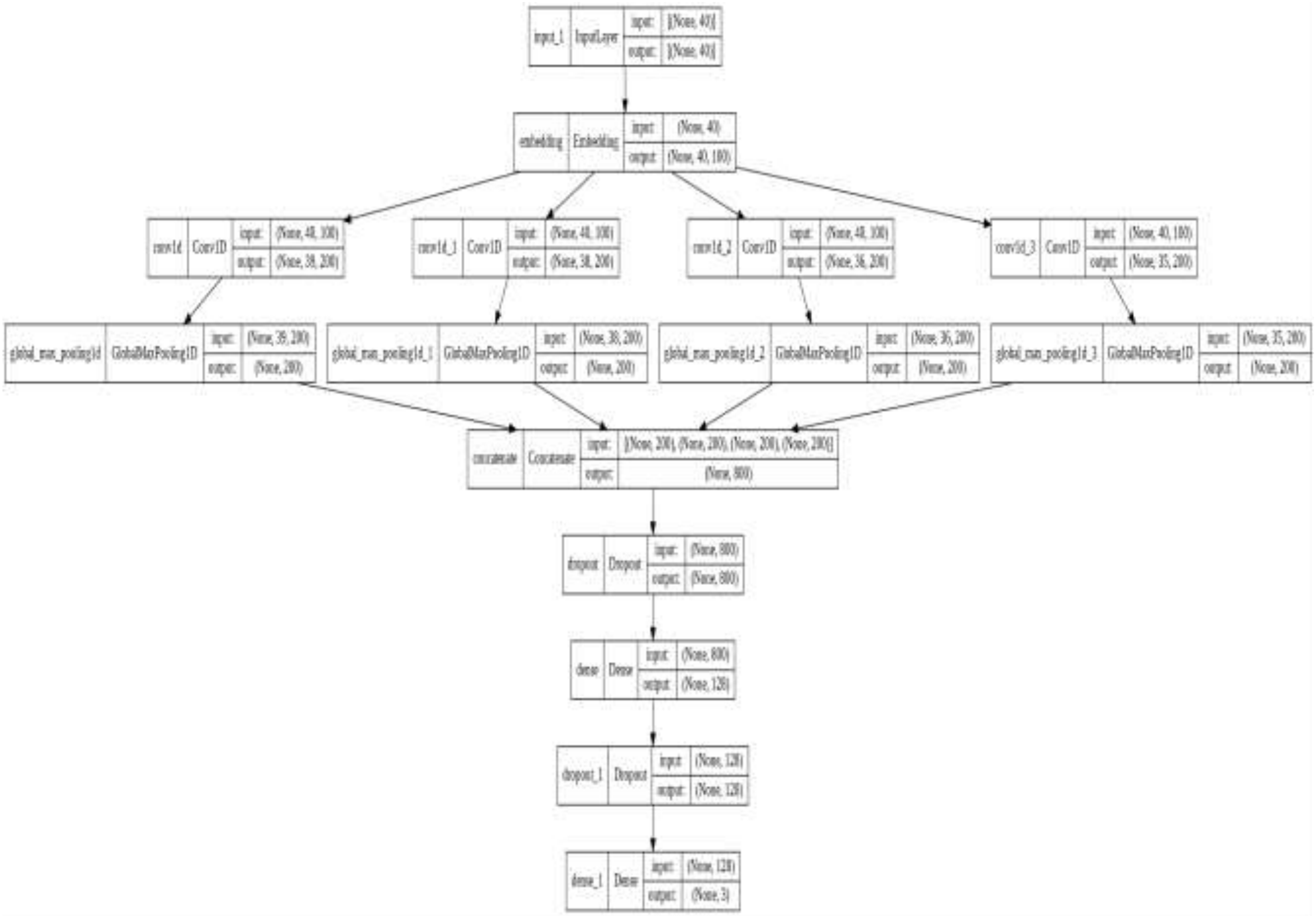


Figure 19:sample output of the CNN architecture

4.5. Accuracy And Validation Test of The Proposed Model

We experimentally tested the accuracy of the model with various parameters to rich up the maximum performance of the designed model. In our experiment we used 10 epochs for evaluating the model accordingly the proposed model achieves the highest accuracy in the 10 iteration or epochs. The experiment shows that when the number of epochs lower the model accuracy become decrease but according to our experiment convolutional neural network model performs better when the dataset volume increases.

Accuracy is used to test the performance of the overall implemented model; how much the model is accurate in classifying the reviews into its correct category as per the labeled dataset feed to the network. In the proposed convolutional neural network its performance is measured with accuracy, the ablity of the model on the provided dataset to classify the reviews into extereme,anti -extreme and neutral classes. The accuracy of the model is visualized in the below graph.

Figure 20:Training and validation accuracy

We experimentally test the performance of the model as shown in figure 21 to analysis the training accuracy and validation accuracy of designed model for the state of the art. As it is experimentally visualized the model performs better on tenth (10) epochs at high accuracy on the classification of the review to its correct category.

4.5.1. Accuracy, Precision, Recall and F1 score of the proposed work

The number of classifications a model successfully predicts divided by the total number of predictions is known as model accuracy. It's one method of evaluating a model's performance, but it's far from the only one. In fact, a wide range of rich metrics exist for this purpose like precision, recall and F1 score to evaluate the model accuracy. However, accuracy provides the best view on how well a model performs on a particular dataset when examining many of them at once rather than any single one in isolation. Based our experiment for convolutional neural network the proposed model achieves good prediction accuracy of 80% after training the model with training and testing dataset.

The other measurement metrics for testing the accuracy of the model is through precision measure. Precision measures the number of correctly identified Positive samples divided by the total number of Positive samples is used to determine precision (either correctly or incorrectly). The model's accuracy in classifying a sample as positive is assessed. Positive incidents as a fraction of total positive cases projected. The denominator is the positive model prediction from the full dataset. We also used a precision to check the performance of our model through experimental.

The other measurement metrics used during our experiment is recall of the proposed model. Recall is a metric that measures how many correct positive predictions were produced out of all possible positive predictions. Unlike precision, which only considers the right positive predictions out of all positive predictions, recall considers the positive predictions that were missed. All the measuring metrics were experimentally tested for our model which is convolutional neural network to check the overall performance of our proposed model. Accordingy the convolutional neural network achieves 76% precision, 77% recall and 77% of F1-score accuracy for the proposed model. The experimental result for accuracy, precision recall, and F1 is summarized in the figure below.

	precision	recall	f1-score	support
0	0.77	0.81	0.79	125
1	0.90	0.86	0.88	258
2	0.62	0.66	0.64	99
accuracy			0.80	482
macro avg	0.76	0.77	0.77	482
weighted avg	0.81	0.80	0.80	482

21: Summary of Precision, recall and fi score

The experimental result shows that the proposed model performs better accuracy for our state of art problem on the prepared review.

4.6. Other Deep Learning Algorithm Implementation

We try to compare the result of our proposed model with other neural network models to investigate the performance of our model experimentally. We develop a Long Short-Term Memory (LSTM) which is one of the well-known models employed in text and sentiment classification mostly. Long short-term memory (LSTM) is a deep learning architecture that uses an artificial recurrent neural network (RNN). LSTM has feedback connections, unlike normal feedforward neural networks. We used the same data set which is used for the proposed model and followed all the data preprocessing steps to train and test the LSTM model to compare its performance with the state-of-the-art model. But the LSTM model achieves lower performance when it is compared with convolutional neural network. The model achieves accuracy of 70% which is too lower accuracy when it is compared with the convolutional neural network.

4.6.1. LSTM Model

We developed a LSTM in our experiment to classify reviews into extreme, anti-Extreme and neutral categories and to assess the effectiveness of our proposed model with it. Concatenated word vectors of the text are required as input for the long short term memory design. We also used word embedding called word2vec with the model for feature extraction of the review to feed the long short term memory model. The LSTM network is implemented with the below python code on google colab platform to experimentally test our collected review classification in to its correct category with long short-term memory model.

4.6.2. Parameteres Used For Long Short Term Memory(LSTM)

Parameters are configured for the long short term memory (LSTM) to train and build the model as well as for coparison purpose of the proposed model with the LSTM model for the sentiment classification. We implemented values of batch size, number of iteration, optimizers, dropout rate and activation functions. We train the model using the training data and optimizing parameters on test split. To train our model, we used the back-propagation algorithm which updates parameters on every training process. For parameter optimization we employ relu activation function because relu function is used for determining the sentiment class and in our proposed model we have three classes we used relu function for determining the output of the model to one of the three classes and adam optimizer algorithm with dropout rate of 0.2 or 20% percent of the input randomly dropout from the model to minimize the overfitting problem happens during the implementation process of the model and for each input to regularize during input data. We also used softmax function parameter in LSTM model because it is used in multi class classification and in the proposed model we have three classes extreme,anti-extreme and neutral class. Parameters we implemented in the long short term memory is summarized as in the table below:

Parameters used in the LSTM model	Value
Batch size	68
Dropout	0.2 or 20%
Number of iteration(epochs)	10 epochs
Optimization algorithm	Adam Optimizer
Activation function	ReLu activation function
Activation function	Softmax function
Embedding dimension	100 embedding dimension is used
Evaluation metrics	Accuracy, recall,precision and F1-score

Table 11:Parameters used in the LSTM model

The long short term developed model is implemented on the colab online platform with the below pyhton code to classify the review into extreme ,anti-extreme and neutral classes.

```
def rnn(embeddings,
        max_sequence_length,
        num_words,
        embedding_dim,
        labels_index):

    embedding_layer = Embedding(num_words,
                                embedding_dim,
                                weights=[embeddings],
                                input_length=max_sequence_length,
                                trainable=True)

    sequence_input = Input(shape=(max_sequence_length,),
                                  dtype='int32')
    embedded_sequences = embedding_layer(sequence_input)
    lstm = LSTM(128)(embedded_sequences)

    x = Dense(128, activation='relu')(lstm)
    x = Dropout(0.2)(x)
    preds = Dense(labels_index, activation='softmax')(x)
    model = Model(sequence_input, preds)
    model.compile(loss='binary_crossentropy',
                  optimizer='adam',
                  metrics=['acc'])
    model.summary()
    return model
```

Figure 21:LSTM model

The sample output of the developed long short term memory architecture used for the extremist affiliation sentiment analysis to classif the dataset into the proposed classes.

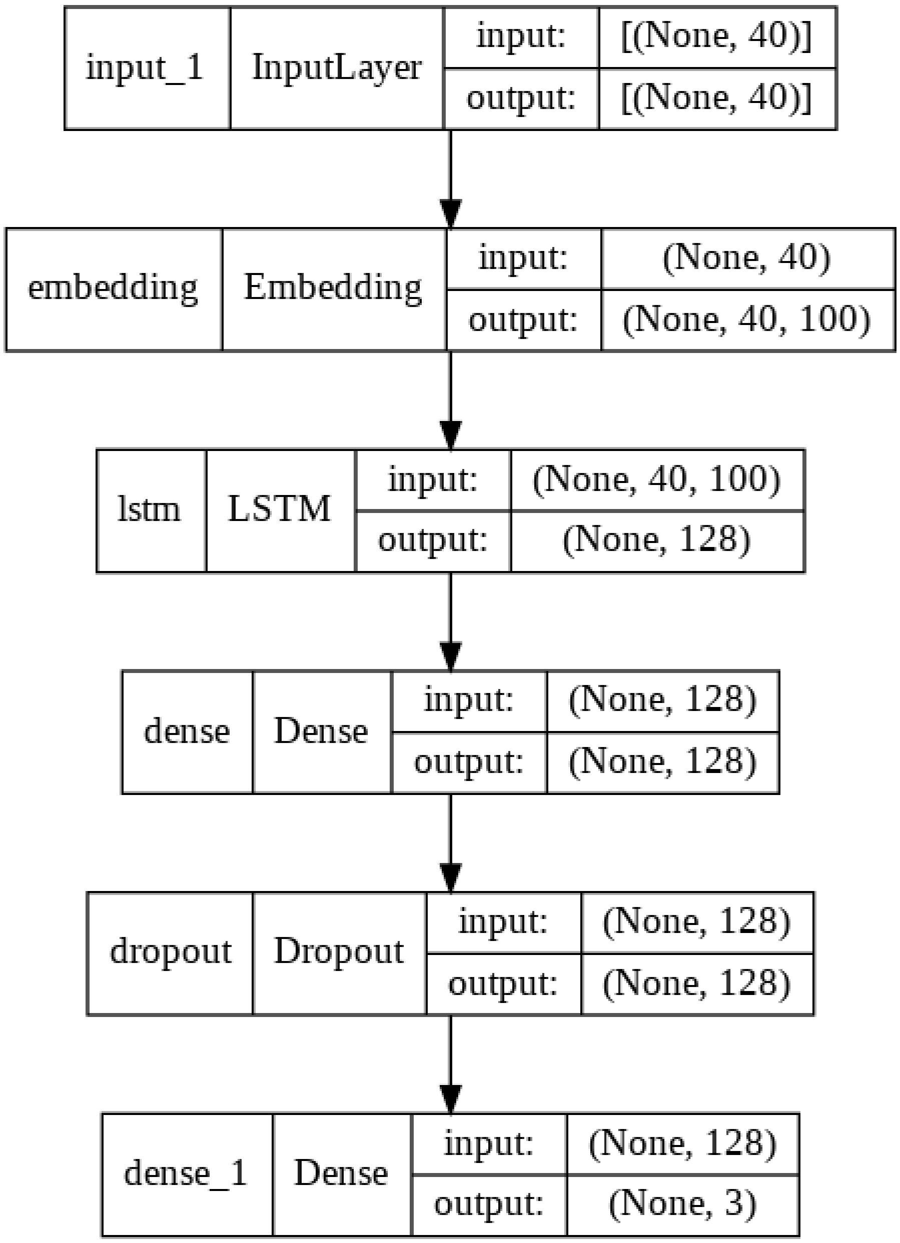


Figure 22:Architecture of the lstm model

4.6.2. Experimental Result of LSTM

We experimentally tested the accuracy of the model with various parameters to rich up the maximum performance of the Long Short-Term Memory. In our experiment we used epochs which is similar with our proposed model for evaluating the model. Even though the proposed model convolutional neural network achieves the highest accuracy with the same epochs than LSTM. The experiment shows that when the number of dataset small the model accuracy become decrease but according to our experiment convolutional neural network model performs when the number dataset is large the performance also increase. Based on the experiment long short-term memory achieves accuracy of 70% on the dataset of our proposed model. We also used various metrics used to evaluate the performance of the LSTM model such as precision, recall, F1-score. As we experimentally investigated the LSTM score precision of 71%, recall of 71% and F1-score of 70% respectively. Finally, we can conclude that convolutional neural network performs better accuracy than the long short-term memory based on our experiment on sentence level sentiment analysis for Afaan Oromo.

	precision	recall	f1-score	support
0	0.60	0.77	0.67	125
1	0.79	0.68	0.73	148
2	0.75	0.68	0.71	154
accuracy			0.70	427
macro avg	0.71	0.71	0.70	427
weighted avg	0.72	0.70	0.70	427

Figure 23: Accuracy, Precision, Recall and f1 score of LSTM

4.6.3. Accuracy and Validation test of the LSTM Model

We experimentally tested the accuracy of the model with various parameters to rich up the maximum performance of the designed model. In our experiment we used 10 epochs for evaluating the model accordingly the LSTM model achieves the highest accuracy in the 10 epochs. Finally, the long short term memory(LSTM) model achieves accuracy of 70% which 10% lower performance than the proposed model as dipcted in figure 25. From this we can genralize that the convolutional neural network shows promising result in extremist review classification than long short term memory from the investigated experiment prepared for the mmodel.

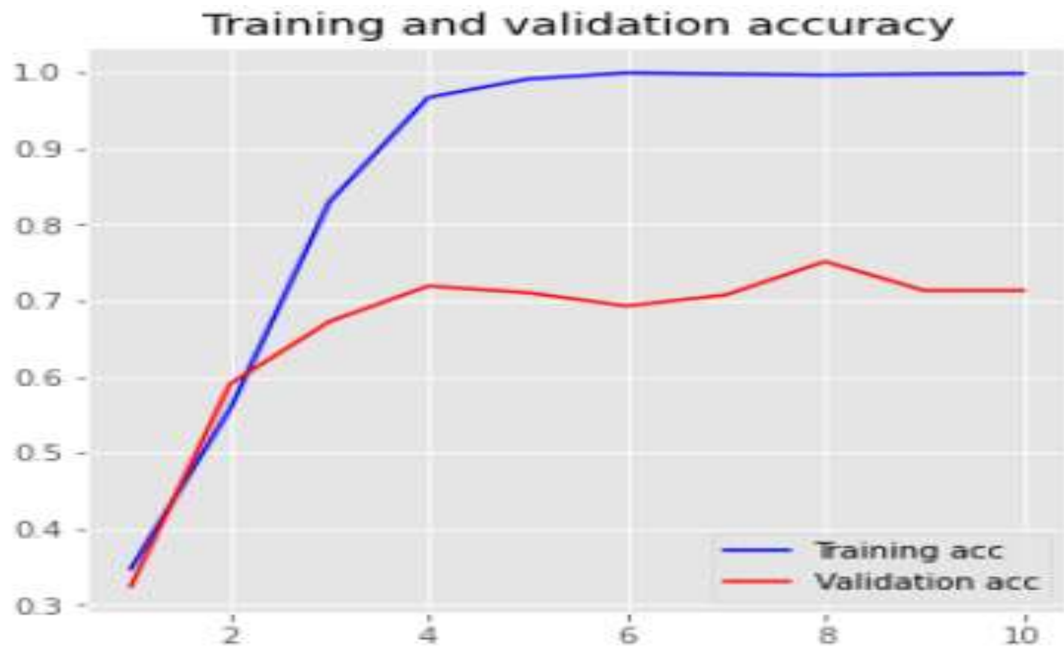


Figure 24: Training and validation accuracy of LSTM model

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1. Conclusion

The internet has drastically altered how people express their thoughts and beliefs. More and more people are able to freely share their thoughts on a wide range of topics, including political, economic, religious, ethnic, movie, product, etc. They freely express their comments, attitudes and feelings toward those various issues posted on social network such as Facebook, twitter, YouTube and any other social media platforms without any limitation. Service providers; the social networks and governments utilize these reviews to make more informed judgments and improve their services. However, the volume of reviews is expanding at such a rapid rate that people are finding it increasingly difficult to sift through them all and extract relevant information.

This study attempted to examine sentiment mining approaches for Afaan Oromo writings that were opinionated, posted and commented in various social platforms by different users on three domains mainly politics, religious, and ethnic written by Afaan Oromo only.

One of the challenging parts of sentiment analysis model developments is about the review or dataset for the model due to the unavailability of annotated dataset. It takes time to label dataset for the designed model this was one of the challenging activities during our work. In our case it was very tedious and time consuming to prepare the corpus for our model. The opinionated text in the online platforms is mostly have spelling error since any one can simple comment and post without any consideration about the spelling error of the writing system of Afaan Oromo. To overcome the mentioned problem, we try to correct the spelling error manually to increase the quality of the corpus. We collect the extremist related contents, anti-extreme and neutral as a dataset for our proposed work. The study only concerned with text data than using audio, videos, and images that have extremist related content. Pre-processing like stop word removal, conversion to lower case, tokenization and polarity classification techniques are all used to classify a given opinionated sentence or text into predetermined classifications. The dataset was labeled manually to three different classes extreme, anti-extreme and neutral; tagged with numerical values 0, 1, and 2 respectively by experts. After it is manually labeled the dataset goes through various preprocessing such as tokenization each text is split to individual words, stop words are removed all the stop words which cannot cause change in meaning of text is removed from the dataset and also the dataset is converted to lower case. Finally, the dataset is split in to training and testing we employed 80% of the dataset for training purpose and 20% of the dataset for testing the accuracy of the model. The study attempted to use word embedding which is word2vec as a method to find the word similarity with in the review dataset to increase the accuracy of the developed model with dimension of 100. We developed word embedding from our dataset because there is no pretrained word embedding for Afaan Oromo. As stated on state-of-the-art convolutional neural network model used to experimentally classify the review into its correct category. We proposed convolutional neural network with different layers because it have higher ability to classify sentiment to its correct class than other deep learning algorithms that why we implemented to overcome the state-of-the-art problem. We used softmax function in the proposed model. The Softmax Activation Function, also known as Soft Argmax or Normalized Exponential Function is a fascinating activation function that takes vectors of real numbers as inputs, and normalizes them into a probability distribution proportional to the exponentials of the input numbers. The softmax function is used for multi class classification and we have three classes in the proposed model we used softmax function for classification purpose. We also used gradient descent optimizer function to optimize our model. The favored method for optimizing neural networks and many other machine learning methods is gradient descent Adam optimizer function. For deep learning model training, Adam is an alternative for stochastic gradient descent. Adam combines the finest features of the AdaGrad and RMSProp methods to create an optimization technique for sparse gradients on noisy issues.

After the dataset passed through various pre-processing stages it is feed to the convolutional neural network and the proposed model achieves a promising accuracy of 80% on sentence level sentiment analysis for extremist-based content. For a comparison purpose we implemented long short-term memory with same data set, training and testing parameters and the model achieves accuracy of 70% which is lower than our proposed model for the state of the art. Finally, our model shows a promising accuracy of 80% for the state-of-the-art problem. In the future by developing a standard corpus for Afaan Oromo the model able to predict better than the current work. So, by increasing the quality and volume of the dataset it is possible to increase the accuracy of the model in the near feature. Finally, we conclude that convolutional neural network good at classifying sentence level sentiment analysis for Afaan Oromo.

5.2. Recommendation

This research aims to construct a sentence-level sentiment mining model for opinionated Afaan Oromo language texts based on extremely affiliated comments and posts regarding political, religious, and ethnic domains written in Afaan Oromo from social media platforms. To develop a more efficient sentiment analysis system, a coordinated team effort is required that includes linguistic professionals, computer scientists, and others with experience collecting large numbers of comments from the web due to a lack of prepared corpus in the Afaan Oromo language to develop sentiment analysis models. This study has put in a lot of effort in that direction of sentiment analysis. Even though, there are a number of difficulties that should be researched further in the future in area of Afaan Oromo sentiment analysis development.

There are other research areas that could be pursued in the future to further the current study to develop more accurate model than the current study in the future.

- Other deep learning models and algorithms should be investigated to improve the performance of the sentiment classification technique.
- Preparing Afaan Oromo standard corpus is one of the most difficult aspects of opinion mining. As a result, future research should investigate developing a huge corpus that may be used in future studies in the field of sentiment analysis.
- In addition, sentiment reviews and comments in other sectors such as product, commerce, news, and film could be a future study in Afaan Oromo.
- It is also better to incorporate sentiment analysis of audio/voice data for Afaan Oromo as a future work in the area.
- This research's result can be utilized as an input for a recommendation system, so it's possible to expand this work to create a full-fledged recommendation system based on user review.

References

[1] M. Park and P. Facebook, “Facebook Reports Fourth Quarter and Full Year 2020 Results,” pp. 1–12, 2021.

[2] B. Dean, “Active Users of YouTube in 2021” *Backlinko*, pp. 1–22, 2021, [Online]. Available: https://backlinko.com/youtube-users.

[3] B. Dean, “Active Users of Twitter in 2021?” *Backlinko*, no. November 2013, pp. 1–21, 2021.

[4] Brian Dean, “ Users Active Social Media users Worldwide in 2021” *Backlinko*, no. July, pp. 1–28, 2021, [Online]. Available: https://backlinko.com/social-media-users.

[5] D. Chaffey, “Global Social Media Statistics Summary 2016,” *Smart Insights*, vol. 2016. pp. 1–16, 2015, [Online]. Available: http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/.

[6] Y. E. Sagduyu, A. Grushin, and Y. Shi, “Synthetic social media data generation,” *IEEE Trans. Comput. Soc. Syst.*, vol. 5, no. 3, pp. 605–620, 2018, doi: 10.1109/TCSS.2018.2854668.

[7] M. Oljira, “Sentiment Analysis For Afaan Oromoo Using Combined Convolutional Neural Network And Bidirectional Long Short-Term Memory,” Vol. 11, No. 11, Pp. 101–112, 2020, Doi: 10.34218/Ijaret.11.11.2020.010.

[8] V. P. Verma, “Extremist Text Post Detection on Social Media : Review,” vol. 7, no. 02, pp. 87–92, 2021.

[9] Sofea Azrina Et Al.“Terrorism Detection Based On Sentiment Analysis Using Machine Learning”2017 .

[10] P. Sreevidya, S. Veni, A. V. Vidyapeetham, A. V. Vidyapeetham, and A. Info, “Sentiment analysis by deep learning approaches 1,3,” vol. 18, no. 2, pp. 752–760, 2020, doi: 10.12928/TELKOMNIKA.v18i2.13912.

[11] A. Kaur et al.“Detecting Radical Text over Online Media using Deep Learning.”,30 July,2029.

[12] C. Core, “6 Media and Extremism,” no. 1, pp. 157–180.

[13] J. Baldauf *et al.*, “Hate Speech and Radicalisation Online: The OCCI Research Report,” 2019, [Online]. Available: https://www.isdglobal.org/wp-content/uploads/2019/06/ISD-Hate-Speech-and-Radicalisation-Online-English-Draft-2.pdf.

[14] D. K. Citron, “Notre Dame Law Review Extremist Speech, Compelled Conformity, and Censorship Creep,” *Notre Dame Law Rev.*, vol. 93, no. 3, 2018, [Online]. Available: https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=4772&context=ndlr.

[15] C. Sentiment, M. Model, F. O. R. Opinionated, A. Oromo, T. On, and E. Politics, “Constructing Sentiment Mining Model For Opinionated Afan Oromo Texts On Ethiopian Politics ” Unpublished Thesis ,2017.

[16] E. S. Usop, R. R. Isnanto, and R. Kusumaningrum, “Part of speech features for sentiment classification based on Latent Dirichlet Allocation,” *Proc. - 2017 4th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2017*, vol. 2018-Janua, pp. 31–34, 2017, doi: 10.1109/ICITACEE.2017.8257670.

[17] J. vom Brocke, A. Hevner, and A. Maedche, “Introduction to Design Science Research,” no. November, pp. 1–13, 2020, doi: 10.1007/978-3-030-46781-4_1.

[18] K. Jilcha Sileyew, “Research Design and Methodology,” *Cyberspace*, no. August, 2020, doi: 10.5772/intechopen.85731.

[19] N. Wayessa and S. Abas, “Multi-Class Sentiment Analysis from Afaan Oromo Text Based On Supervised Machine Learning Approaches,” vol. 7, no. 7, pp. 10–18, 2020.

[20] M. Taboada, “Sentiment Analysis : An Overview from Linguistics Sentiment Analysis : An overview from Linguistics Maite Taboada,” no. February, 2017, doi: 10.1146/annurev-linguistics-011415-040518.

[21] A. Khattak *et al.*, “Fine-grained sentiment analysis for measuring customer satisfaction using an extended set of fuzzy linguistic hedges,” *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 744–756, 2020, doi: 10.2991/ijcis.d.200513.001.

[22] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, “Cross-cultural polarity and

emotion detection using sentiment analysis and deep learning on covid-19 related tweets," *IEEE Access*, vol. 8, no. October, pp. 181074–181090, 2020, doi: 10.1109/ACCESS.2020.3027350.

[23] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-Based Sentiment Analysis using BERT," *Proc. 22nd Nord. Conf. Comput. Linguist.*, pp. 187–196, 2019, [Online]. Available: https://www.aclweb.org/anthology/W19-6120.

[24] T. A. Consultant, "An Introduction to Sentiment Analysis About me !!," *Adv. Soft Comput. Its Appl.*, 2013, [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-45111-9_41.

[25] V. K. Bongirwar, "A Survey on Sentence Level Sentiment Analysis," *Int. J. Comput. Sci. Trends Technol.*, vol. 3, no. 3, pp. 110–113, 2015, [Online]. Available: www.ijcstjournal.org.

[26] J. Shen, X. Liao, Z. Tao, V. Processing, V. Processing, and A. Intelligence, "Proceedings of spie," no. November 2019, 2021, doi: 10.1117/12.2550215.

[27] I. Chaturvedi, E. Cambria, R. E. Welsch, and F. Herrera, "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," *Inf. Fusion*, vol. 44, no. December 2017, pp. 65–77, 2018, doi: 10.1016/j.inffus.2017.12.006.

[28] H. Qin, R. Gong, X. Liu, X. Bai, J. Song, and N. Sebe, "Binary neural networks: A survey," *Pattern Recognit.*, vol. 105, no. 5, 2020, doi: 10.1016/j.patcog.2020.107281.

[29] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, no. c, pp. 53040–53065, 2019, doi: 10.1109/ACCESS.2019.2912200.

[30] Hans-Delter Wehle "Deep learning " conference paper, july 2017.

[31] T. Hopfield, J. Hopfield, and T. Bam, "Recurrent Neural Network " vol. 23, no. 4. pp. 1–6, 2018.

[32] Subarno Pal. et al."Sentiment Analysis in The Light of Lstm Recurrent Neural Networks" ,2017.

[33] Kumar Bandl,"Convolutional Neural Network "pp. 5–6, 2018.

[34] J. M. Czum, "Dive Into Deep Learning," *J. Am. Coll. Radiol.*, vol. 17, no. 5, pp. 637–638,

2020, doi: 10.1016/j.jacr.2020.02.005.

[35] H. Kim and Y. S. Jeong, “Sentiment classification using Convolutional Neural Networks,” *Appl. Sci.*, vol. 9, no. 11, pp. 1–14, 2019, doi: 10.3390/app9112347.

[36] C. Ko and H. Chang, “LSTM-based sentiment analysis for stock price forecast,” pp. 1–17, 2017.

[37] H. A. Elnemr, N. M. Zayed, and M. A. Fakhreldein, “Feature Extraction Techniques,” pp. 264–294, 2015, doi: 10.4018/978-1-4666-8654-0.ch013.

[38] L. Ma and Y. Zhang, “Using Word2Vec to process big text data,” *Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015*, no. October 2015, pp. 2895–2897, 2015, doi: 10.1109/BigData.2015.7364114.

[39] K. O. Owoeye, “Classification of Extremist Text on the Web using Sentiment Analysis Approach.”,2017.

[40] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and I. Awan, “Detection and classification of social media - based extremist affiliations using sentiment analysis techniques,” *Human-centric Comput. Inf. Sci.*, 2019, doi: 10.1186/s13673-019-0185-6.

[41] “Ethiopians_ Amhara and Oromo - International Institute of Minnesota”,International Institute of Mensota,2021 .

[42] “University of Namibia | Language Centre.” [Online]. Available: http://www.unam.edu.na/language-centre.

[43] Ibrahim Bedane, “The Origin of Afaan Oromo : Mother Language,” *Double Blind Peer Rev. Int. Res. J.*, vol. 15, no. 12, 2015.

[44] E. Beekan G., “Oromo People.” [Online]. Available: https://scholar.harvard.edu/erena/oromo.

[45] A. Oromo, “Afaan Oromo,” pp. 10–13, 2017.

[46] D. Bijiga, “The Development of Oromo Writing System School of European Culture and Languages,” *Http://Kar.Kent.Ac.Uk*, 2015.

[47] M. Oljira, “Sentiment Analysis of Afaan Oromo using Machine learning Approach,” vol. 7, no. 9, pp. 7–15, 2020.

[48] A. Hunegnaw, M. Meshesha, A. Ababa, and E. A. Science, “Sentiment Analysis Model for Afaan Oromoo Short Message Service Text : A Machine Learning Approach Turkish Journal of Computer and Mathematics Education,” vol. 12, no. 13, pp. 332–342, 2021.

[49] S. M. Yimam, H. M. Alemayehu, and U. Hamburg, “Exploring Amharic Sentiment Analysis from Social Media Texts : Building Annotation Tools and Classification Models,” 2018.

[50] Y. Getachew, “Deep Learning Approach For Amharic Sentiment Analysis Deep Learning Approach For Amharic Sentiment Analysis” *ReaserchGet*, no. March, 2019.

[51] A. F. Tela, “Sentiment Analysis for Low-Resource Language : The Case of Tigrinya,” no. May, 2020.

[52] T. Liu, S. Fang, Y. Zhao, P. Wang, and J. Zhang, “Implementation of Training Convolutional.”

[53] M. A. Nasichuddin, T. B. Adji, and W. Widyawan, “Performance Improvement Using CNN for Sentiment Analysis,” *IJITEE (International J. Inf. Technol. Electr. Eng.*, vol. 2, no. 1, 2018, doi: 10.22146/ijitee.36642.

[54] R. Yamashita, M. Nishio, R. Kinh, G. Do, and K. Togashi, “Convolutional neural networks : an overview and application in radiology,” pp. 611–629, 2018.

[55] B. Mele and G. Altarelli, “Lepton spectra as a measure of b quark polarization at LEP,” *Phys. Lett. B*, vol. 299, no. 3–4, pp. 345–350, 1993, doi: 10.1016/0370-2693(93)90272-J.

[56] W. Etaiwi and G. Naymat, “The Impact of applying Different Preprocessing Steps on Review Spam Detection,” *Procedia Comput. Sci.*, vol. 113, pp. 273–279, 2017, doi: 10.1016/j.procs.2017.08.368.

Appendix A: Afaan Oromo stop words

The following lists are some of the Stop words in Afaan Oromo.

immoo haa mee akka ishii akkasi utuu kan kee hanga jechuun ol waan akkam henna oliif waggaa akkasumas hoggaa kanaaf oliin akkuma hogguu kanaaf yammuu hoo kanaafu osoo yemmuu ammo ille otoo yeroo an keenya otumalle ykn ani innaa keenyaa otuu yommi booda inni keeti otuullee yommu booddee isaa keetii saniif yoo dura isaan koo silaa isee kun simmoo yookaa eega yookiin eegana iseen malee sun yookinim oo eegasii ishee moo ta`ullee yoom ennaa isheenu tahullee garuu erga itumalle nuti tanaaf Jechuu fakkeenya af ituu nuyi tanaafi oggaa fi ituullee odoo tanaafu fkn Jechaa ofii tawullee kan ini ini isaa ofii yoom ammo akkasu mas booddee erga eegasii eega jechuu kanaafi jechaa otuu otoo ituu akkum dura saniif waan tahullee ituullee ta'ullee otuulle henna innaa waggaa hoggu yommuu yemmuu yommiisimmo oo woo akkam ituu hanga aanee eenaa irra isiniif keessatti oo tawullee agarsiisoo erga irraa isiniin osoo teenya akka ergii irraan isinirraa koo otoo teessan akkam f isa isinitti kun otumalle tiyya akkasumas faallaa isaa ittaanee lafa otuu too akkum fagaatee isaaf itti lama otuullee tti akkuma fi isaan ittumallee malee saaniif utuu ala fkn isaani itu manna sadi waa’ee alatti fullee isaanii ituullee maqaa sana waanalla fuullee isaanitiin jala moo saniif waggaa amma gajjallaa isaanirra jara na si wajjiin ammo gama isaanitti jechaan naa sii warra an gararraa isaatiin jechoota naaf siif woo ana garas isarra jechuu naan siin yammu ani garuu isatti jechuun naannoo silaa yeroo ati gidduu isee kan narra simmoo yommii bira gubbaa iseen kana nati sinitti yommu booda ha ishee kanaa nuu siqe yoo boodde hamma isheen kanaaf nu’i sirraa yookiin dabalatees hanga ishii kanaafi nurra sitti yoom dhaan henna ishiif kanaafuu nuti sun ufuu dudduuba hoggaa ishiin Kanaan nutti tahullee dugda hogguu ishiirra kannatti nuu tana dura hoo ishiitti karaa nuuf tanaaf duuba illee isii kee nuun tanaafi eega immoo isiin keenna nuy tanaafuu eegana ini isin keenyaa odoo ta’ulle eegasii innaa isini keessa ofi ta’uyyu yoolinimoo inni isinii keessan ogga ta’uyyuu