JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY (JiT)

SCHOOL OF COMPUTING

DEPARTMENT OF INFORMATION TECHNOLOGY

Sentiments Analysis for Afaan Oromoo Socio-Politics Text: Deep Learning Approach

BY: - Amanuel Assefa Workineh

A Thesis Submitted to Faculty of Computing of Jimma Institute of Technology in Partial Fulfillment for the Degree of Master of Science in Information Technology

Jimma, Ethiopia

October 25, 2019

JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY (JiT)

SCHOOL OF COMPUTING

DEPARTMENT OF INFORMATION TECHNOLOGY

Sentiments Analysis for Afaan Oromoo Socio-Politics Text: Deep Learning Approach

BY:-Amanuel Assefa Workineh

Name and Signature of Members of the Examining Board

Name	**Title**	**Signature**	**Date**
Dr. Wondwossen Mulugeta	Advisor	________________	________________
________________	Co-Advisory	________________	________________
________________	Chairperson	________________	________________
________________	External Examiner	________________	________________
________________	Internal Examiner	________________	________________

Jimma, Ethiopia

October 25, 2019

Acknowledgements

First of all, I would like to thank the Almighty God, for his countless blessings throughout my life and this research.

This research work would not have been accomplished without the support of many people. Hence, I would like to express my gratitude to all the people who helped and supported me during this study.

I would like to express my deep and sincere gratitude to my research advisor, Dr. Wondwossen Mulugeta for agreed to supervise my thesis; without his guidance and support this thesis would not have been successful.

I would also like to thank my co-advisor Kibret Zawude for his continual support, encouragement, and patience.

I would also like to express my gratitude to all my friends for giving me support and continuous encouragements throughout my thesis.

Words are not ample to express my love and appreciation to my beloved wife for her endless love, care and great support through the duration of my studies. Last but not least, I would express my special thanks to all of my family members for their love, prayers, caring and sacrifices for teaching me.

Dedication

This work is dedicated to my son, Milto Amanuel.

Abstract

Sentiment analysis, also known as emotion analysis, has recently become one of the growing areas of research related to natural language processing and machine learning. It is a method to determine emotional feedback of people towards an event. Much sentiment about specific topics are available online; it allows several parties such as customers, companies and governments to explore these opinions. An opinion may be positive, negative or neutral depends on individual's judgment or evaluation towards a topic. The analysis of natural language text for the identification of subjectivity and sentiment has been well studied in terms of the English language, Japanese and others. Conversely, the work that has been carried out in terms of Afaan Oromoo remains in its infancy; thus, more cooperation is required between research communities in order to offer a mature sentiment analysis system for Afaan Oromoo.

This thesis addresses the problem of sentiments classification of Afaan Oromoo language reviews on Facebook social media and provides the rationale behind the proposed methods to enhance the performance of sentiment analysis in the Afaan Oromoo language. The first step is to increase the resources that help in the analysis process; the most important part of this task is to have annotated sentiment corpora. We also describes the work undertaken by the author to enrich sentiment analysis in Afaan Oromoo by building a new Afaan Oromoo Sentiment Corpus. The data is labeled not only with two polarities, but the neutral sentiment is also used during the annotation process. The second step includes apply preprocessing on textual data makes ready for learning features and classifying Afaan Oromoo text into different polarity.

Generally the model is generated by a neural network variance called Convolutional Neural Network because it have shown great promise in the task of sentiment classification .The classifier model obtains a classification accuracy of 79.99%, which is encourage the researchers to continue in this direction of research.

Keywords Word embeddings, Sentiment analysis, Natural language processing, classification, machine learning, Afaan Oromoo, Convolutional Neural Network

Contents

List of Tables

List of Figures

List of Abbreviations

Adam	Adaptive Moment Estimation
API	Application Program Interface
CBOW	Continuous Bag of Words
CNN	Convolutional Neural Network
FN	False Negative
FP	False Positive
HTML	HyperText Markup Language
IR	Information Retrieval
LSTM	Long Short Term Memory
ML	Machine Learning
MLP	Multi-Layer Perceptron
NB	Naive Bayes
NLP	Natural Language Processing
NN	Neural Network
POS	Part of Speech
ReLU	Rectified Linear Units
RNN	Recurrent Neural Network
SA	Sentiment Analysis
SVM	Support Vector Machines
SGD	Stochastic Gradient Decent
TF	Term Frequency
TF-IDF	Term Frequency-Inverse Document Frequency
TP	True Positive

CHAPTER ONE: INTRODUCTION

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

1.1. Background

The World Wide Web technology provides a great place, opportunity and platform that allows a people to share their feeling and gain knowledge from the information. In fact, this technology plays a vital role on presenting and the expanding of web forums, social network websites, reviews, blogs and others; hence it provided for researchers with a huge source of user-generated online contents.

Nowadays, people are using social media platforms like Facebook and Twitter to broadcast their opinions, feelings, thoughts and ideas on different issues. These online comments or opinions can be about several topics like movies, news, electronic products, cars, politics, and many others. According to[1], internet contains a wealth of information that people can use to help them make a decision about a given issue. Several parties such as individuals, organizations, customers, and governments are analyze and explore these opinions to know what other people think and feel about their products, services and new rules or policy they have set. In fact, it is not common for many people to ask a friend / relatives in the manual way of sentiment analysis, when someone wants to some services. The only thing required is internet needed to surf through the unstructured information from feedback, comments, and blog-posts. People usually try to ascertain other people's opinions that are found online about politics, news, products, countries that they are considering traveling to and spending time in, or movies that they are thinking of watching in a cinema. For instance, the rapid growth of e-commerce has caused the people to buy more from online shops and stores, thus people started to review comments about these products and learn from other people's experiences to get a general idea about these products in order to help them in making the best choice [2].

A plentiful opinion and sentiment about specific issues could be collected and analyzed from websites and social media. It is a challenging task for a people to analyses and draw conclusions manually from opinionated content because the data increasing rapidly.

Therefore, sentiment analysis should be able to surf such information and bring it in structured format to the end users.

According to Pang and Lee [3], sentiment Analysis is one of an important application of Natural Language Processing (NLP) and has proved beneficial for several tasks such as answering systems and information extraction. It is a process of finding people's opinion, attitude, views and detecting emotions towards any entity [4]. The entity can be individuals, products, events or topics. In another way it is a computational task for automatically detecting and classifying sentiment from text, document or from sentences to finding its polarity or orientation. The polarity of the documents can be positive, negative or neutral. Many researchers are engaged and started to become more interested in automatic analyzing public opinion for a better investigation and classification reviews to acquire beneficial information by identifying the sentiment expressed in the form of text.

In most of research works, mainly two approach were employed to analyze the sentiment expressed in a text, those are semantic orientation approach and machine learning approach to analyze the sentiment expressed in a text. Machine learning is a method where algorithms are designed and trained to be efficient and accurate as they can be when doing different predictions on data. Machine learning has successfully been applied in a variety of applications such as natural language processing, text or document classification, speech and image recognition as well as recommendation systems to name a few. One area within machine learning is deep learning, it is a technique where the model is learning by using several non-linear processing layers to process the data.

A deep learning model can be described as a model of two neurons, input, and output, where data is sent through the input layer. The input layer sends the data onto the hidden layers, where it is examined at different levels and features. Deep learning can be used in different learning scenarios such as unsupervised, supervised, and hybrid networks and for different problems like classification and regression.

As many of the articles witnessed among these two approaches, the machine learning has shown good result[4]–[8]. Consequently, in this research work we also implement a deep machine learning approach in supervised learning scenarios.

Afaan Oromoo is one of the major African languages that is widely spoken and used in most parts of Ethiopia and some parts of other neighbor countries like Kenya and Somalia [9][10]. It is used by Oromoo people, who are the largest ethnic group in Ethiopia, which amounts to 34.5% of the total population[10].

The number of individuals and organizations that are using Afaan Oromoo language 'qubee' are rapidly increasing from time to time to express their feeling over internet and social media. They are publishing a huge amount of important information concerning to any business, company and government in many websites by using this language.

Because of the scarcity of resources and tools in Afaan Oromoo language is big challenging to do successful research especially in sentiment analysis field, it is mandatory to conduct studies to build model that identifying sentiment associated with in the text.

1.2. Motivation

A social networking service is allows users to create, share information, and ideas that express sentiment and influence others. With the growth of these social websites, an attention is given to information extraction on these messages. Henceforth, due to the high number of users of Afaan Oromoo in social media it is need to understand people's emotion especially in socio-politics.

Choosing to work with the Afaan Oromoo language is due to several factors. First, naturally peoples have ability to understand emotions, analyze the situations and the sentiments associated with it. But, how efficiently can we train a machine to exhibit same phenomenon becomes an important and vital question to be explored from Afaan Oromo language perspective.

Second, Afaan Oromoo sentiment analysis is importance due to its already large scale audience and huge availability of data on the internet.

And last but not least, most of the tools are available for the English and other language, it is important to build a model for Afaan Oromoo to classify sentiment of text, in order to reduce the effort of analyze, suggest and provide important information for the user and motivated researcher to take this problem as our research work.

1.3. Statement of the Problem

Sentiment analysis is one of the newest emerging research fields caused by the great opinionated web contents, and it's due to the emergence of the Web 2.0 technology. This web 2.0 generated a massive amount of raw data by enabling Internet users to post their opinions, reviews, and comments on the web. For instance, peoples around the world are widely using social networking websites in order to communication with each other, government or public organization used to announce if any changes or modification exist in policies or strategies, mass media use social media networks to broadcast news and other service. In fact those websites are an influential medium for sharing ideas, feelings and knowledge. Presently, there are several social networking websites, such as Facebook, and Twitter. Facebook is the most popular social networking website, which covers around 2.32 billion monthly active worldwide users as of 4th quarter 2018 [11]. It has become a common communication tool in people's day-to-day life. In the Facebook world, people can express or write their feeling or opinions on any discussion topics, such as socio-political issues, economic issues, and entertainment issues like movie reviews, technology issues and many others. This open style encourages people to naturally express their perceptions[12].

Due to the typically large volume of reviews, manually looking through them can be infeasible or processing such raw data to extract useful information is very challenging task. An example of important information that can be automatically extracted from the users' posts and comments is their opinions on different issues, events, services, products. This problem of sentiment analysis has been studied well on the English language with very limited research done for other languages. Due to this reason, the scientific community is interested in studies in other languages.

As far as researcher knowledge is concerned, there are only two studies done by[13],[10] on Afaan Oromoo using lexicon based and rule based approach. Since the granularity of text or level sentiment classification vary in nature; the previous work done is not tackled document level text classification of Afaan Oromoo. To the best of our knowledge, this is the first attempt to explore deep learning models for sentiment classification in Afaan Oromoo language.

Publically unavailability dataset/resource for Afaan Oromoo sentiment analysis is another challenge that restricts research community for further studies; so, providing a dataset also useful for research community.

This work have aimed to address the following major questions.

1. What challenges are observed during labelling Afaan Oromoo text into different polarities?
2. What preprocessing tools and procedures applied to come up with quality data sets for training and testing set of Afaan Oromoo text?
3. How can deep learning be used to detect polarity of user reviews in Afaan Oromoo?

This research work attempted to address the above stated problems and answer the outlines research questions.

1.4. Objectives of the study

General objective

The general objective of this study is to build sentiment analysis model for Afaan Oromoo using convolutional neural network architecture.

Specific objective

In order to meet the main goal of the study, the following specific objectives was identified.

- To explore extensive study of literatures on the sentiment analysis approaches.
- To collect and prepare corpus of socio-political domain for the performance evaluation.
- To construct model for sentiment analysis.
- To conduct an experiment as a means of evaluating the usefulness of the extracted information and the accuracy of classifications.

1.5. Scope and Limitations of the study

The key goal of this study is to develop sentiment analysis of Afaan Oromoo language text or reviews. This research is done based on data collected from FBC official Facebook page of Afaan Oromoo news. Specifically, we focus on socio-political news domain of people's displacement, immigration, public relations and political parties. Currently, in our country peoples are expressing their feeling freely on social media by different mechanism like text, emoji, picture and voice. In this study we treated only text based user generated content.

The approaches we used in order to meet our study goal was included word embedding and deep neural network. This investigation gives a devotion for direct opinion, for instance let see direct expression "*Kun raajjiidha lammiin lammii miidhuu haa dhabuu*" this sentence is shows the direct feeling of writer.

This research covers Facebook platform sentiment classification of Afaan oromoo text into positive and negative polarity. This study is not cornered with aspect level sentiment analysis and question answering related opinion. Since sentiment is highly domain dependent, this study is only deal with specific of socio-politics domain. The opinion holder identification and reasons for positive and negative classifications are not covered in this research work.

1.6. Methodology

In order to solve a research problem in a systematically way it a required to follow an appropriate research methodology. This research was conducted to constructing Sentimental analysis model for Afaan Oromoo socio-political text of Facebook platform. The research was adopted the following methodologies as the main focus of achieving its objectives:

1.6.1. Research design

In this research work we follows design science research methodology. The design science research paradigm is a widely used by Information Systems researcher and designate their work as design science for many years. It is referred to as a problem-solving paradigm because it aims at building model that are aimed at addressing a problem[14]. As stated in [14] ,the artefacts address the problems or enhance existing solutions and are important tools for arriving at research outcomes and reviewing them to decide how the artefact adopted can be further utilized.

In this research, the goal is to build an automated analysis of Facebook social media user's comments to identify their sentiment and opinions about the socio- politics news. Hence, the six steps design science process were followed.

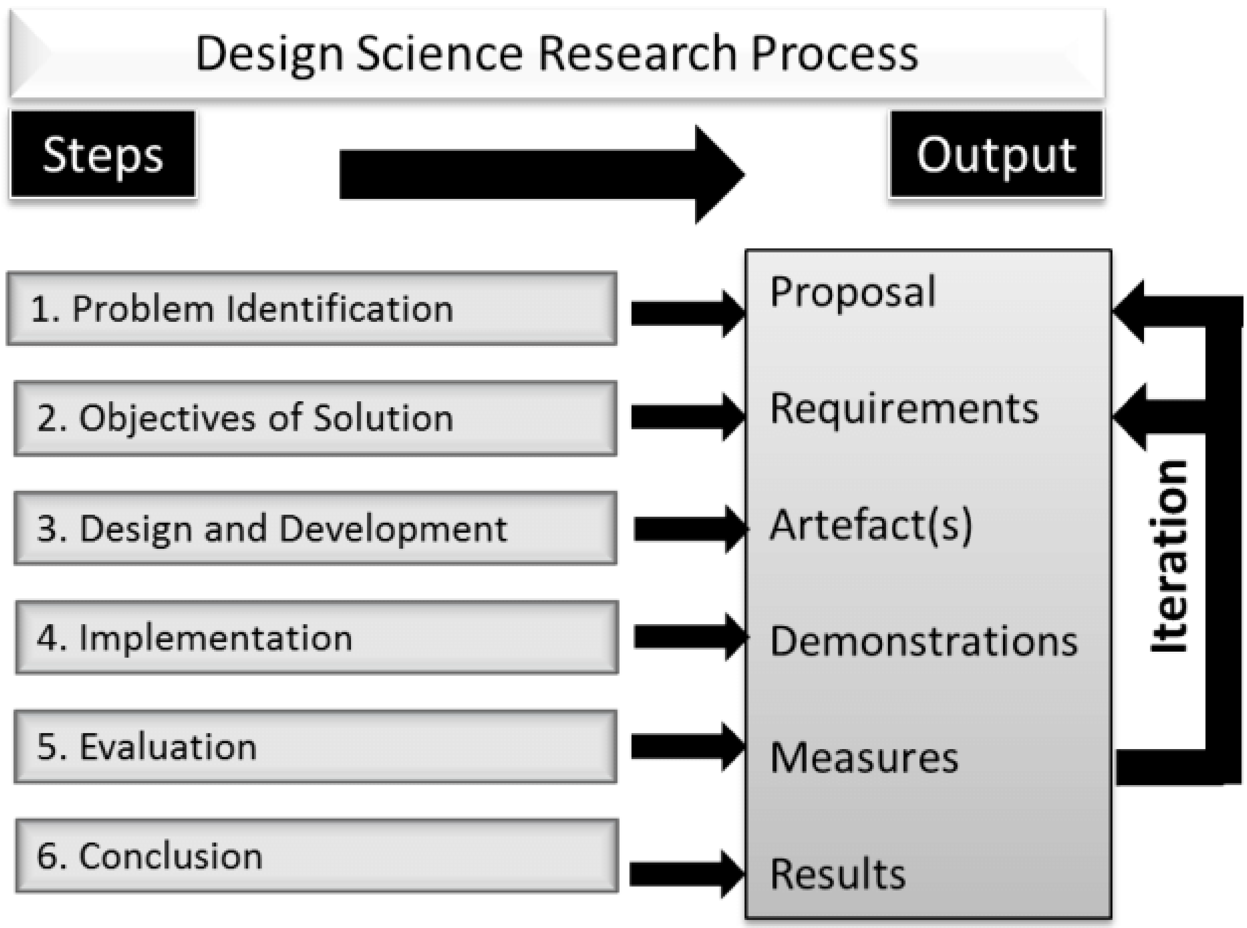


Figure 1: Steps in the Design Science Research Methodology (adapted from)

1.6.2. 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 Oromoo text which passed through many phase of development task. In the first phases, collect or extract the opinion from Facebook platform. Next, we understands the nature of socio-politics 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 social media we gave more attention on data cleaning. Then vocabulary was constructed. Moreover, designing an algorithm, implementing with programming language and training was done. In the final phase evaluation was done.

1.6.3. Data source and data collection methods

To realize the objective of our study, the researcher were collected different reviews manually from Facebook social media (https://www.facebook.com/afanoromofana), on selective socio – political Afaan Oromoo news. Our data source was FBC Afaan Oromoo official Facebook page because it is legal under the Facebook company terms and condition. Even though the news posted on this page covers many topics, we pay special attention to socio-politics domain focusing on displacement/flee, migration, public relation, and political parties. The reason behind choosing these domains are the availability of user generated content in Afaan Oromo language pretty good than in others domain.

We collected the data by using a tool called export comments by creating developer account. We was collected totally 12928 reviews from thus issue and then reduced to 12662 during preprocessing steps because of some of reviews was written by non Afaan Oromoo language while they was commented the news. We used 12662 comments and common 8675 vocabulary file for the purpose of training and testing the model.

Table 1: reviews collected from FBC Afaan Oromoo

		FBC	
S.No	Domain/issues	Num. Posted	Num. of comments
1.	People's Displacement	12	2,469
2.	Public relation	10	3,957
3.	Immigration	8	2,384
4.	Political parties	52	4,118
Total num. of post and comments		82	12,928

1.6.4. Design Approach

To model sentiment classification of Afaan Oromoo 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.

1.6.5. 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 to store data collected by export comment tool, Python 3.6.6 , Keras ,Tensor flow, and CNN are used for implementation of the model.

1.6.6. Evaluation techniques

This research focus on designing and developing sentiment analysis model using deep neural network approach where evaluation of the model is an important task.
Model Evaluation is an integral part of the model development process. Measure the performance of a machine learning model can be done in a several ways. For classification task a straightforward approach is to calculate the percentage of correctly classified and or incorrectly classified data points. This measure is called accuracy and as the name indicates measure how accurately the machine learning model can classify unseen data points [15]. For performance evaluations after training models, usually training and test accuracies are considered. It helps to find the best model that represents our data and how well the chosen model will work in the future. Accuracy is a reliable metric of performance when the dataset used is balanced, i.e. the number of data points in each class are evenly distributed. Say that a dataset contains 50% positive and 50% negative data points [15]. In order to evaluate our sentiment analysis model, we used the accuracy score evaluation metric. Consider that each testing instance, after processed by an automated system to detect if it expresses a positive or negative sentiment, can either be:

-True Positive (TP) - system prediction is positive, as well as the real value.

-False Positive (FP) - system prediction is positive and the real value is negative.

-False Negative (FN) - system prediction is negative and the real value is positive.

-True Negative (TN) - system prediction is negative, as well as the real value.

A better performing system will as a result produce a higher accuracy score than a worse system. Accuracy can be computed as the overall correctness of the model i.e., the decisions that model got right divided by all the total number of decisions made by the model and calculated by the following formula.

$$\text{Accuracy} = \frac{TP+TP}{TP+TN+FP+FN} \qquad 1.1.$$

True positive and true negative are the amount of correctly predicted comments/reviews. False positive and false negative are the amount of reviews of which the sentiment was predicted incorrectly.

Therefore, we evaluate the model using two experiments by changing training and testing split 90%, 10%:80%, 20% of size of the dataset, setting different number of epochs and network layers.

1.7. Significance of the study

One of the main goals of sentiment analysis is facilitating users' information search on a specific issue because people usually consider others' opinions before making a decision. Thus, sentiment is important in studies of news, public opinion, and political polarization. In this situation sentiment analysis has useful tool for organization, government and individuals in order to make effective decision.

Individual’s users often spend some time surfing the internet in order to establish the opinions of other users on several issues such as politics, sports, news, and other service, but such tools easily identify people’s sentiment on news posted Facebook social media on specific issues of socio-politics. For government and public organization, such tool very useful to analyzing political trends, identifying ideological bias e.g.in news texts, evaluate of public opinions (on policies, parties, government agencies, politicians) to overcome on negative judgement and resume their activities of positive feelings.

Because of the high volume of data posted to internet every day reaching and processing all the data, analyzing the sentiment behind it concerning some issues on the internet almost impossible.

For example providing such tools in socio-politics domain very chief, because recently in our country socio-politics issues very close to people's heart. It is important for government to understand people's sentiments on socio-politics issues and accordingly to change the strategy to provide good service for citizens. Thus, a tool that can obtain and analyze user reviews in order to understand the final sentiment is valuable to government, public organization and individual. Hence, this research work is investigate sentiment analysis for Afaan Oromoo text on socio-politics domain.

Contributions

The main contribution of this research is an investigation of classification algorithms for extracting sentiments from Facebook FBC news comments. For this purpose, CNN methods are studied. In a convolutional neural network that converts words into word embeddings and then passes these embeddings through the layers to extract the polarity of comments. Therefore, the aim of the thesis is to perform experiments and investigate the performance of algorithms detecting positive and negative comments/review of Afaan Oromoo language.

Since annotated corpus not available for Afaan Ormoo Sentiment analysis, we start this work with building our annotated corpus of sentiment data. Hence, this study is somehow fill the limitation of the resource in this language.

1.8. Thesis Structure

Relevant background theory and fundamental concepts involved in the sentiment analysis/classification, in addition to the tools and data sets used are described in Chapter 2. Chapter 3 describes the methodologies used in the sentiment classification of other related language. Chapter 4 describes the detailed implementation of related technologies like word segmentation, word embedding, and Convolutional Neural network, pooling, softmax. Experiments conducted on this 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

Sentiment analysis, also known as opinion mining[16], [17] is a rapidly developing field of study that analyzes people's opinions, feelings, and emotions towards whole range topics. The words 'sentiment' and 'opinion' are often used interchangeably, also in this thesis. It is one of the most active research areas in natural language processing, yet is still a relatively young field with a long way to go, and progress to be made[16]. Even though a manually sentiment analyses have been used earlier in societies, the web 2.0 is extremely contribute for quick development of sentiment analysis of this days.
Platforms such as Facebook, Twitter, discussion forums, and blogs nowadays provide a vast amount of user-generated content that can be collected, processed and used for scientific research such as sentiment analysis or opinion mining. There are many applications based on sentiment analysis techniques have been utilized for various purposes such as determining the polarities of customer reviews[18] tracking political opinions[19], predicting stock market movements[20].
This chapter provides an overview of the studies in the field of sentiment analysis, it begin with introducing the notion of sentiment analysis. Meanwhile, many researchers have been conducted studies at various level of sentiment analysis; each level i.e. document level, sentence level and aspect of sentiment analysis is reviewed respectively in this chapter. Since previous works related with sentiment analyses have laid the groundwork for the current research activity, reviewing of different literatures covered in this chapter. In order to have a detailed understanding of the background and techniques of sentiment analysis, several evolutionary stages of sentiment analysis are presented. Furthermore, two main approaches for sentiment analysis, which are the semantic orientation approach and the machine learning approach, will be discussed in detail. In the end, the research gaps will be identified through overall discussions.

2.2. The notion of sentiment analysis

Sentiment analysis measures the polarity or tonality of texts by identifying and assessing expressions people use to evaluate or appraise persons, entities or events[3].According to [21] the word '*sentiment*' is defined as 'an attitude, thought or judgment prompted by feeling', it is also defined as 'a specific view or notion: opinion' and 'emotion'. The word '*opinion*' is usually referred to as 'a view, judgment or appraisal formed in the mind about a particular matter'. Sentiment analysis [16], [17],[22], [3] is a research area that aims at detecting the author's sentiment, emotions or opinion about the events, topics or individuals expressed in text. Sentiment analysis is sometimes also referred to as opinion mining, and usually these two terms have identical meaning [17]. Nevertheless, some researchers explain that there exist small differences in notions of these two terms [17] . For example, [23] point out that opinion mining came from the information retrieval (IR) community and aims at first extracting and then processing opinions about an entity, while sentiment analysis originates from the natural language processing (NLP) community and is concerned with detecting the sentiment in given text. Moreover,[3] provide information about the terminology and the history of appearance of these two terms, but conclude that in the broad context these two terms represent the same field of study. As Liu [16] state that, there are also other many names and slightly different tasks, for example, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining etc. Accordingly, both the terms 'sentiment analysis' and 'opinion mining' used as a synonym in this thesis.

The field of sentiment analysis or opinion mining has recently gained a lot of attention from the researchers and marketers; there has been a steady undercurrent of interest of analyzing opinions[21].Before explosion of web 2.0 only a little opinioned text available, this is main reasons for lack of the studies in the field of Sentiment. This circumstance was leads much of the early research on textual information processing to focused on mining and retrieval of factual information, such as information retrieval, text classification or text clustering [24].

Liu [16] states that an individual usually asked his/her friends or family for opinions before making a decision and an organization normally conducted opinion polls, surveys and focus groups to find out the sentiments of the general public about its products or services. Thereafter, many platforms were designed and developed to realize the vision web 2.0 technologies. Also societies were begin to be familiar with the platforms such as blogs, forums, Twitter, Facebook and various other types of social media; this is create a enormous chance to express their opinions and emotions by posting reviews of products or services online. Likewise, the companies, partnerships, industries' and organizations can modify their marketing strategies through social media monitoring and analysis. However, it can still be a formidable task to find opinions sources and monitoring them on the World Wide Web, because there are a large number of diverse sources such as online forums, discussion groups, and blogs. Also each source may have a huge volume of user-generated content indicating feelings or emotions [21]. In the manually mechanism of sentiment analysis it is difficult to identify and extract the relevant information from the texts with opinions and summarize them. Accordingly, an automatically system is needed to discover and analyze the online opinionated texts.

In natural language processing (NLP), sentiment analysis includes diverse aspects concerning how information about emotions, attitudes, perspectives and social identities is conveyed in language[21].

2.3. Levels of sentiment analysis

As the previous study indicates, sentiment analysis can be carried out at three different levels: document level, sentence level and aspect level[16],[3].

2.3.1. Document level

At this level the main task is to define opinion of the whole document, opinion should be expressed about specific topic [22]. Let's take an automatic sentiment analysis system which serves for e-commerce as an example. In e-commerce, every product has its own page. A document-level sentiment analysis task would be read all reviews in that page or about that particular product and expresses an overall sentiment or opinion for that product. The assumption indicates that a sentiment analysis system classifies the overall polarity of a customer review about a specific product. This means ,one document expresses opinions on a single object, such as customer reviews of products and services, because usually the result of sentiment analysis only have two (positive and negative) or three outputs (positive, negative and neutral). However it is common that there might be a few different opinions in one document, thus it is not applicable to documents in which opinions are expressed on multiple products. There are a number of researchers who have carried out document-level sentiment analysis [22] [3]. They mainly focus on how to separate the positive texts from negative texts automatically and they also have presented different methods to improve the accuracy.

2.3.2. Sentence level

At this point, every sentence is considered as a short document which can be subjective or objective. As author is argues in[25],subjective is an opinionated sentence that expresses sentiment. The aim sentence level sentiment analysis is to recognize sentiment or opinion for that particular sentence, instead of overall sentiment. In another word, it involves determining whether each sentence expressed a neutral, positive, or negative opinion. For instance,

1. *'qafoo bilbiilla teekinoo kana kaleessan bitaadhe; qarshiin isaas* ***gariidha****,qulqullini kaameera isa immoo* ***bayee mishaadha'***
2. *torbaan darbee obboolleettiin koo bilbillaa teekinoo biitatee turte,garuu immoo sabaaba qulqullinni kaameeran isa* ***gadi buha*** *ta'eef ni* ***jibbitee****'*
3. *bilbilla teekinoo kana bakka heeddutti arge.*

The first review is categorized as positive, the second is categorized as negative and third is as neutral because positive indicator word **gariidha** and **mishaadha,** negative indicator **jibbite,** and no sentiment terms are attached to these reviews. According to[16],the sentences are considered as short documents and which indicate as there is no fundamental difference between document-level and sentence-level sentiment analysis. In sentence level sentiment analysis commonly two sub-tasks are performed. In the first stage, determining whether the sentence is a subjective sentence or an objective sentence; next if the sentence is subjective, determining whether it expresses a positive or negative opinion[16] .This level is related to the subjectivity classification which is to distinguish the subjective sentences that express sentiments or views from objective sentences that express factual information. The subjectivity classification is very important, because it filters out those sentences that contain no opinions. The value of neutral usually indicates the objective sentences or sentences absent of opinions.

2.3.3. Aspect level

Aspect level, sometimes also called feature level or feature level in some studies [3].In this level, the main task to infer the sentiment polarity of a specific aspect by a given sentence. Both document level and sentence level is worthwhile in various cases, but they are unsatisfactory to providing the necessary details for an application, because they do not identify sentiment targets or assign opinions to these targets [16]. At the document level, a positive document on an object does not mean that the author has positive opinions on all aspects of this topic. Besides the sentence-level classification of sentiment is often an intermediate step since it is more useful to know what features or entities of the object the opinions are on. The aspect level of sentiment analysis focuses on opinions itself instead of looking at the constructs of documents, such as paragraphs, sentences and phrases. It is not enough just to find out the polarity of the opinions; identifying the opinion targets is also essential. The aspect-level sentiment analysis can be decomposed into two sub-tasks: aspect extraction and aspect sentiment classification [16]. The task of aspect extraction can also been seen as an information extraction task, which aims to extract the aspects that opinions are on.

For our pervious mobile reviews example, if we have a sentence from **teekinoo** review *'qafoo bilbiilla teekinoo kana kaleessan bitaadhe; baatiirin isaas **dafee dhuma**,qulqullini kaameera isa immoo **bayee mishaadha**'*. Aspect level task is try expresses positive emotion for the aspect "kaameera/camera" and negative emotion for the aspect "baatiiri/battery". The basic approach of extracting aspects is finding frequent nouns or noun phrases, which are defined as aspects. Then the text containing aspects are classified as positive, negative or neutral[26]. Liu [16] points out that the accuracy at aspect level sentiment is still low because the existing algorithms still cannot deal with complex sentences well.

2.4. Approaches of Sentiment analysis

According to the type of sentiment and the levels of classification, the employed techniques for sentiment classifications are vary. Sentiment analysis techniques can be applied on different kinds of data, such as news, reviews, blogs, or social networking and microblogging messages. Every data type has its own characteristics, which must be taken into account during data collection, preprocessing, and feature construction. Likewise, sentiment analysis is applied in many situations such as tracking and aggregating opinions about a product, a company, detecting opinions for or against some movement or political party, predicting a sentiments of book, movie, financial assets, and in many other situations. Data from social network and microblogging Web sites is especially interesting for research and applications because of its large volume, popularity, and capability of near-real-time publishing of individuals' opinions and emotions about any subject[25].

Subsequently many researchers have been studying sentiment analysis, especially in the era of Web 2.0 they use different methods and algorithms that help to performing sentiment analysis. There are two main approaches for sentiment classification: machine learning and lexicon based methods[3].We discuss an approach in detail in the following sections. As [27] argues ,lexicon-based approach is focuses on words and phrases as the indicators of semantic orientation and the overall of the polarity of the text is an averaged sum of the polarities of indicators. Lexicon based approach is also called sematic orientation. The second approach is machine learning, which is concentrating on selecting the appropriate machine-learning algorithm and the right features of texts to classify the polarities of the text [5]. The following figure is adopted from Medhat [17],shows the details of sentiment analysis approach of both a machine learning and lexicon based with it sub-division or types.

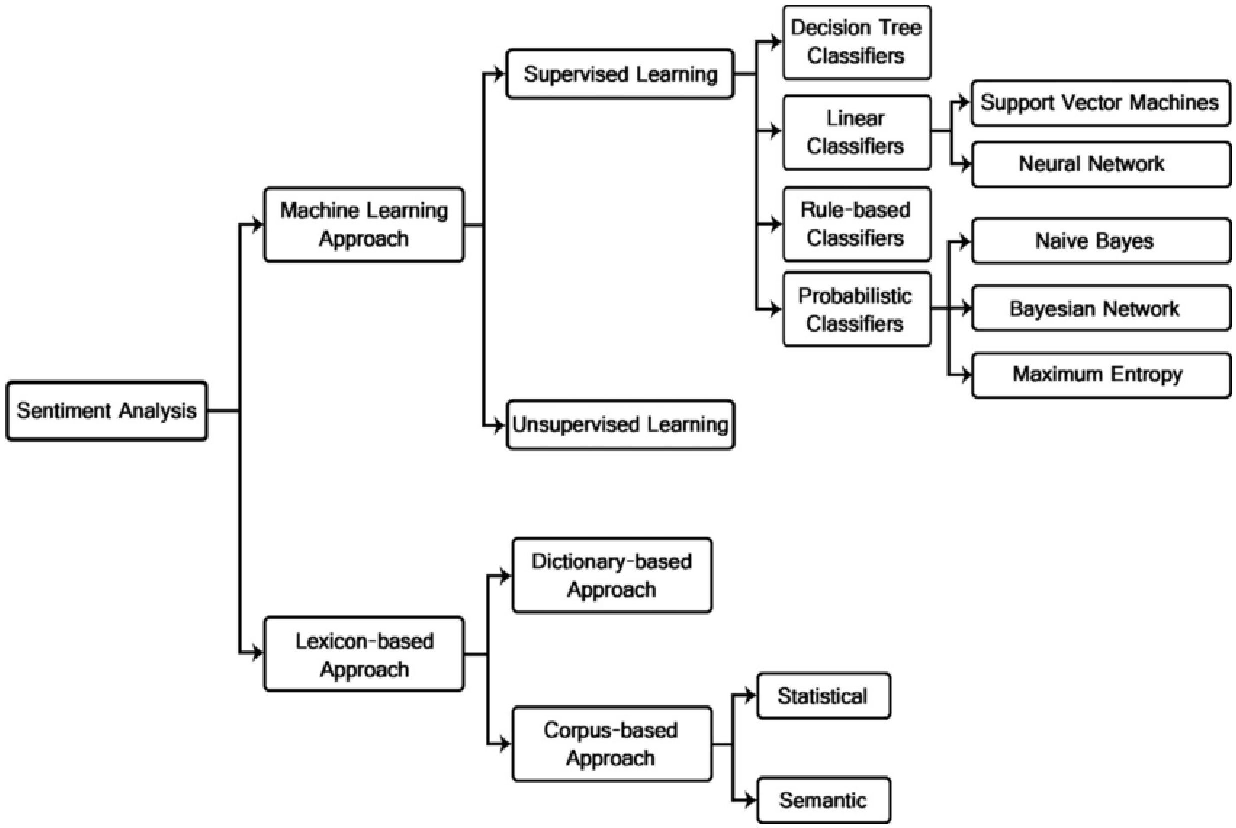


Figure 2: Sentiment analysis approaches (adopted from [12])

2.4.1. Lexicon-based approach

In this approach, the fundamental idea is sentiment lexicon; this sentiment lexicon is the plays a vital role during the process of sentiment analysis, such as collections of sentiment phrases or dictionaries of sentiment words.

As Turney and Littman [28], lexicon based approach has varied directions of polarity and intensity. Positive semantic orientation of a word denotes a desirable state for instance a words such as beautiful, wonderful, etc. While negative semantic orientation of a word represents undesirable states such as hate, disgusting, etc.

The studies show that words with polarities, especially adjectives are used as good indicators of subjectivity[27].Hence, the lexicon based approach focuses on words and phrases as the bearers of polarities, and the overall semantic orientation of the whole text is determined by the sum of indicators with polarities. The sentiment words are also called opinion words, which are commonly used to express positive or negative sentiments. For instance, words such as good, beautiful, amazing are positive sentiment words while other words such as horrible, disgusting, bad are negative sentiment words.

2.4.2. Machine learning approach

Machine learning is one part of artificial intelligence, contains many algorithms that aims at examining and evaluating machine behaviors. Machine learning (ML) algorithms deal with different tasks to allow computers to learn. Usually, machine learning algorithms work well on inferring information about the properties of sets of data; features are often captured and analyzed. For the task of sentiment analysis, the success of machine learning also relies on selection and extraction of sentiment features, which are especially from natural language processing (NLP) techniques. In the approach of machine learning, a textual feature representation has been utilized coupled with several algorithms such as Naïve Bayes, Support Vector Machines (SVM), Maximum Entropy, which are commonly used to build the classifiers for sentiment analysis. These classifiers built based on different algorithms can learn the rules or decision criterion of sentiment classification based on training data, then they are used to conduct sentiment analysis automatically[29]. This clearly indicates that the machine learning approach for sentiment analysis is a kind of supervised learning paradigm, where a large number of labeled training data are required to train the classifier before it is used for classifying the new data[5].

In the notion of machine learning, two main activities are performed: the first one is learning the model from a corpus of labeled training data by using algorithms; secondly, classifying the new data based on the trained model.

During the process of the classification task various sub tasks, such as data preprocessing, feature selection, representation, classification and post processing are performed.

In this work, the deep Machine Learning approach will be followed. This approach usually starts with a set of training data. The data should be chosen and categorized properly in order to achieve good prediction results. Then features are chosen to represent the review and train a classifier has been done. Finally the performance of the classifier is evaluated on the testing data.

2.4.2.1. Feature selection

A feature selection is extremely important in machine learning primarily because it serves as a fundamental technique to direct the use of variables to what's most efficient and effective for a given machine learning system. It is the process of selecting a subset of relevant features for use in model construction and usually integrated as the important part of treating the corpus training data in the machine learning approach [30]. It means converting a piece of text into a feature vector or other representation for computational processing [3]. First of all, the training data are labeled as positive, negative or neutral and then a set of features is extracted from the labeled training data. Then the collection of features can be encoded using simple value types, such as Booleans, numbers and strings. For example, the presence or absence of words that appear in the text can be seen as features. Since the training data usually consists of two group data (positive and negative), each word in each group can be seen as a feature vector. Some of words such as Stop Words (for English language “a”, “is”, “the” and for Afaan Oromoo “akka”,”utuu ”,” ega ” ,” koo”,” Kun”,”akkasi”,”osoo”,”moo”) might not provide any sentiment information thus these words are usually filtered out.
The objective of the feature selection is to decrease the dimensionality of the feature space and thus make the computational processing easier. As we observed in several articles a researchers and scholars have tried to exploit different feature types with different algorithms in order to improve the performance of sentiment classification [5], [30], [31].Different combinations of feature selections applied in the machine learning approach could lead to various performances.

For example one of most widely known research to apply the classification approach on document-level sentiment analysis was conducted by Pang and Lee [5]. They used 700 positive and 700 negative pre-tagged documents as training data to build a model with: Naive Bayes, maximum entropy and SVM, respectively. They also discussed the performances under different feature types (unigrams, bigrams) and data forms (frequency or presence), obtaining accuracy rates between 72.8 % and 82.9 %. The best analysis result is obtained by using SVM with the feature of unigrams features than bigrams. This result is indicating that classifier based on SVM algorithm made better prediction with unigrams features than bigrams.

On the divergent, *Dave et al.* [32] conducted an experiment to classify the product reviews from the websites *CNET*8 and *Amazon*9 into two polarities (positive and negative) and found that bigrams and trigrams perform better than unigrams in some setting. *Peng et al.,* [33] presented a simple approach for text categorization by applying n-gram models and reports that the classifier based on the bigram outperforms than its counterpart based on the unigrams. N-gram features offer a simple way to capture the context, if *n* is too small, the model is not able to capture enough context. Conversely, if n is too large (for n > 3), it is computationally very difficult to generate the features directly and also it will create severe sparse data problems[33], [34].

2.4.2.2. Algorithms

This step is start after a feature is extracted from the training data; an algorithm is applied to learn those features. If a particular feature tends to be true consistently when the training data belongs to a certain group (neutral, negative, positive), the classifier based on this algorithm will learn that feature is a good indicator of that group. For instance, if the word 'good' appears more often in known positive training data, the algorithm will learn the feature word "good" is an effective indicator of positive orientation. When the classifier is used to test the new data, it will derive values according to the features based on the new data and multiply those values by the weights learned from the training data. The sum of the value will show the results of the sentiment classification[3], [32],[5]. For example, if the word "good" always appeared in positive training data set, whereas the word 'bad' showed up more often in the negative training data set during the training, the algorithm gives a +1 weight on the feature word 'good' and a -1 weight on "bad".

According to Brooke [36],if a new text that contains the word 'good' and 'bad' is fed to the classifier, the number of appearances of the word 'good' (say 3 times) and 'bad' (say 5 times) will be calculated and summed up (score: 3-5 = -2), which is indicate as the result is negative.

The machine learning approach for sentiment analysis has been epitomized by Pang *et al.* [5]. Since it is not obvious to see which particular algorithm can make better prediction, Pang *et al.* [5]compared Naïve Byaes, Support Vector Machines (SVM), and maximum-entropy-based classifier to classify movie review into two classes: positive and negative. Neutral reviews were not used in this study, which made the classification task easier. They report 81.0%, 82.9% and 80.4% accuracy for Naïve Byaes, SVM and Maximum

Entropy algorithm respectively. They also indicate that either Naïve Bayesian or SVM algorithm performed well with unigram features.

2.4.3. Artificial Neuron

Artificial Neural Network (ANN) is widely used in the field of machine learning to solve problems such as predictive modeling, classification and function approximation. An ANN is a network of nodes connected to each other. The nodes of the network, also called Artificial Neurons (AN) or perceptrons, are the basic building blocks of ANNs. The neurons in a network are connected by weights and the structure of a simple AN can be seen in Figure 1. An AN essentially takes an input, x, associated with a weight, w, plus a bias term, b, calculate the *net* input which usually is a weighted sum, (Equation 2.1), apply an activation function that decides whether the neuron should "fire" or not. The role of the weights associated with the input is to either strengthen or weaken the input signal. The operations of an AN can be seen as a nonlinear mapping from R^N to usually either an output in the range of 0 and 1 or -1 and 1 where N defines the number of inputs. The choice of activation function determines to what interval the input signal will be mapped and if the neuron should fire. The bias term allows for shifting the activation function to either right or left along the x-axis and is also associated with a weight and an additional biasvalue, see Equation 2.2.[15]

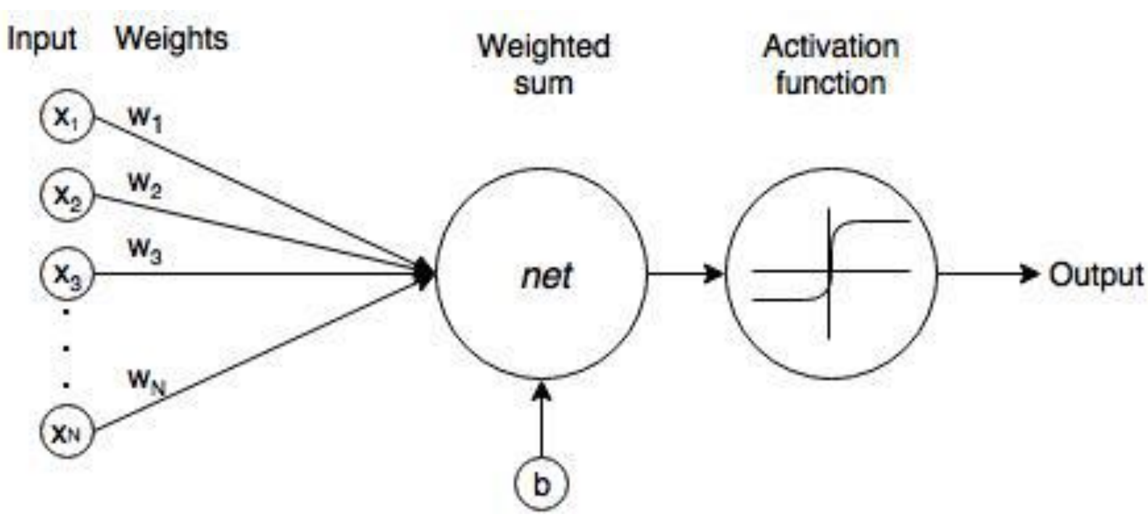


Figure 3: Artificial Neuron

$net = \sum_{i=1}^{N}(XiWi) + b$, where b is as in Equation 2.2   2.1.

b = w_0 *biasvalue, where biasvalue usually is equal to 1   2.2.

Activation function

Activation function is an important feature of neural network. Neuron should be activated or not is decided by activation function. It calculates the weighted sum of inputs and add bias. It's a non-linear transformation of input value. After transformation, this output is sent to next layer. There are various types of activation functions, also called transfer functions; those are Step function, sigmoid function, tanh function or rectifier (ReLU) function are all common activation functions[15].

Step function is a threshold based activation function. Simple, if value of X is greater than a certain value mark it activated otherwise not.

Activation function Y = "not" if X < threshold else "activated"

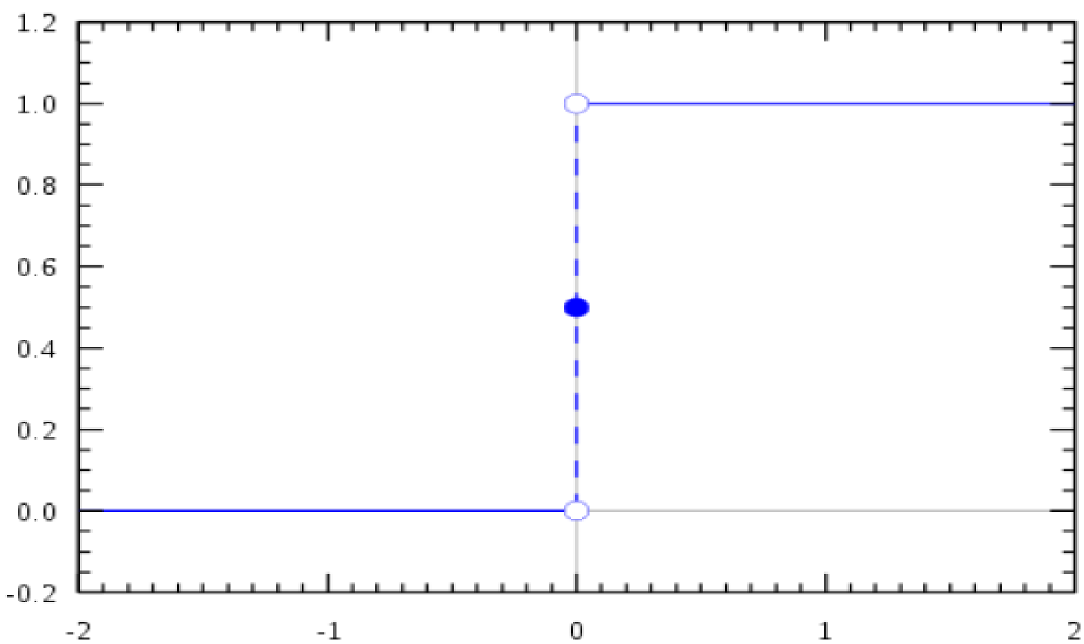


Figure 4: Step function calculation based on threshold

As shown in Figure 3 output is 1 when value is greater than 0 (zero is the threshold value) and outputs is 0 if less than threshold.

A plot of the sigmoid activation function where $\lambda = 1$ can be seen in Figure 2 and its corresponding mathematical expression in Equation 2.1. The activation function maps the net input value to a value in the range of [0, 1]. The λ parameter controls the steepness of the function. In a binary classification problem the output value can be seen as the probability that the input belongs to a certain class. Say that the activation function outputs the value 0.8, this indicate that there is 80% chance that the input belongs to class 1 and 20% chance that it belongs to class 2. A similar activation function used for multi-classification task is the softmax function which will output a probability for every class in the range [0,1] where the sum of the probabilities are equal to 1.

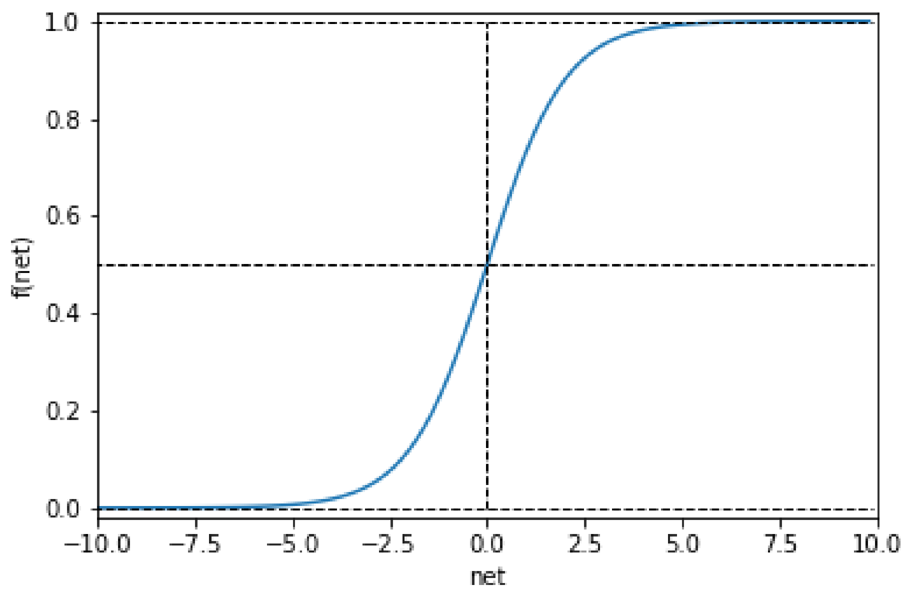


Figure 5: Sigmoid activation function

$$f(net) = \frac{1}{1+e^{-\lambda\, net}} \tag{2.3}$$

The ReLU activation function, Figure 3 and mathematically expressed in Equation 2.2, maps the net input to a value in the interval [0,∞]. This means that all negative values are set to zero. By setting negative values to zero and hence make the neuron inactive will ease the computational load in a network with many neurons since fewer neurons are activated and hence less computations are required. Also computations are linear for ReLU compared to the exponential sigmoid function. It has also been shown in [15]that ReLU activation function is well suited for not only image recognition but also for sentiment analysis with sparse text data.

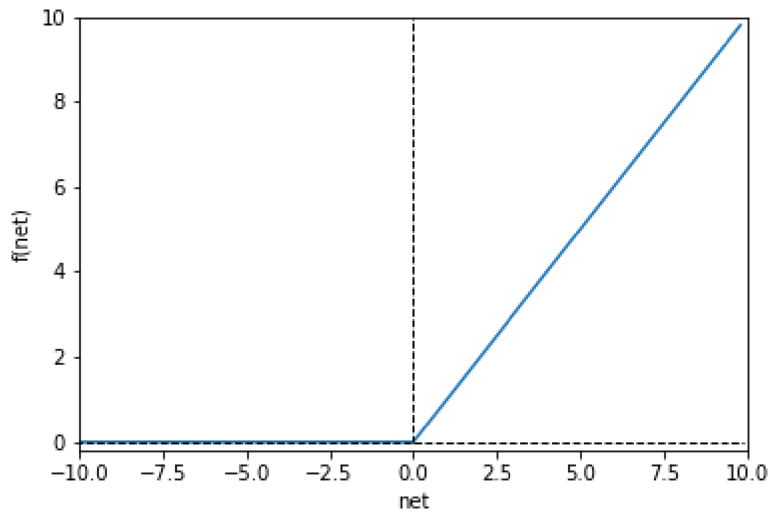


Figure 6: ReLU activation function

$$f\,(\text{net}) = \max(0,\text{net}) \tag{2.4}$$

Neural Network

In terms of computer science, a neural network is an artificial nervous system for receiving, processing and transmitting the information. Collection of neurons that connected together is called a neural network.

Neural networks have shown great promise in several NLP applications, including sentiment analysis[37]. A full overview of neural networks and some family of neural network such as RNN, LSTM, Recursive Neural Networks and others are beyond the scope of this thesis. Therefore, we will focus only on models used in our experiments. Like all machine learning, neural networks also require large amounts of training data in order to generalize well. However, since they have a larger number of free parameters, they are more prone to overfitting than linear models such as SVMs or maximum entropy models. Regularization and early stopping are often used to counteract this problem.

Feed forward neural networks: -

Feed forward neural networks (FF) are the first and simplest kind of neural network. As explained in [15] regular feed forward neural network has at least three layers; an input, hidden and output layer where every layer consist of a few or several neurons.

Information flows from the input layer, through an optional number of hidden layers, finally coming to the output layer, which has a dimensionality of the number of classes k that we want to classify. An example of a feed forward neural network with two hidden layers, also called a Multi-Layer Perceptron (MLP) [15]. An MLP receives an input signal and passes that signal through the network, layer by layer to finally reach the output layer. The output of a layer is the input of the subsequent layer. In supervised learning, the goal is to train a machine learning model so that it can classify or predict unseen and unlabeled data by letting the model train on known labeled data.

A general figure of a deep Feedforward Neural Networks is shown in Figure 5.

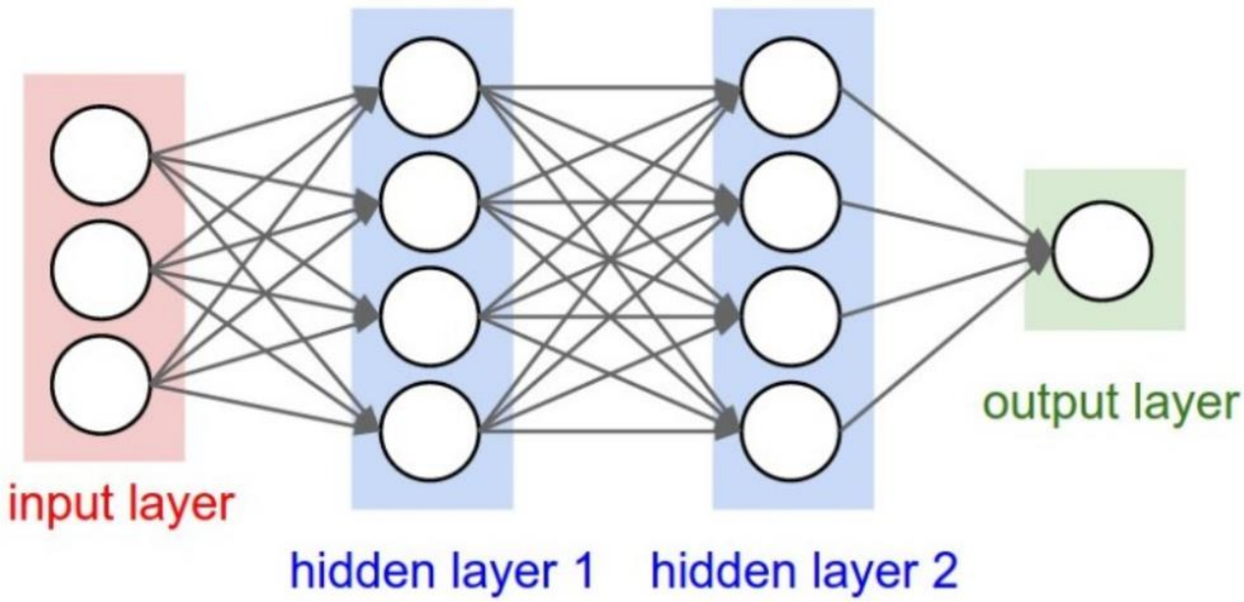


Figure 7: Feedforward Neural Networks

Neural networks can be used for supervised learning where the model, during training, is given both an input as well as a label/target associated with that input. The network then tries to minimize the error between the networks predicted output value and the target value by adjusting its weights. The weight adjustments of the neurons are made by utilizing optimization and a method called backpropagation. Once a network is trained it can be saved and used for later hence there is no need to retrain the network with the same data every time it shall be used. What saving a trained network actually means is that the architecture as well as the trained weights are saved then either the entire network or just desired parts, i.e. layers or specific neurons of the network can be used for other occasions.

Cost function

Measuring how well a machine learning or statistical model predict the outcome of an event can be done using a cost or objective function, sometimes also called a loss function[15]. Essentially what the cost function is measuring is the error rate between the predicted value and the correct target value. The goal of the machine learning model is to obtain the smallest possible error, i.e. to minimize the error and hence the cost function. There are several types of cost functions. A common function is the Mean-Squared Error (MSE) that is used to calculate the average squared difference between the predicted values, p, and the target values, t, see Equation 2.3. Cross Entropy (CE) (Equation 2.4), or binary cross-entropy in binary classification problems, is another common objective function that has proven to converge faster as well as yield better outcome in classification error rates compared to MSE for ANNs[15].

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (pi - ti)^2 \qquad 2.5$$

$$\text{CE} = \frac{1}{N} \sum_{i=1}^{N} ti \; log \; (pi) \qquad 2.6$$

Where p is the predicted value, t is the correct target value and N number of samples.

2.4.4. Deep learning

Deep learning is a subfield of machine learning that also allows one to address text classification. The concept originates from the idea that machines can store and process information in a similar way to the brain. Information is processed by many simple computation units. Many computational units in sequence can process complex information. The deep word means that a number of layers in the network. So having more layers lead to a better and deeper network. Supervised, semi-supervised or unsupervised learning can perform on different classification tasks such as images, text, or sound.

Recently, Convolutional Neural Networks have been improved the state-of-the-art in text polarity classification by a significant increase in accuracy [38][39][40]. The main idea of these methods is to use supervised training on a network with convolutional filters, acting as a sliding window over the sequence of given words, followed by maxpooling.

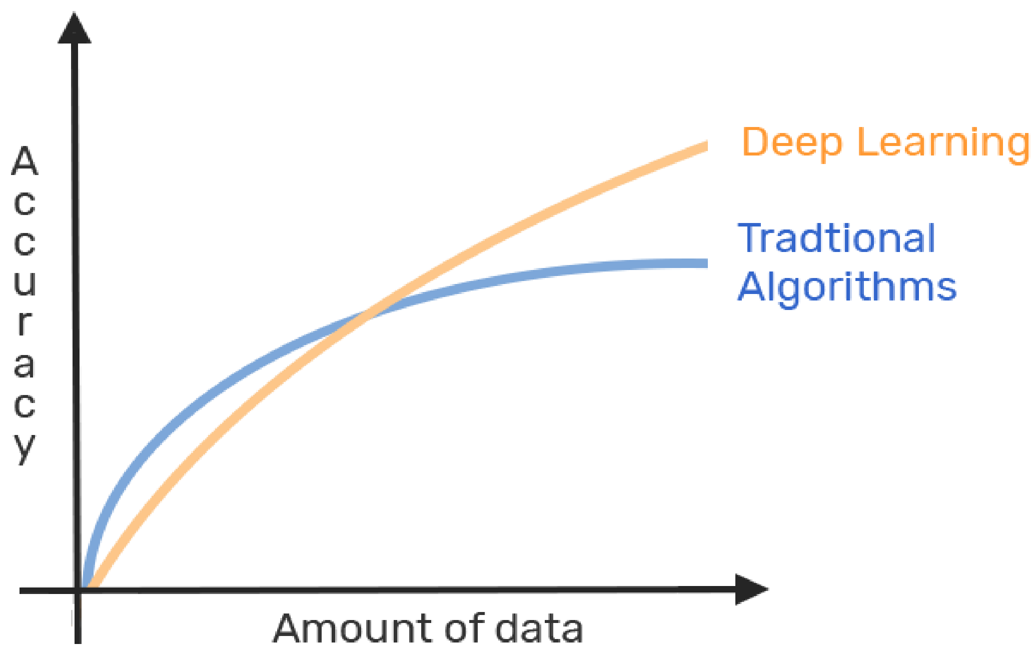
Typically, common word embedding techniques are used to represent each input word. Such neural network architectures share a lot of similarity with CNN methods used in computer vision, which have achieved excellent results in several tasks, e.g. object detection and image segmentation[41].

- **Comparison of machine learning and deep learning**

This section presents some important points of deep learning and machine learning then compare the two techniques.

- **Data dependencies**

When it comes for data dependencies the Machine Learning can perform well even with a small size of data while Deep Learning doesn't perform that well with small set of data. On the one hand, deep learning algorithms require much more training data than traditional machine learning algorithms. On the other hand, traditional machine learning algorithms such as SVM and NB reach a certain threshold where adding more training data doesn't improve their accuracy.



This is because deep learning algorithms need a large amount of data to understand it perfectly or in order to make a brief conclusion.

❖ Hardware dependencies

Regarding the hardware dependencies, deep learning algorithms heavily depend on high-end machines, contrary to traditional machine learning algorithms, which can work on low-end machines. This is because the requirements of deep learning algorithm include GPUs which are an integral part of its working. Deep learning algorithms inherently do a large amount of matrix multiplication operations. These operations can be efficiently optimized using a GPU because GPU is built for this purpose[42].

❖ *Feature* engineering

Feature engineering is a process of putting domain knowledge into the creation of feature extractors to reduce the complexity of the data and make patterns more visible to learning algorithms to work. This process is difficult and expensive in terms of time and expertise. In Machine learning, most of the applied features need to be identified by an expert and then hand-coded as per the domain and data type. The performance of most of the Machine Learning algorithm depends on how accurately the features are identified and extracted.

Deep learning algorithms try to learn high-level features from data. This is a very distinctive part of deep learning and a major step ahead of traditional Machine Learning. Therefore, deep learning reduces the task of developing new feature extractor for every problem.

Consequently, in the feature engineering, deep leaning outperformed Machine learning since Deep learning has the ability to make the feature itself while the Machine Learning requires the user to identify the exactitude of the feature.

❖ Problem solving approach

The problem-solving method of deep learning is better than Machine Learning one because Deep Learning the problem are solved end to end while Machine Learning solve the problem by splitting into small pieces then solve one by one and get a result by combining all of them[42].

- **Execution time**

Regarding the execution time, usually a deep learning algorithm takes a long time to train. This is because there are so many parameters in a deep learning algorithm that training them takes longer than usual. Whereas machine learning comparatively takes much less time to train, ranging from a few seconds to a few hours, but this is completely reversed on testing time. At test time, deep learning algorithm doesn‘t take much time compare to machine learning to run[42].

2.5. Word Embeddings

In order to use a neural network in NLP task converting words to vectors are carried out as initial step. The task of converting words or sentences in vectors is called pre-processing. Among several approaches, word embedding is a modern approach for representing text while working on natural language processing with neural networks algorithm. It is capable of capturing context of a word in a document, semantic and syntactic similarity, and relation with other words**.** In other word embeddings helps to understand how dissimilar words are related based on their context. It is techniques use to represent every word in the vocabulary as a real-valued vector in a predefined vector dimension, usually hundreds of dimensions. In this approach, every phrase gets map to individual vector and the vector values learn the way to look like a neural network. The vectors are learned in such a way that words that have similar meanings will have similar representation in the vector space or close in the vector space, which is indicate that Word embeddings maintain Semantic Similarity.

For example, the function that transforms *Man* into *Woman* will return *Queen* if we apply it to *King*.

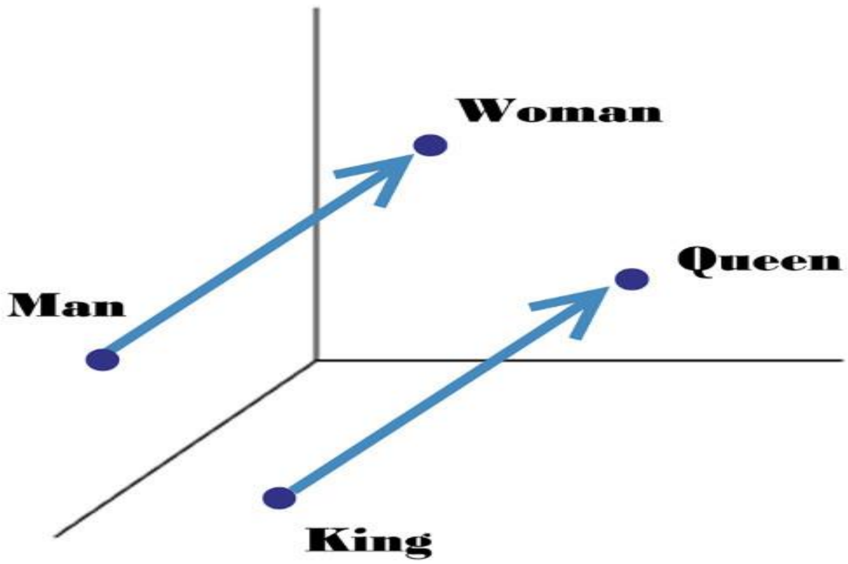


Figure 8: word2vec

This is a more expressive representation for text than more classical methods like bag-of-words, where relationships between words or tokens are ignored. Generally, there are three techniques used to learn a word embedding from corpora of text those are embedding layer which will be discussed in chapter four but, word2vec and gloves, discussed in the following section.

2.5. Word2vec

The word2vec is a collection of similar models that are used to make word embeddings. These models are simple, neural networks with two-layer that are used to train and build linguistic contexts of phrases. Word2vec gets an input of large corpora of text and creates a vector space, generally hundreds of dimensions, where each individual word in the corpora is being designated to a corresponding vector space. Those word vectors are placed in the vector space so that words with similar contexts in the corpus are placed close one to another in the space. It was developed at Google by Tomas Mikolov and his team. The algorithm developed was consequently analyzed and many other researchers have explained it too. The technique used two different learning approach to learn the word embedding: CBOW model and Skip-Gram Model for distributed representations of words whose aim is to minimize computational complexity. The Continuous Bag-of-Words is a model that predicts the actual word based on its content while learning the embedding. The CBOW uses the average value of the word embedding of the context to predict the current word.

And the skip-gram is similar to CBOW, but instead of predicting the word from context, it tries to maximize the classification of a word based on another word in the same sentence. It is model that predicts the words around for a given actual word. The following figure 10 shows Word2Vec Training Models [43] .

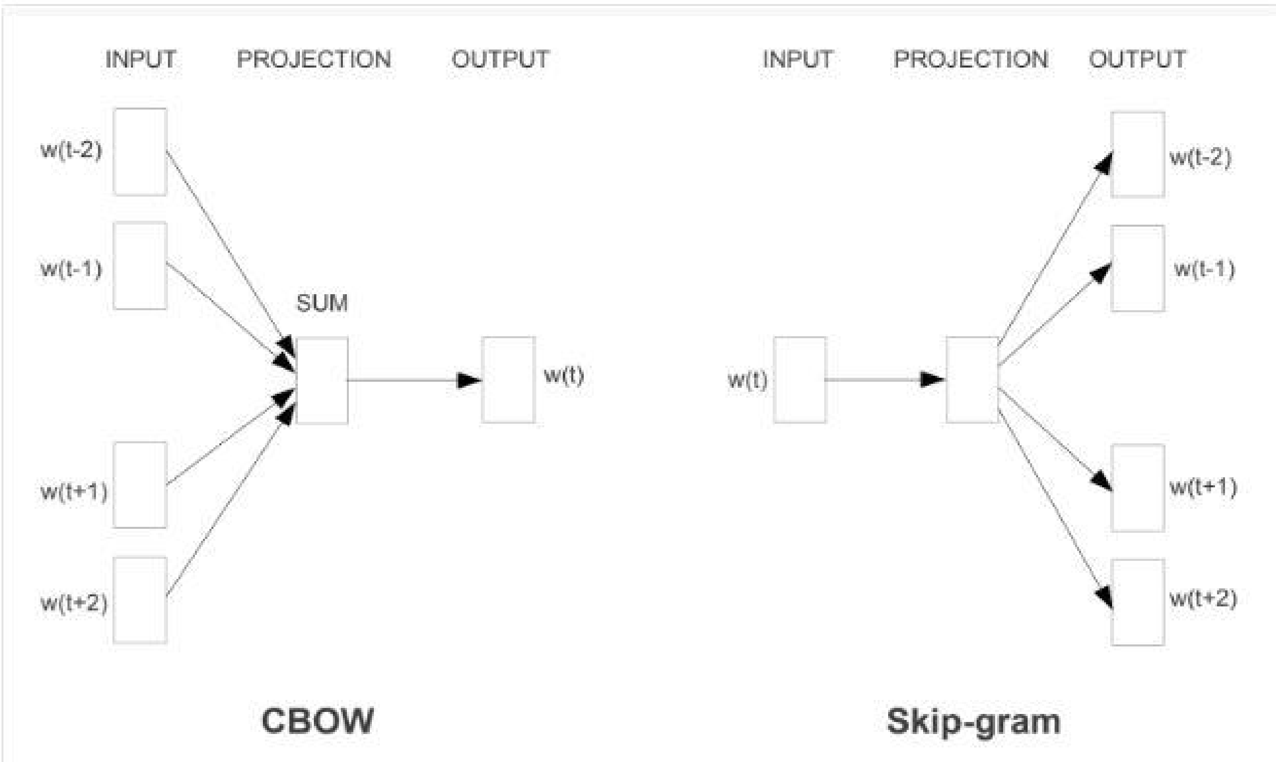


Figure 9: Shows Word2Vec Training Models

Gensim provides a word2vec class to use word2vec model[44]. It just involves organizing and loading text into sentence and pass those sentences to the constructor of word2vec() instance. This constructor has many parameters but there are few which will need to be configure according to defined problems. To say few; Size one which default value is 100, it's the dimension of embedding's. The other Sg and argument that is important for training algorithm in order to use *CBOW* (0) or *skip-gram* (1) and default value is 0. Also Min-count an argument which decided its minimum number of words to consider while training model also has default value 5. The other argument is window which is the maximum distance words around the target word and target word and used 5 as default value. The main advantage of using this approach is that word embeddings learn effectively in high-quality word and allow great size of embeddings to be learned from much bigger corpus of text[43].

2.5.1. Gloves

Global Vectors for Word Representation or GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space [43].It was created by team of researchers led by Pennington at Stanford in 2014. It is a methods that put the global statistics of matrix factorization methods such as Latent Semantic Analysis with the content-based learning in word2vec. It is learning of word representations that outperforms other models on word analogy, word similarity, and named entity recognition tasks[43].

2.6. Afaan Oromoo Language

Afaan Oromoo is one of the most widely spoken and used language in Ethiopia ;by more than 34.4% of the total population of the country [9]. It has also widely used as both written and spoken language some neighboring countries like Kenya, Egypt and Somalia[10]. Afaan Oromoo belongs to an East Cushitic language family of the Afro-Asiatic language super family. It is the second most widely spoken indigenous language in Africa next to Hausa in Nigeria[45].

Currently, it is an official language of Oromia regional state, which is the largest regional state among the current Federal States in Ethiopia. It has also employed as a medium of instruction in in primary schools, elementary schools and teacher training college of Oromia regional state. Similarly, taught as an examinable subject called Barnoota Afaan Oromoo in high school and offered at some universities in Ethiopia as a Minor and Major subject of study at both undergraduate and postgraduate level.Moreover, a number and variety of books published in Oromoo using the *Qubee*. Besides school student textbooks, a lot of grammar books, dictionaries, novels, short stories, poetry and plays have been written in Oromoo during the last decades. Additionally, a number of social media and websites are rapidly increasing to broadcast news and other events using Afaan Oromoo and these have provided greater opportunities for Oromoo people to speak, read and write in their language.

2.6.1. Afaan Oromo Alphabet and Writing System

The writing system of Afaan Oromoo language is designed based on the Latin script which called Qubee in Afaan Oromoo; which has been accepted and became the official scripting system of the language since 1991[45].This writing system was adopted from the fact that its characters explicitly represent the vowels and the consonants of the language. The Qubee writing system of Afaan Oromoo has a total of thirty-three letters that consists of all the twenty-six English letters (a…z) and the seven combined consonant letters (ch, dh, sh, ny, ph, ts, zh). The combined consonant letters are known as „**qubee dachaa'**. Afaan Oromo alphabet is characterized by capital and small letters like English alphabet. All the vowels in English (a, e, i, o and u) are also vowels in Qubee Afaan Oromoo. It has five short and five long vowels, which is lead to look the results or meanings from two different perspectives. For instance: *Bona (summer) and laga (river)* are short vowels, whereas *laagaa (throat) and Boonaa (pride)* are long vowels [10]. Doubling of a consonant is a phonemic in Afaan Oromoo. For Example: *Callaa (product), Damma (honey), Ganna (winter)*. But "*h*" character is not geminated at all [10].

2.6.1. Punctuation marks in Afaan Oromoo

Punctuation is placed in text to make meaning clear and reading easier. As Wegderes stated that, analysis of Afaan Oromo texts reveals that different punctuation marks follow the same punctuation pattern used in English and other languages that follow Latin Writing System [13]. Like English, the following are some of the most usually used punctuation marks in Afaan Oromoo language [13].
☐ "***Tuqaa***", Full stop (.): Like English full stop is used at the end of a sentence and also in abbreviations.

☐ *"**Mallattoo Gaafii**",* Question mark (?): is used in interrogative or at the end of a direct question.

☐ ***“Rajeffannoo”***, Exclamation mark (!): is used at the end of command and exclamatory sentences. ***“Qooduu”,*** Comma (,): is used to separate listing in a sentence or to separate the elements in a series. ***“Tuqlamee colon”,*** Colon (:): the function of the colon is to separate and introduce lists, clauses, and quotations, along with several conventional uses, and etc. Unlike English language apostrophe (***’)*** is not punctuation mark in Afaan Oromoo, rather it is part of words. For example, *har’a* (today), *re’ee* (goat) etc.

2.6.2. Afaan Oromoo Word Categories

Language is made up of words and a series of rules that connect words together[45]. A word is the basic stuff of language. The arrangement of word depends on the grammar of that language. Putting words together based on the rules of the language is yield sentence; the meanings of these sentences depend on each word of the sentence. All words do not have equal contributions to sentence meaning[45]. Their contribution depends on their category and their feature. Some of the linguistics experts are classified grammatical word categories of Afaan Oromoo into eight classes, afterward a series of improvement in terms of its word categories and other syntactic features. Hence, the word categories of Afaan Oromoo words are reduced from eight (Noun, Verb, Adjective, Adverb, Adposition, Pronoun, Conjunction and Interjection) to five (nouns, verbs, adverbs, adjectives and adposition) grammatical categories whereas pronouns included under the noun category, conjunctions and interjections under adposition [46].

2.7. Afaan Oromoo sentiment analyses challenge.

There are several challenges that restrict automated ways of understanding, analyzing and classifying Afaan Oromoo text into different polarities. The major challenge of Afaan Oromo language in sentiment analysis is lack of available resource such as preparation lexicon database, part of speech tagging used to identify the noun, noun phrase, adjective, verb and adverb, used to detect aspect of the opinions and polarity of opinions. Similarly, there is no ready-made dataset /corpus available for Afaan Oromoo sentiment analysis in order to conduct or investigate different machine learning or deep learning experiments. The other challenge is that users express their fleeing in context based or/and indirect manner which makes the task of identify the polarity of the comments more difficult.

CHAPTER THREE: RELATED WORK

3.1. Introduction

This chapter briefly discusses some of the sentimental analysis related works for some languages. In addition, the three mostly employed approaches for sentiment analysis i'e, Lexicon, 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 Oromoo[10] ,Amharic [47], French[48], Turkish[49].

3.1. Sentiment Analysis using Rule Based Approach

A lot studies done on sentiment analysis for local language and international language using lexicon and rule based approach. We also reviews some of the thesis and article that has been on both local and international language using this approach as follows.

3.1.1. Sentiment Mining for Afaan Oromoo Texts

Eshetu [10] proposed a sentiment analysis model for opinionated Afaan Oromoo texts. The author's attempted to identify and categorize sentiments expressed in a piece of texts into positive, strongly positive, weakly positive, negative, strongly negative, weakly negative or neutrals. In order to realize the objective of study the author used manually constructed rules and subjectivity lexicon of the language. The model detects subjectivity words of a review from the developed lexicon and assigns an initial polarity weight for each sensed sentiment terms in order to determine the polarity classification of the opinionated Afaan Oromoo texts.

The author conducted an experiment on three different datasets manually collected from peoples of Addis Ababa house renters, A.A Anbesa bus users and English to Afaan Oromoo online dictionary. Precision, recall and F-measure parameters were used for evaluating the effectiveness of the developed prototype. For house rental reviews a precision, recall and F-Measure of 0.670, 0.744 and 0.699 respectively were achieved. For A.A Anbesa bus reviews a precision, recall and F-Measure of 0.745, 0.709 and 0.718 respectively were achieved. For English to AO online dictionary reviews a precision, recall and F-Measure of 0.678, 0.890 and 0.716 respectively were achieved.

Another studies also done by Wegderes Tariku [13],the researcher used rule based approach to detect the polarity of opinion and summarization. The researcher develop sentiment mining and aspect based opinion summarization of service review in Afaan Oromoo language for Oromia Radio and Television Organization (ORTO). They collected 300 reviews from Oromia Radio and Television Organization in news domain. Their experimental result indicated that for positive class precision of 90% and recall of 87% whereas for negative class precision 86.6%and 89.6% recall is achieved.

We identified the limitation of both paper is need performance improvement. The data used for those experiment not sufficient in order to conduct deep learning experiment hence collecting and annotating more data is very important since it somehow fulfill the resource gap in sentiment classification area.

3.1.2. Sentiment Mining for Amharic language

Tulu [47], employed manually crafted rules and lexicon approach to develop feature level opinion mining and summarization model for Amharic language. The author deal with a variety of components required for constructing the model such as text operator, morphological analyzer, feature extractor, opinion extractor, and feature-Opinion Summarization. The author used Michael Gasser"s HornMorpho 2.2 to analyze the reviews, then sends analyzed text to feature extractor component in order to extracts the nouns as features.

The author has been conducted two experiments for features extraction and opinion words determination by using 484 reviews from three different domains. The first experiment indicated that an average precision of 95.2% and recall of 26.1% were achieved in the features extraction and an average precision of 78.1% and recall of 66.8% were achieved in the determination of opinion words. The precision of the second experiment in features extraction gets lower by 15.4% whereas the precision of opinion words determination gets higher by 1.9% and the recall of both features extraction and opinion words determination gets higher by 7.8% and 25.9% respectively when compared to the first experiment. The author stated that experimental results demonstrate the effectiveness of the techniques they applied.

Wondwossen[50] present a multi-scale sentiment analysis model for Amharic using supervised machine learning approach. Author developed their own corpus by collecting around six hundred posts from online sources. Author have been employed preprocessing techniques to clean the data, to convert transliterations to the native Ethiopic script and to change words to their base form by removing the inflectional morphemes. After preprocessing, the corpus is manually annotated by giving polarity and sentiment intensity scale values. Author employed Naïve Bayes machine learning algorithm and used unigram, bigram and hybrid variants as features. The experiment results show that, among the three learning setups, the accuracy of the bigram approach is found to be promising. Generally, author reported as the results are encouraging despite the morphological challenge in Amharic, the data cleanness and small size of data. The also convinced that the results could improve further with a larger corpu.Selama G.Meskel [51] also proposed sentiment mining model for determining the sentiments expressed in an opinionated Amharic reviews. The polarity classification of the opinionated texts was positive, negative or neutral. The proposed model detects positive and negative sentiment terms including contextual valence shifters such as negations and assigns an initial polarity weight to all detected sentiment terms in order to determine the polarity classification of the opinionated text. The author was test the model using movie and newspaper reviews total 254 movie reviews and 49 reviews newspaper collected manually. Author reported that, the result obtained with these test data was very much encouraging

Even though the proposed models are address the identified problems and shows the promising result, but impossible to apply for Afaan Oromoo reviews because of the text granularity and nature the language.

3.1.3. Sentiment Analysis for Tigrigna Language

Weldu [52],proposed a sentiment analysis model for opinionated Tigrigna texts. The major focus of the research is sentiment analysis of Facebook activists opinionated of Tigrigna political reviews. The researcher was used manually constructed rules and subjectivity lexicon of the language. The developed prototype detects subjectivity words of a review from the developed lexicon and assigns an initial polarity weight for each sensed sentiment terms in order to determine the polarity classification of the opinionated Tigrigna texts.

Researchers conducted an experiment on political domain reviews or datasets of 582 political reviews. The achieved result in the experiment was 0.905, 0.801, and 0.848 of precision, recall and f-measure respectively.

3.1.4. Sentiment Analysis of Chinese Language

Zhang et al. [28] proposed a sentiment analysis system for Chinese depending on rule based system with no-annotation cost for Chinese articles in multiple domains. Their approach is based on using the sentiment lexicon and the syntactic structure of each sentence. Their method consists of two main steps: The first step is computing the sentiment of the sentences. The second step is aggregating the sentiment of the sentences to get the score of the sentiment of the document sentiment. The objective sentence is excluded by scanning the document for subjectivity sentences only using the occurrence of the subjective words. To compute the polarity of each sentence, the researchers depend on computing the modified polarity of the words.

3.2. Sentiment Analysis using Machine Learning Approach

We will see some of the studies performed using classical machine Learning and deep learning as follows:

3.2.1. Sentiment analysis of French movie reviews

Ghorbel and Jacot[48] classifying the polarity of conveyed opinions from French movie reviews. A supervised learning approach is used to train the classifier on annotated data of French movie reviews extracted from the web. Unigrams feature served as the baseline in experiments. Also further linguistic features such as lemmatized unigrams, POS tags and semantic orientation of selected POS tags was included in order to improve the results. The authors were addressing the problem of loss of precision in defining the semantic orientation of word unigrams from English lexical resources. Lexical features with composed of word unigrams are served as baseline of their experiments. Each unigram feature formulates a binary value indicating the presence or the absence of the corresponding word at the review level. In order to improve the relevance of unigram features, authors propose certain further variants. The authors are point out, like other language also French language contains a lot of stop words (”de”, ”du”, ”`a”, ”le” and ”la”). Since stop words don’t hold polarity information so they aren’t relevant for the classification, so that removing stop words improve the results. Another assumption is grouping all inflected forms of a word in a single term are another useful technique in sentiment analysis. For example, consider the words ”aim´e”, ”aimait”, ”aimer”, ”aiment” and ”aime”, all these words share the same polarity but will be considered as five separate features during the classification. The authors stated that applying lemmatization helps to obtain a unique feature and restricting features to very important Part of Speech would improve performance.

For this experiment, authors collected data from web. They extracted a corpus of 2000 French movie reviews of 1000 positive and 1000 negative from 10 movies. 1600 were used for training, 400 for testing and Support Vector Machine (SVM) classification method applied to train and classifies French movie reviews. Different features have been incorporated and improved the accuracy and performance of the proposed system. So generally Features reduction will improve the tuning of the training process and applying POS is also important to aid word sense disambiguation.

3.2.2. Urdu Sentiment Analysis

Mukhtar and Abid Khan[53] studied sentiment analysis of Urdu language reviews using machine learning approach. Data was collected from various online blogs; totally 6025 sentences are collected from 151 blogs. These sentences are annotated by two human Urdu annotators; each annotated all the 6025 sentences. A sentence is labeled in case of agreement of same label by the two annotators. In case of disagreement, the decision of the third annotator was used to annotate the sentence. Urdu stop words play important role in sentence completeness. They are, not removed earlier in the annotation phase as the annotators may face problems in predicting the correct classification of the sentence due to incomplete sentence. Support Vector Machine, Decision tree and k-Nearest Neighbor (k-NN) are tested, their outputs are compared and their results are ultimately improved in several iterations after taking a number of steps that include stop words removal, feature extraction, identification and extraction of important features. The authors concluded that k-NN is performing better than Support Vector Machine and Decision tree in terms of accuracy, precision, recall and f-measure.

3.2.3. Sentiment Analysis for Arabic

This[7] paper deals with sentiment analysis in Arabic reviews from a machine learning perspective. The generated the dataset by collecting tweets and Facebook comments from the internet. These tweets and comments address general topics such as education, sports, and political news. Three classifiers were applied on an in-house developed dataset of tweets/comments. The Naïve Bayes, SVM and K-Nearest Neighbor classifiers were run on this dataset. The results show that SVM gives the highest precision while KNN (K=10) gives the highest Recall.

3.2.4. Sentiment Analysis for English Language

In publications, there are many works interested in SA of English Language. Li and Jain[54] used four methods for document classification: decision trees classifier, naive Bayes classifier, nearest neighbor classifier and subspace method. Yahoo news data set was used for their studies. Features were represented using standard bag of words (BOW). They indicated that naive Bayes classifier and the subspace method give better results than decision trees and nearest neighbor classifiers on the same data sets.

Pang et al.[5], used ML methods such as ME, NB, and SVM to define the polarity of movie reviews. They downloaded the data from Internet Movie Database (IMDb). The reviews were divided to 700 positive and the other 700 reviews are negative. They applied ML classifiers and standard bag of features (BOW) on documents. The position of word and the part of speech were also treated. They performed many experiments using n-grams approaches, and according to the results they concluded that using of unigrams is the most effective method, and the SVM classifier outperforms the other classifiers used in their work.

Turney and Peter[22] detected document sentiment based on adjectives or adverbs phrases by using proposed an unsupervised learning algorithm. Then, they computed the semantic orientation by using Point Mutual Information (PMI). They assigned a class of "recommended" or "not recommended" to the sentences based on the average semantic orientation of the phrases. Finally, the achieved average of accuracy is 74% when the data were taken from different domains (410 reviews of banks, films, automobiles and travel destinations). While the accuracy is 84% for automobile reviews and 66% for film reviews.

3.3. Sentiment Analysis Using a Deep Learning approach

In the follows we will some of studies employed deep learning on various natural language:

3.3.1. Sentiment Analysis in Spanish

Paredes-valverde.*et.al* [55], proposed a deep-learning-based approach that allows companies and organizations to detect opportunities for improving the quality of their products or services through sentiment analysis. They used convolutional neural network (CNN) and word2vec. They was classified tweets in order to determine the effectiveness of their study. The proposed sentiment classification in this work is has three main modules: preprocessing module, word embeddings, and CNN model. The researcher's performed firstly, tokenization and normalization of the text. And then used word2vec to obtain the feature vectors. At the last step trained a convolutional neural network to classify tweets as positives or negatives. The researchers was conducted experiments with different sizes of a Twitter corpus composed of 100000 tweets. And obtained encouraging results with a precision of 88.7%, a recall of 88.7%, and a F-measure of 88.7% considering the complete dataset.

3.3.2. Sentiment Analysis of Swedish reviews

Kristoffer Svensson[56], conducted study on Swedish reviews using deep learning and neural networks. Several convolutional neural network models were implemented with different settings to find a combination of settings that gave the highest accuracy on the given test dataset. There were two different kind of models, one kind classifying positive and negative, and the second classified the previous two categories but also neutral. The training dataset and the test dataset contained data from two recommendation sites, www.reco.se and se.trustpilot.com.The final result shows that when classifying three categories (positive, negative and neutral) the models had problems to reach an accuracy at 85%, were only one model reached 80% accuracy as best on the test dataset. However, when only classifying two categories (positive and negative) the models showed very good results and reached almost 95% accuracy for every model.

3.3.3. Sentiment Analysis for Bangla Sentences

Alam [57], propose a framework that analyzes sentiments from texts written in Bangla. Researcher use Bangla comments and generate a classification model. The model is generated by a neural network variance called Convolutional Neural Network. The proposed system consists of two processes, preprocessor and model generator. The researcher is collect about 850 Bangla comments from different sources and process them to make them suitable to train the network. Among the comments, 500 comments are positive and the rests are negative. The researcher copy-paste the comments repeatedly to increase the size, because the amount of comments is not sufficient to train the proposed network.
The classifier model obtains a classification accuracy of 99.87%, which is 6.87% better than the available state-of-the art Bangla sentiment classifier.

3.3.4. Sentiment Analysis for French language

Joel Kanku[43] present, a deep learning approach for sentimental analysis using French text. His is based on Convolutional Neural Networks for text classification. The goal of the paper is to use convolutional neural network architecture to classify the sentiment expressed in a given French text about services and products. The model build get a brief text of restaurant and laptop reviews as input and determines the sentiment polarity in each sentence as output. To determine the effectiveness of this approach for classifying French text, they conducted experiments in two-channels: CNN-rand and CNN-static. They obtained encouraging results for CNN-rand with 75.44% for restaurant domain and 74.64% for laptop domain. And with CNN-static they got 82.46% for restaurants and 81.16% for laptops.

3.3.5. Sentiment analysis for Finnish Language

Nukarinen[58], proposed automatic sentiment analysis of product reviews written in Finnish language. Done using the latest deep learning techniques in the field of machine learning. The researcher is aimed to answer the question "Is it feasible to do sentiment analysis for Finnish language automatically?". They collected product review from website called www.verkkokauppa.com using API of the shop. Simple pre-processing consisting of lemmatization, fixing the size of the vocabulary, and padding were applied. The actual implementation of the neural network was done using Keras. The researcher was concluded that automated sentiment analysis for Finnish language is feasible.

3.3.6. Sentiment analysis of Amharic: A deep Learning Approach

Yeshiwas and Abebe[59], has analyzed some Facebook postings to understand socio-political sentiments of Amharic language using deep learning approach. The researcher collected the data from Facebook social media. After collecting thus data from the FBC all preprocessing steps tokenization, stop word removal, stemming of the sentence was undertaken. To evaluate the performances of the systems they was collected 1600 reviews from immigration, war and public relation domains. Then evaluate the system training and validation accuracy using three experiments by changing training and testing split 90%,10%:80%,20%:70%,30%, the size of the dataset, the number of epochs and network layers. Accordingly, on the first experiment register 90.1 % average training accuracy and 90.1% average validation accuracy performances were obtained by the first method. The second method achieves an average training of 82.4% and an average validation accuracy of 40% performances obtained. The third experiment conducted by increasing the number of data set 1600 and five network layers we get 70.1 training accuracy and 40.1 validation accuracy.

Table 2: shows a summary of some previous work/ literature

Author	Year	Approach/Models	Data source and domain	Data set	Text granularity	Language
Eshetu G.	2017	Manually constructed rules and subjectivity lexicon.	Self-collected customer reviews From A.A peoples on three issues.	375 reviews	Sentence	Afaan Oromoo
Wegderes T.	2017	Rule based	News reviews	300 reviews	Aspect	Afaan Oromoo
Tulu T.	2012	manually crafted rules and lexicon	Used hotels and universities reviews	484 reviews	Aspect	Amharic
Selama G.	2010	Lexicon	Movies and newspaper reviews	303 reviews	Document	Amharic
Weldu A.	2018	Rules based and lexicon	Politics reviews	582 comments/reviews	sentence	Tigrigna
Ghorbel and Jacot	2009	Machine Learning	Movie reviews	2000 reviews	Document	French
Yeshiwas and Abebe	2018	Deep Learning	News reviews	1600 reviews/ comments	Sentence	Amharic
Paredes *.et.al*	2017	Deep Learning /CNN	Product reviews	100000 tweets	Sentence	Spanish
Joel Kanku	2018	CNN	Product and reviews	2514 reviews	Aspect level	French
Alam,et.al	2017	CNN	Service and product	500 comments	Sentence	Bangla

As the above table shows a majority research done on non Afaan Oromoo lanague as well as various domain dataset and different text granularity. Hence our study is focused on Afaan Oromoo document sentiment classification by using socio politics dataset using a deep leaning approach.

CHAPTER FOUR: DESIGN AND IMPLEMENTATION OF SENTIMENT ANALYSIS OF AFAAN OROMOO REVIEWS

This chapter describes the general proposed approach, techniques used and algorithm selection of proposed sentiment classification in Afaan Oromoo language.

4.1. Proposed Approach

Our proposed approach for Sentiments Analysis for Afaan Oromoo text using deep learning approach is presented in the following section. Our work divide into three main stage. Those are: - firstly, we collect dataset need to be used to feed our network, secondly we use the preprocessing method in order to clean our text then use word embedding to obtain the feature vectors. Finally, configure the CNN model to train and classify document as positives or negatives.

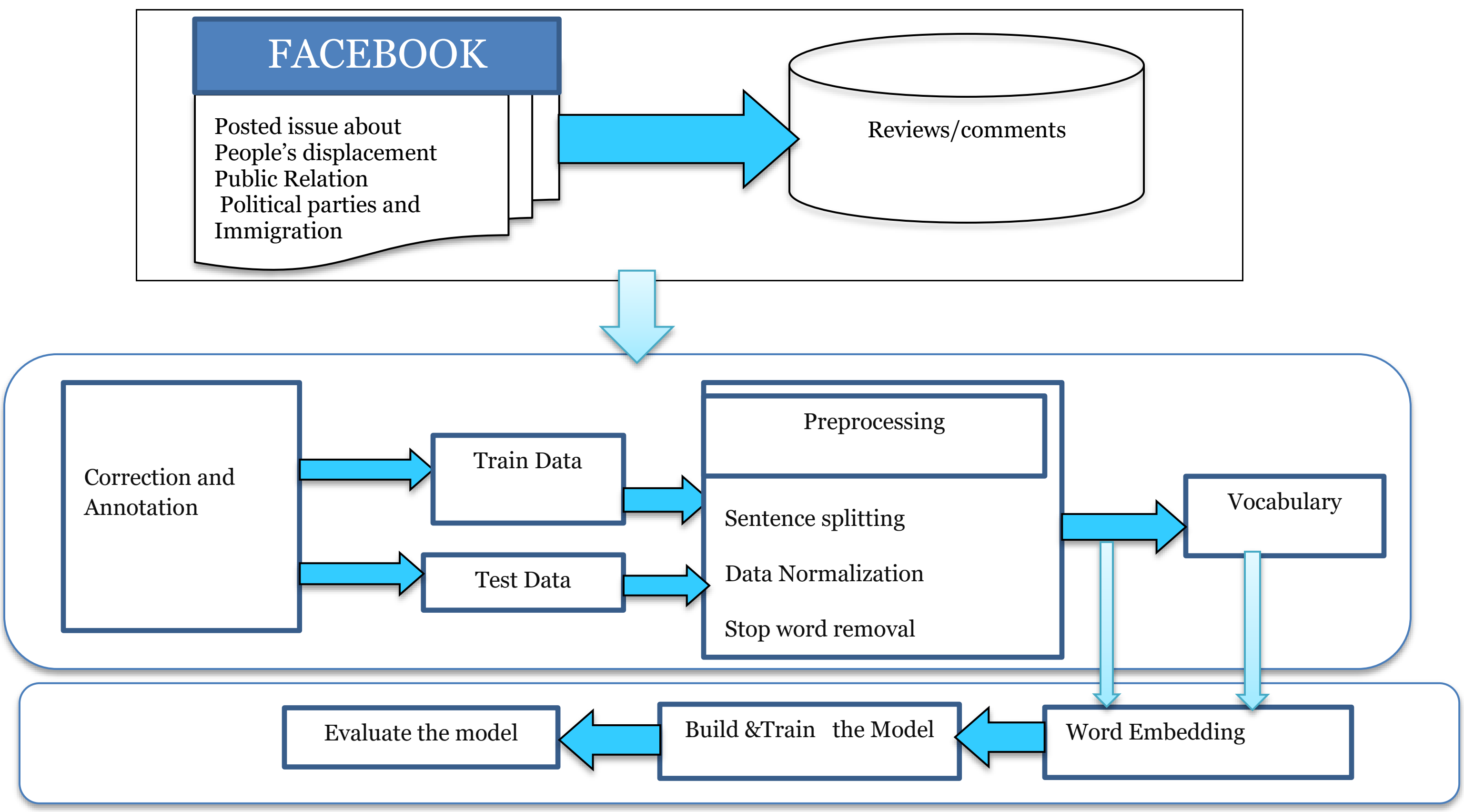


Figure 10: Architecture of our proposed work

4.1.1. Collection of Afaan Oromoo blogs

A lot of peoples are vigorously expressing their opinion on Facebook using Afaan Oromoo. Hence it is resulted for available several Afaan Oromoo public page and groups on Facebook world. Even though gigantic amount Afaan Oromoo data is existed in social media they are not annotated. In order to annotate this data various steps are engaged. First, Afaan Oromoo socio-political news post specifically several issues such as people's displacement , public relation, immigration and political parties post with respective audience comments are collected from verified Facebook public page of Fana broadcasting Corporation. We used an export comments tool to extract the comments from a public Facebook post and export them to a CSV/Excel file. Unwanted data, such as IDs, dates and URLs were removed, keeping the comments of the articles only.

Table 3: shows the detail of data source and the number of page per issues considered in this research.

		FBC	
S.No	Domain/issues	Num. Posted	Num.of comments
5.	People's Displacement	12	2,469
6.	Public relation	10	3,957
7.	immigration	8	2,384
8.	Political parties	52	4,118
Total num. of post and comments		82	12,928

4.1.2. Correction and Annotation of the collected data

In order to perform text classification through supervised learning, an annotated dataset will be needed for the training of a classifier. As we discussed in earlier section 4.1.1 our dataset consist out of socio-politics comments. A total of 12928 comments are collected from Afaan Oromoo FBC Facebook public news page. Texts generated by humans in social media sites like Facebook contain lots of noise that can significantly affect the results of the sentiment classification process. In Facebook social network, sometimes users use informal language while expressing their opinion. Usually, natural language like English uses automatic ways to correct spelling and grammar, translate from one language to another language. However, for Afaan Oromoo such tools and resource are not available in online even if developed for seek of fulfillment for the Degree.

So the researcher taken this into account and addressed several issues such as; correct spelling errors and grammar of some text manually, correct short form or contraction (Lakk., Fkn , Ykn , H/Bulaa , Q/Bulaa) and replication of characters, because some of them may useful in the polarity of the sentence, and a word written in Amharic or English mixed with Afaan Oromoo was translated into Afaan Oromoo using dictionary and language Expert.

Then, sentences are annotated independently by two human Afaan Oromoo native speakers, each annotated all the 12662 sentences. These sentences are classified into three classes for example, positive, negative and neutral.

Table 4: shows the number of positive, negative and neutral sentences as annotated by the two annotators.

Annotator 1		Annotator 2	total
Positive	6228	6211	
Negative	6049	6072	
Neutral	391	379	
Total			12662

A sentence is labeled/annotated based on the agreement of two annotators with the same label. For example, annotator 1 may label sentence X with positive sentiment and annotator 2 may label sentence X with positive, then sentence X is taken as having positive sentiment. In case of disagreement; if both annotator's assigned different labels to the same sentence, third annotator was taken into account. Depending on the decision of the third annotator, whether s/he agrees with the first annotator or the second one, a sentence was labeled accordingly. Sentence was rejected, in case the third annotator neither agrees with the first one nor with the second one.

Additionally we measure the inter-annotator agreement is calculated by using Kappa statistic. Kappa statistic is frequently used to test the inter-rater reliability, where inter-rater reliability is the measurement of the range up to which, same score is assigned to the same variable by different raters is called interrater reliability [60]. The importance of rater reliability is the fact that it represents the extent to which the data collected in the study; are correct representations of the variables measured. While there have been a variety of methods to measure interrater reliability, in this study we used a robust statistic Cohen's kappa since our annotators are two people. As stated in [60]Cohen suggested the Kappa result be interpreted as follows: values ≤ 0 as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

Table 5 : Interpretation of Cohen's Kappa

Value of Kappa	**Level of Agreement**	**% of Data that are Reliable**
0–.20	None	0–4%
.21–.39	Minimal	4–15%
.40–.59	Weak	15–35%
.60–.79	Moderate	35–63%
.80–.90	Strong	64–81%
Above.90	Almost Perfect	82–100%

Cohen's kappa measures the agreement between two raters who each classify *N* items into *C* mutually exclusive categories. In our case, the annotation was done independently by two native speakers of Afaan Oromoo as suggested by previous researchers and the conformity between the annotators was statistically measured. We found the value of the

Kappa (K) parameter to be 0.89 which is an evidence that there is a very good strength of agreement between the two annotators.

4.1.3. Preprocessing

In order to perform sentiment analysis the text or document collected from Afaan Oromoo Fana news page. Texts generated by humans in social media sites contain lots of noise that can significantly affect the results of the sentiment classification process. However, in order to acquire a good results this step plays a very important role in our system because preprocessing of text proved in the area of text classification and natural language processing. The preprocessing is increase the data quality to some extent and also needed to transform the data into a format that the model can interpret. We also employed preprocessing in our study and it will discussed in next section.

Before training our model with labeled data we removed punctuations and apply other preprocessing in the text. The basic approach to deal with this is to remove everything that isn't a letter. It should be borne in mind that sometimes punctuations can be really useful, like web addresses, where the punctuation often defines the web address. Therefore, the removal of punctuation should be tailored to the specific problem. In our case, we will remove all punctuations like removal of HTML-tags, non-alphanumeric tokens, blank lines, whitespace from the collected comments.

In the next stage, we change the case of the word to lowercase so that same words are not counted as different because of lower or upper case. Afterwards the data was saved and stored on disk in a Tab delimited or txt format for easy access later on.

Additional, in data preparations the comments with text shorter than a certain threshold was removed. The threshold used in this study was to remove all reviews that contained less than two character. Since reviews with a text length shorter than two character were directly removed a lower threshold could never be used, no restrictions were however set for an upper threshold at this point. Only positive and negative comments/reviews were considered in this thesis, the neutral reviews were removed.

Then, the dataset files are split into words and returns split sentences. After this, data are tokenized then turn them into lowercase and build vocabulary file from training data followed by padding all sentences to the maximum sentence length which is defined by the longest sentence and padded sentences are returned. It is very useful to do padding sentences to the same length because of it able us to proficiently batch our data. In our model implementation the python code clean_doc().*py* will take care of all these processes. After these processes is completed our input data will be ready to be used in our deep neural network.

Accordingly a techniques like, sentence splitter and stop word removal was applied in this thesis is discussed in detail the following section.

4.1.3.1. Data Normalization

Normalization is a process that converts a list of words to a more uniform sequence. This is useful in preparing text for later processing. By transforming the words to a standard format, other operations are able to work with the data and will not have to deal with issues that might compromise the process. In our case we expand short form into exact forma and converted all words to lowercase in order to reduce different consideration same word because lower and upper case.

4.1.3.1. Sentence splitter

Sentence splitter is one of the pre-processing tasks that can be undertaken when dealing with sentiment analysis in order to split the document of text in to separate sentences. This was supported out by considering punctuation marks that are used at the end of sentence; remove all punctuation from words, like colon (:), semicolon (;), full stop (.), space, and exclamation marks (!),remove all words that are not purely comprised of alphabetical characters, and remove all words that have a length two character.

4.1.3.2. Stop words Removal

It is another preprocessing task performed in order to remove obstructive terms. Many words are frequently used but are only meaningful in a sentence. These are called stop words. A stop-list is the name commonly given to a set or list of stop words. It is typically language specific, although it may contain words. Some of the commonly used stop words from English include “a”, “of”, “the”, “I”, “it”, “you”, “and”, also for Afaan Oromoo we collect some stopword by discussing with linguistics and take some form pervious done thesis.("haa", "akka","ishii","akkasi","utuu","kan","kee","mee,"irra","nuu","itti","kun","kana","irraa"," yoo","malee","isaa","waan","ni",akkasumas","booddee","erga","eegasii","eega","jechuu", "kanaafi","jechaa","otuu","otoo","ituu","akkum","dura","saniif","waan","tahullee","ituull ee","ta'ullee","otuullee","henna","innaa","waggaa","hogguu","yommuu","yemmuu","yo mmii","simmoo","oo","woo","akkam","ituu","hanga") are generally regarded as functional words which do not carry any meaning . The complete list of the stop word applied in this study is found in Appendix I. So, in our approach, we removed stop words that expand sentient words and enhance discrimination degree between documents.

In general, in preprocessing phase a lot task were performed. Spelling checks /word corrections were applied in reviews; also cleaned from emoticons and HTML tags. For our experiments applied later on, all sentences were converted to lowercase, punctuation and stop words were removed.

4.1.4. Keras Embedding layer

This is word embedding which learns collectively along with neural network model on a particular task of natural language processing for example document classification. Embedding layer converts positive integers into vectors. This requires integer encoded data as input so that each word have a unique integer. In this technique a corpus text is needed to be cleaned and get ready every word as one-hot encode. Keras Tokenizer API can be used to perform this step of data preparation. An embedding layer is initialized with random weights and then it will learn embedding's for all words in training dataset. This layer has always to be the first layer while model preparation and it requires three mandatory arguments. The first arguments is input_dim which used to determine the size of vocabulary in text data. Suppose the data is integer encoded in range of 0-10 then 11 would be the size of vocabulary. In our case we are calculating this value at runtime. Second argument is output_dim; it's the size of vector which we will get as output for each word from this layer. This value could be 100, 64, 32 or even larger, so we try this with different values for defined problem. In our case we have embedding_size of 100. The last is input_length which is the length of sequences as input. In our case we calculated using a function max_length.

Since there are differences of word embedding techniques exist and described in section 2.5, in this study we have used word embedding with the Keras deep learning library by using the Embedding layer and set size of the vector space to learn from scratch.

In order to obtain a vector that represents the sentences, we passed through a numbers steps and pre-process them. The first step, we tokenized each sentence, means convert them into a list of tokens. Then we load vocabulary that accomplished in preprocessing stage. Next, all of the training data of our socio-politics reviews loaded. This all are accomplished by sp*_process_docs()* in order to load the documents, splitting each review based on white space, removing punctuation, and then filtering out all tokens not in the vocabulary, and return them as a list of strings with one document per string. We need each document to be a string for easy encoding as a sequence of integers later. In the next step we have encoded each document as a sequence of integers since Keras embedding layer requires integer inputs, where each integer maps to a single token that has a specific real-valued vector representation within the embedding. We also used Tokenizer class in the Keras API to encode the training documents as sequences of integers. Then, we have constructed an instance of the class to train it all documents in the training dataset. Likewise, a vocabulary of all tokens in the training datasets are developed by a consistent mapping from words in the vocabulary to unique integers. Hence mapping words developed by our vocabulary file performed under preprocessing stage.

Now, the mapping of words to integers has been prepared, and it is useful to encode the reviews in the training dataset. Moreover, we have confirmed that all documents have the same length because of necessity for Keras for efficient computation. despite the fact that, truncated reviews to the smallest size or pad with the value '0' reviews to the maximum length is possible to have the same length of documents, in our case we padded all reviews to the length of the longest review in the training dataset. Additionally, it expected to encode and padded the test dataset, which is need later to evaluate the model after training process accomplished.

Furthermore, we have defined the class labels for the training dataset, which is needed to fit the supervised neural network model to predict the sentiment of reviews. In generally since the embedding requires the specification of the vocabulary size, the size of the real-valued vector space, and the maximum length of input documents we used total number of words in our vocabulary and plus one for unknown words as the vocabulary size, and 50, 100,150-dimensional vector space are used in order to observe different experiment result.

4.1.5. Convolutional Neural Networks

Convolutional Neural Network (CNNs), were designed to process images quickly, widely used in image classification problems and are the core of most Computer Vision systems today, from Facebook's automated photo tagging to self-driving cars.

More recently also started to apply CNNs to problems in Natural Language Processing and achieved some interesting results[38].

CNNs are basically just several layers of convolutions with nonlinear activation functions like ReLU or tanh applied to the results. In a traditional feedforward neural network need to connect each input neuron to each output neuron in the next layer. It is also called a fully connected layer, or affine layer. In CNNs no need to have connect each input neuron to each output neuron. Instead, it use convolutions over the input layer to compute the output. This results in local connections, where each region of the input is connected to a neuron in the output. Each layer applies different filters, typically hundreds or thousands, and combines their results.

The components of an image are simply pixels represented by integer values within a specific range. On the other hand the components of a sentence have to be encoded before fed to the CNN. For this purposed we may use a vocabulary. In this study we employed a Convolutional Neural Network, in order to feed a CNN with text data the first step was to map every word in the dataset to an integer or index value representing the word in a vocabulary. The vocabulary is constructed as an index containing the words that appear in the set of document texts, mapping each word to an integer between 1 and the vocabulary size. The variability in documents length or number of words in a document need to be addressed as CNNs require a constant input dimensionality. For this purpose the padding technique is adopted or reviews have different length the sequences are padded to equal length by filling with zeros the document matrix in order to reach the maximum length amongst all documents in dimensionality. Longer reviews are taken without truncated but, shorter reviews are padded with 0 values. The threshold for review text length were set to the maximum length of input documents.

CHAPTER FIVE: EXPERIMENTATION

In order to achieve the goal of this study we used several tools and evaluation metrics to make sure that the model are performing effectively. Effectiveness refers to the extent to which a system fulfills its objective. A Models trained with a good dataset with sufficient amount of data are an approximation of the true model. Thus we create a dataset to do experiment with our proposed framework. We experiment our proposed methodology with different hyperparameters and observe the results. Then we test our dataset with other classification algorithms and available Afaan Oromoo sentiment classifiers to compare their performance with our obtained highest accuracy. The experimental setup, sentimental analysis dataset, software tool used for implementing, evaluation metrics, experimental results and discussions are presented in the following sections.

5.1.Data

In this research, sentiment classification is conducted on Afaan Oromoo corpus. The dataset we used collected from Facebook Afaan Oromoo Fana news. The study is focused on socio-politics domain, since it the goal to classify the sentiment polarity socio-politics related issues of people displacement, public relation, immigration, and political parties. These are no ready-made dataset where each review is tagged with polarity; which is one of the most difficult parts of building a model. Therefore, the researcher was done labelling of the reviews into positive and negatives in order to meet the objective of the study.
The socio-politics corpus consists of 12928 review. Each sentence is annotated by two annotators. For each sentence, we took annotation intersection from two annotators. In the end, we obtained 6067 negative comments and 6200 positive comments.

5.2. Socio- Politics reviews

Socio-politics reviews are known to be more difficult with sentiment analysis. Although, the political reviews contain multi description than product reviews or movie reviews which brings more hidden comments and results in low performance. when a person writes a socio-political comment, he/she probably comments not only political but also political related people (e.g. Prime minister, president, deputy prime minister) while in product reviews, few people will care the issues like who has designed or manufactured the product. Therefore, commented features in a socio-political review are much richer than those product reviews. As a result socio-political review analyzing is more challenging than other domains.

5.2. Tools

Regarding the implementation of our proposed model, we use a laptop PC (with Windows 10 operating system, 2.40 GHz Intel CPU, 8 GB RAM and 1TB hard disk), python 3.6.6.for implemented our algorithm/coding, a deep learning framework mainly Keras. And install different machine learning functions and utilities from Numpy, itertools, scikit-learn, gensim, pandas.

5.3. Model and CNN configuration

In our experimental we used a CNN for classification of comments into positive/ negative classes and to evaluate the effectiveness of our model approach. The convolutional neural network architecture requires concatenated word vectors of the text as input. Regarding the implementation of this model, Tensorflow was used as backend.

We also set and tried different number of convolutional filters, kernel, filter or window sizes and pooling sizes. Other settings that was tuned were settings like number of epochs to train the model, batch size and whether certain weights should be trainable or not. Although a lot of different settings such as ReLU activation function, sigmoid function for classification in the output layer, binary cross-entropy as cost function and Adam as optimization function were used and tested. For our CNN model, we set an embedding vector size to follow a maximum length of the sequence and a vocabulary size.
We have configured CNN with 32 filters and a kernel size of 8. The output of the convolutional layer is passed through a non-linear activation function, before entering a pooling layer hence we used relu activation function. Next, the max-pooling layer is applied right after the convolutional layer to extract the most significant elements in each convolution and turn them into a feature vector. We also used, convolution followed by a pooling layer that reduces the output of the convolutional layer by half. It is common to periodically insert a Pooling layer in-between successive Conv layers in a ConvNet architecture. Its function is to progressively reduce the spatial size of the representation

to reduce the amount of parameters and computation in the network, and hence to also control overfitting.
Afterwards, the flatten operation takes the output and flattens the structure in order to create a single long feature vector, so that it can be used by the following dense layer for the final classification. The final dense layer with the activation function sigmoid transforms the output into a single output in order to indicate the sentiment or the output layer uses a sigmoid activation function to output a value between 0 and 1 for the negative and positive sentiment in the review. In the end we used the binary-cross-entropy loss function because of we are dealing with binary classification problem and accuracy as evaluation metric. Again, the Adam optimization algorithm is performed, since it is known to be very fast, efficient and had become very popular in recent deep learning model applications. We keep track of accuracy in addition to loss during training our model was trained for 50,100,150 epochs through the training data.

Regarding the implementation of this model, we illustrate in the following example how we setting up the model. A CNN with one convolution layer, several filters, a pooling layer and a fully connected layer using Python and Keras with Tensorflow backend is shown in the code below along with additional comments. Note that this is merely an example of how to set up a CNN in Keras, no dropout layer is for instance present in this example.

```
# Let's first assume as all the necessary libraries are loaded.
from Keras import ...  # Next the parameters Specified
EMBEDDING_DIM = 100,150
MAX_SEQUENCE_LENGTH = max_length, TRAINABLE = True
# Instantiate a sequential model
model = Sequential()# Embedding layer as the first layer
model.add(Embedding(vocab_size ,100
input_length=MAX_SEQUENCE_LENGTH,
trainable=TRAINABLE)) # One dimensional convolutional layer with specified number
of filters# filter size, padding and activation function
model.add(Conv1D(filters=32,100, kernel_size=4,8
padding='post', activation='relu'))
```

```
# One dimensional max pooling layer with window size 2
model.add(MaxPooling1D(pool_size=2))
# Flatten the output from the pooling layer
# so it fits the fully connected layer
model.add(Flatten())
# Fully connected layer with ReLU activation function
model.add(Dense(units=250, activation='relu'))
# Output layer predicting the output using sigmoid activation function
model.add(Dense(units=1, activation='sigmoid'))
# Other settings such as loss function,
# optimization algorithm and desired metrics
model.compile(loss='binary_crossentropy',
optimizer='adama',
metrics=['accuracy'])
# Train the model with training data
# for a given number of epochs and batch size
model.fit(X_train, y_train,
validation_data=(X_val, y_val),
epochs=10,
batch_size=32)
# Evaluate model on new data...
scores = model.evaluate(X_test, y_test)
```

In the following table 6, the most important features of the model is summarized. Dropout is used after the embedding layer and prior to the fully connected layer with dropout probability of 0.2 and 0.5 respectively. A dropout layer close to the output layer can also increase the robustness of the network and is mainly used to prevent overfitting. Choosing number of convolutional filters and their kernel sizes as well as other hyperparameters were determined by trial-and-error along with inspiration of previous work in the area. One epoch means a full training cycle on the training set and batch size is the number of training samples that are fed to the model at a time.

Table 6: model overview

embedding vector size	maximum length of sequence	vocabulary size	filter windows	pooling	dropout	batch size	epochs
50,100,150 observed	Calculated using max_length()	Our vocab_size	8	2	0.2,0.5	128	10

5.4. Experimental Result

This section presents the results for the tasks addressed in this work i.e., for the sentiment analysis. In the first iteration of testing, we use an Embedding layer as the first hidden layer. Since Embedding requires the specification of the vocabulary size, the size of the real-valued vector space, and the maximum length of input documents, we used a vocabulary size of 25128, 50-dimesional vector space, and the maximum document length was calculated by the *max_length* variable used during padding. We used a data set of around 12662 labeled comments, with a training split of 90%, and test of 10% size, 32 filter /parallel fields for processing words and a kernel size of 8 with a rectified linear ('relu') activation function. Additionally, we used a binary cross entropy loss function because the problem we are learning is a binary classification problem. Also, the efficient Adam implementation of stochastic gradient descent is used and we keep track of accuracy in addition to loss during training. Our model is trained for 10 epochs, or 10 passes through the training data. Our model, that were chosen to be evaluated in two classes performed well. When it comes to classifying two classes on new data that collected from other domains the models did not perform as good as they did on the dataset containing the training data.

In our experiment is observed the performance of the CNN classification without applying the preprocessing of text, by trained on 5460 positive comments and 5460 negative comments without preprocessing, then the classifiers are returned 46 training accuracy and 50.9 tests accuracy. Thus, as a first run, we use a small data set of 12662 comments and a vocabulary file of size 8675 which returns an initial training accuracy of 100% and 50% tests accuracy. Of course, this is a result that can be seen as expected since the models should perform better on the data that they were trained on. In the second iteration we also used the same data set and vocabulary size of the first iteration 100% training accuracy obtained with 79.99% tests accuracy. Which is shows as correlation exist between train and tests dataset in a domain. However, which suggest a training data is not sufficient. Then we increased our data by double to 25324 and experimented with parameters by increasing number of epochs. After evaluation of the CNN, the accuracy tests set has slight growth to 84.58% which is indicated as the significant change observed in the results.

However, it was discovered that the CNN is extremely resource demanding.

In general, CNN performs better but it requires solid computational resources and large amount of training sample. Additionally, this study has shown that in order to achieve meaningful performance of the classifier it has to be trained and tested on the same type of the dataset because the correlation exists between the classifier performance and domains, which are used for collecting training and testing samples.

5.5. Discussion

After a long period of preparation and lots of experiments, the results which has been shown in under experiment section are satisfying. Although there are still space to be improved, the goals which were proposed in the beginning of our study is achieved to some extent. So the study was guided by certain research question, we attempt to discuss in the following section. What challenges are observed during labelling Afaan Oromoo text into different polarities? As of the main goal this study is to detecting the polarities Afaan Oromoo text; corpus required.

Due to unavailability this resource it mandatory to label the text into different classes. In our case we collected opinion / data from Facebook social and classifying into three classes' i'e positive, negative and neutral. Even though, it is very hard to collect socio-polotics reviews or comments from social media, during labelling thus comments into three polarities we observed some challenges. For instance user are explained their feeling by without consider their writing spelling, which means they simply focused on forward their opinion by incorrect spelling. Hence, labelling thus text without correcting the spelling is result on low accuracy of our model we correct spelling and labeling in appropriate polarities. Additionally, spelling variation – e.g. "breda,brda,bareeda" *etc.* and Elongation –like gariiiiiiiiiiiiiii, gammadeerrrrrrrrrrrrrra,etc used by writer in comments, then we correct thus problems by understanding the reviews and considering the language structure. The other challenges are; user was used inappropriate punctuation marks in sentence for example the used comma/qoodu/,/ instead of using hudhaa/'/ in sentence, then we corrected in appropriate marks because it may provide some information for annotator. Since there is no ready resource is not available for this language we corrected them by manually.

The other research question is "What preprocessing tools and procedures applied to come up with quality data sets for training and testing Afaan Oromoo text?"

During evaluate our model without preprocessing, we only done splitting the training and test datasets since its prior to any data preparation and mandatory to evaluate the model. Then when we trained got 46.8 training accuracy and 50.8 test accuracy, which unsatisfying result in both training and tests set. This fact results from the unstructured and the noise in the text which decreases the performance of the classifier to identify the class of the comments.

The possible solution could be preprocessing the text in order to get satisfying results on the sentiment classification of the text. So, applying normalizing text, stemming and stop words be removal from these comments is very important. In our case have applied various procedure to enhance the quality of our dataset. We done splitting of tokens on white space, removed all punctuation from words and remove all words that are not purely comprised of alphabetical characters. And also we removed the stopword from our dataset, since the list of Afaan Oromoo stopword is not public available we identified the by ourselves and discussing with linguistics.

The more words, the larger the representation of documents, therefore we constrain the words to only those believed to be predictive. This is done by removing punctuation and numbers from all documents. A vocabulary of known words defined as a counter means dictionary mapping of words and their counts which is allows us to easily update and query. Each document added to the counter by defined functions, and removed all words that have a low occurrence, in our case once occurred word are removed.

The other research question is "What is the outcome of the size of the corpus used to train the model? "

In fact deep neural networks are data intensive and it performs better if trained with a dataset with large amount of data. As dataset in Afaan Oromoo language is publicly not available, we collect about 12928 Afaan Oromoo comments from different socio-politics news sources and process them to make them suitable to train the network. Among the comments, 6200 comments are positive, 6067 comments are negative, 395 are neutral and the rests are mixed comments and we rejected them. Since our aim is to deal with binary classification problems we ignore the neutral class, and only attempted to detect positive and negative textual sentiments of Afaan Oromoo language. In the first experiment we take only equally amount for both positive and negative class i'e 5460 positive comments and 5460 negative comments which 90% of the data used for train our network. After evaluated our network on test dataset, we increase the size of the comments to get a sufficient amount of data to train our model.

In the last we deal with "How can deep learning be used to detect polarity of user reviews in Afaan Oromoo?" Based on the result of the tests and training of the convolutional neural network it showed that this type of neural network could be used for sentiment analysis on Afaan Oromoo texts. The result over the models covering two classes, positive and negative, showed some very good results.

Since the models show a really good result on both training data and test data, this indicates that the networks have learned from the training data and can take the knowledge with them when new real-time data has to be analyzed.

Eshetu [10] proposed a sentiment analysis model for opinionated Afaan Oromoo texts.

The author conducted an experiment on three different datasets manually collected from peoples of Addis Ababa house renters, A.A Anbesa bus users and English to Afaan Oromoo online dictionary. For house rental reviews an F-Measure of 0.699, for A.A Anbesa bus reviews an F-Measure of 0.718 were achieved, but our classifier model obtains a classification accuracy of 79.99%, using word embedding technique and deep learning which is show that as the accuracy was improved. But, it is important to mention that we did not carry out a direct comparison of our results with those reported in related works because there is a lack of deep learning approaches for sentiment analysis in Afaan Oromoo.

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS

This chapter introduces conclusions of the work that was done as well as discussions about possibilities for performing future work.

6.1. Conclusions

Sentiment analysis task is under research since the early 2000s and it is still in developing stage. Social media networks such as Facebook is generate gigantic amount of data, which more facilitate for being studied of sentiment analysis. For example if we take Facebook, it is less informative opposed to usual review or comment since it contains a lot of noisy data which makes classification of comments more challenging.

This thesis investigates the algorithms that used for Afaan Oromoo sentiment classification. As we observed from the literature review, the majority of sentiment analysis approaches on rely on supervised machine-learning methods. Consequently, convolutional neural network approaches studied in this thesis since these method a current state of the art as witnessed by many researchers and it provided a blameless results on the experiments. Hence, we implemented CNN algorithms and their performance was evaluated based on experiments. The classification model was trained on socio-politics datasets collected from FBCafanoromo, and the significance of the preprocessing stage was discussed since the comments are utilized as training data.

In this thesis binary classification is considered, namely, the comment is assigned a positive or negative label according to the sentiment conveyed in it. Then experiment was conducted and illustrate that the performance increases along with the quality of the data. It shows sufficient results on the sentiment classification achieved and the best result of the accuracy that was achieved 79.99 % when convolutional neural network was learned from the whole set of comments.

In general, CNN performs better, but it requires solid computational resources and large amount of training sample. Additionally, this study has shown that in order to achieve meaningful performance of the classifier it has to be trained and tested on the same type of the dataset because the correlation exists between the classifier performance and domains, which are used for collecting training and testing samples.

The recommendations that can be applied for the model to improve the performance of the classifier are described in the following section.

6.2 Future Perspectives

There are still space to be improved which can be proposed in the future. Therefore, future work will involve investigation of other approaches for preprocessing comments because they have to be more thoroughly filtered to achieve the higher accuracy.

There are several directions that can be performed. As mentioned earlier, Facebook comments may contain a lot of spelling mistakes, hence, an automatic spelling corrector can be applied to exclude typos.

Additionally, Facebook comments contain huge amount of emoticons and expressions that convey laugh, such as kik..ki-kik,qis-qis-qis , ha-ha-ha, that have to be generalized and labeled whether emoticon/expression refers to a positive or negative meaning, the ones that are ambiguous have to be removed from the training dataset.

Another imperative point for future researchers it better to apply stemming in preprocessing step in order to increase the performance of the model ,

Besides, another experiment that may be carried out is the replacement of the abbreviations with their full meaning. It obviously will increase the size of the training corpus but may add more sense to the comments.

Moreover, it would be interesting to add neutral class and check the performance of the classifier.

Finally, we also recommend other researchers to considering and conducting experiments other neural network models such as Recursive Neural Tensor Networks (RNTN), Recurrent Neural Networks (RNN), and Long Short Term Memory (LSTM) to observe if it will achieve better performance and able to encode long-range dependencies since for some language modeling tasks a long-distance dependence matters.

References

[1] Saud Saleh Alotaibi, “Sentiment Analysis in the Arabic Language Using Machine Learning, Fort Collins, Colorado,” 2015.

[2] M. Rushdi-saleh, M. T. Martín-valdivia, L. A. Ureña-lópez, and J. M. Perea-ortega, “OCA : Opinion Corpus for Arabic,” vol. 62, no. July, pp. 2045–2054, 2011.

[3] B. P. and L. Lee, *Opinion Mining and Sentiment Analysis*. Foundations and Trends in Information Retrieval, 2008.

[4] G. Kaur and K. Kaur, “Sentiment Analysis on Punjabi News Articles Using SVM,” *Int. J. Sci. Res.*, vol. 6, no. 8, pp. 414–421, 2017.

[5] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” *Empir. Methods Nat. Lang. Process.*, vol. 10, no. July, pp. 79–86, 2002.

[6] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Found. Trends® InformatioPang, B., Lee, L. (2006). Opin. Min. Sentim. Anal. Found. Trends® Inf. Retrieval, 1(2), 91–231. doi10.1561/1500000001n Retr.*, vol. 1, no. 2, pp. 91–231, 2006.

[7] R. M. Duwairi, “Arabic Sentiment Analysis using Supervised Classification,” *1st Int. Work. Soc. Networks Anal. Manag. Secur. (SNAMS - 2014), August 2014, Barcelona, Spain Arab.*, no. August, pp. 1–10, 2014.

[8] A. Tripathy, A. Agrawal, and S. K. Rath, “Classification of Sentimental Reviews Using Machine Learning Techniques,” *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015.

[9] CSA, “2007 Population and Housing Census of Ethiopia: Administrative Report,” no. April, pp. 1–117, 2012.

[10] G. Eshetu, “Sentiment Analysis for Opinionated Afaan Oromoo Texts, M.sc Thesis, Addis Abeba University,Addis Abeba ,Ethiopia,” 2017.

[11] D. Lamoureux, “Number of monthly active Facebook users worldwide as of 4th

quarter 2018 (in millions) February 26, 2019." .

[12] A. Shrivatava and B. Pant, "Opinion Extraction and Classification of Real Time Facebook Status," *Glob. J. compuer Sciemce Technol.*, vol. 12, no. 8, 2012.

[13] W. Tariku, "Aspect based summarization of opinionated Afaan Oromoo news text,M.Sc ,Debre Berhan University,Debre BerhanEthiopia," 2017.

[14] M. Bahja, "Automated Analysis of Patient Experience Text Mining using a Design Science Research (DSR) Approach," *Eighth Int. Conf. Bus. Intell. Technol.*, no. c, pp. 21–24, 2018.

[15] J. Sundström, "Sentiment analysis of Swedish reviews and transfer learning using Convolutional Neural Networks," Uppsala University, 2018.

[16] B. Liu, "Sentiment Analysis and Opinion Mining," *Morgan & Claypool*, no. May, 2012.

[17] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications : A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.

[18] M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews," *Am. Assoc. Artifical Intell.*, pp. 755–760, 1998.

[19] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Election Forecasts With Twitter : How 140 Characters Reflect the Political Landscape," vol. 29, no. 4, pp. 402–418, 2016.

[20] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market .," *J. Comput. Sci.*, vol. 2(1), pp. 1–8.

[21] Z. Qin, "A Framework and practical implementation for sentiment analysis and aspect exploration," University of Manchester, 2016.

[22] P. D. Turney, "Thumbs Up or Thumbs Down ? Semantic Orientation Applied to Unsupervised Classification of Reviews," *Inst. Inf. Technol. Natl. Reseach Canada*, 2001.

[23] M. Tsytsarau and T. Palpanas, “Survey on mining subjective data on the web,” *Data Mining and Knowledge Discovery*, vol. 24, no. 3. pp. 478–514, 2012.

[24] V. Gupta, L. C. Science, and G. S. Lehal, “A Survey of Text Mining Techniques and Applications,” *J. Emerg. Technol. web Intell.*, vol. 1, no. 1, pp. 60–76, 2009.

[25] J. Smailović, “Sentiment Analysis in Streams of Microblogging Posts, International Postgraduate School Ljubljana, Slovenia.,” 2014.

[26] S. Blair-goldensohn *et al.*, “Building a Sentiment Summarizer for Local Service Reviews,” *Int. World Wide Web Conf. Com- mittee (IW3C2).*, 2008.

[27] V. Hatzivassiloglou and K. R. Mckeown, “Predicting the Semantic Orientation of Adjectives,” *35th Annu. Meet. Assoc. Comput. Linguist. 8th Conf. Eur. Chapter Assoc. Comput. Linguist.*, pp. 174–181, 1997.

[28] P. D. Turney and M. L. Littman, “Measuring Praise and Criticism : Inference of Semantic Orientation from Association,” pp. 1–37.

[29] M. Ghiassi, J. Skinner, and D. Zimbra, “Expert Systems with Applications Twitter brand sentiment analysis : A hybrid system using n -gram analysis and dynamic artificial neural network,” *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6266–6282, 2013.

[30] O. Kummer and J. Savoy, “Feature Selection in Sentiment Analysis,Institut d’informatique, Université de Neuchâtel,” no. January, 2012.

[31] E. Riloff, S. Patwardhan, and J. Wiebe, “Feature subsumption for opinion analysis,” *Proc. 2006 Conf. Empir. Methods Nat. Lang. Process. - EMNLP ’06*, no. July, p. 440, 2006.

[32] K. Dave, K. Dave, S. Lawrence, S. Lawrence, D. M. Pennock, and D. M. Pennock, “Mining the peanut gallery: Opinion extraction and semantic classification of product reviews,” *Proc. 12th Int. Conf. World Wide Web*, pp. 519–528, 2003.

[33] F. Peng, D. Schuurmans, and S. Wang, “Language and task independent text categorization with simple language models,” *Proc. 2003 Conf. North Am.*

Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - NAACL '03, vol. 1, no. June, pp. 110–117, 2003.

[34] D. Bespalov, B. Bai, Y. Qi, and A. Shokoufandeh, "Sentiment classification based on supervised latent n-gram analysis," *Proc. 20th ACM Int. Conf. Inf. Knowl. Manag. - CIKM '11*, p. 375, 2011.

[35] Bo and S. Lee, "sentiment classification using machine learning techniques," *Empir. Methods Nat. Lang. Process.*, 2012.

[36] Brooke, "A sematic approach to automated text sentiment analysis," 2009.

[37] R. Socher *et al.*, "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," no. October, pp. 1631–1642, 2013.

[38] Y. Kim, "Convolutional Neural Networks for Sentence Classification,New York University," 2014.

[39] A. Severyn, "Twitter Sentiment Analysis with Deep Convolutional Neural Networks," *Qatar Comput. Reseach Inst.*, pp. 959–962.

[40] A. Severyn and A. Moschitti, "Training Deep Convolutional Neural Network for Twitter Sentiment Classification," *Proc. ofthe 9th Int. Work. Semant. Eval. (SemEval 2015),* no. SemEval, pp. 464–469, 2015.

[41] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation."

[42] F. Shaikh, "Comparison between Deep Learning & Machine Learning." .

[43] J. K. Ntumba, "Sentiment Analysis in French for Improving Service and Products: A Deep Learning Approach. Master final thesis,Vytautas Magnus University," 2018.

[44] R. Řehůřek, "gensim_ models," *https://radimrehurek.com/gensim/models/word2vec.html*, 2019. .

[45] I. Bedane, "The Origin of Afaan Oromo : Mother Language," vol. 15, no. 12, 2015.

[46] A. S. Genemo, “Afaan Oromo named Entity Recognition using Hybrid Approach,M.Sc. Thesis, Addis Ababa University, Addis Ababa, Ethiopia,” no. March, 2015.

[47] Tulu.T, “Opinion Mining from Amharic Blog,M.Sc. Thesis, Addis Ababa University, Addis Ababa, Ethiopia,” 2013.

[48] H. Ghorbel and D. Jacot, “Sentiment Analysis of French Movie Reviews,Univeristy of Applied Sciences Western Switzerland, Haute Ecole Arc Ingeniere,” 2009.

[49] B.Kama, “Feature based Sentiment analysis on Informal Turkish Texts,” Middle East Technical University, 2016.

[50] W. P., “A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts.” .

[51] S. Gebremeske, “Sentiment Mining Model for Opinionated Amharic Texts,M.Sc. Thesis, Addis Ababa University, Addis Ababa, Ethiopia,” 2010.

[52] W. Ashebr, “Sentiment Analysis Model for Tigrigna Opinionated Texts,M.Sc. Thesis, University of Gonder, Gonder, Ethiopia,” 2018.

[53] N. Mukhtar and M. A. Khan, “Urdu Sentiment Analysis Using Supervised Machine Learning Approach,International Journal of Pattern Recognition and Artificial Intelligence Vol. 32, No. 2 (2018),” vol. 32, no. 2, pp. 1–15, 2018.

[54] A. K. J. Li, “Classification of Text Documents,” *Dep. ofComputer Sci. Eng. Michigan State Univ. East Lansing, Michigan, USA*, vol. 41, no. 8, pp. 537–546, 2005.

[55] M. A. Paredes-valverde, R. Colomo-palacios, M. P. Salas-zárate, and R. Valencia-garcía, “Sentiment Analysis in Spanish for Improvement of Products and Services : A Deep Learning Approach,” *Hindawi Sci. Program.*, vol. 2017, 2017.

[56] K. Svensson, “Sentiment Analysis With Convolutional Neural Networks,” Swedish University, 2017.

[57] H. Alam, "Sentiment Analysis for Bangla Sentences using Convolutional Neural Network," *2017 20th Int. Conf. Comput. Inf. Technol.*, pp. 22–24, 2017.

[58] V. Nukarinen, "Automated Text Sentiment Analysis for Finnish," Tampere University of Technology, 2017.

[59] A. Yeshiwas, "Deep Learning Approach for Amharic Sentiment Analysis Deep,M.Sc. Thesis, Gonder University, Gonder, Ethiopia," 2019.

[60] M. L. McHugh, "Interrater reliability_ the kappa statistic," pp. 276–282, 2012.

Appendix I: Collected stopwords

"immoo"	"haa"	"mee"	"akka"	"ishii"	"akkasi"	"utuu"	"kan"
"kee"	"hanga"	"jechuun"	"ol"	"waan"	"akkam"	"henna"	"oliif"
"waggaa"	"akkasumas"	"hoggaa"	"kanaaf"	"oliin"	"akkuma"	"hogguu"	"kanaafi"
"yammuu "	"hoo"	"kanaafu"	"osoo"	"yemmuu"	"ammo"	"illee"	"otoo"
"yeroo"	"an"	"keenya"	"otumalle"	"ykn"	"ani"	"innaa"	"keenyaa"
"otuu"	"yommii"	"booda"	"inni"	"keeti"	"otuullee"	"yommu"	"booddee"
"isaa"	"keetii"	"saniif"	"yoo"	"dura"	"isaan"	"koo"	"silaa"
"isee"	"kun"	"simmoo"	"yookaan"	"eega"	"yookiin"	"eegana"	"iseen"
"malee"	"sun"	"yookinim oo"	"eegasii"	"ishee"	"moo"	"ta`ullee"	"yoom"
"ennaa"	"isheen"	"nu"	"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"	"hogguu"	"yommuu"	"yemmuu "	"yommii"
"simmo"	"oo"	"woo"	"akkam"	"ituu"	"hanga"		