2022-03

Aspect Level Sentiment Analysis Using Hybrid Deep Learning Approach for Amharic News Comments

Zemenu, Mekonnen

Bahir Dar University

Bahir Dar Institute of Technology

School of Graduate Studies

Faculty of Computing

Aspect Level Sentiment Analysis Using Hybrid Deep Learning Approach for Amharic News Comments

By: Zemenu Mekonnen

Advisor: Abreham Debasu (Ass. Prof.)

March, 2022

Bahir Dar Ethiopia

Bahir Dar University

Bahir Dar Institute of Technology

School of Graduate Studies

Faculty of Computing

Aspect Level Sentiment Analysis Using Hybrid Deep Learning Approach for Amharic News Comments

A Thesis Submitted to Bahir Dar University Institute of Technology School of Research and post graduate studies faculty of computing Department of Information Technology for The Requirement of Masters of Science in Information Technology

By: Zemenu Mekonnen

Advisor: Abreham Debasu (Ass. Prof.)

March, 2022

Bahir Dar Ethiopia

BAHIR DAR UNIVERSITY
BAHIR DAR INSTITUTE OF TECHNOLOGY
SCHOOL OF GRADUATE STUDIES
FACULTY OF COMPUTING
Approval of thesis for defense result

I hereby confirm that the changes required by the examiners have been carried out and incorporated in the fina l thesis.

Name of Student Zemenu Mekonnen Shiferaw Signature ______ Date March 11. 2022

As members of the board of examiners, we examined this thesis entitled " Aspect Level Sentiment Analysis Using Hybrid Deep Learning Approach for Amharic News Comments " by Zemenu Mekonnen Shiferaw. We hereby certify that the thesis is accepted for fulfilling the requirements for the award of the degree of Masters of science in " Information Technology ".

Board of Examiners

Name	Signature	Date
Name of Advisor: Abraham D.		17/03/2022
Name of External examiner: Michael Melese (PhD)		March 11, 2022
Name of Internal Examiner: Abdulkerim Mohammed (PhD)		17/03/2022
Name of Chairperson: Tamir Anteneh		22/03/2022
Name of Chair Holder: Abdulkerim Mohammed (PhD)		17/03/2022
Name of Faculty Dean: Asegahegn E.		22/03/2022

Faculty Stamp

ባህር ዳር ዩኒቨርሲቲ ባህር ዳር ቴክኖሎጂ ኢንስቲትዩት ኮምፒዩቲንግ ፋካሊቲ
Bahir Dar University Bahir Dar Institute Of Technology Faculty Of Computing

Table of Contents

List of Figures

List of Tables

Declaration

I declared that this thesis is my own work. Incompliance with international accepted practices, I have acknowledged and refereed all material used in this work. I understand that non-adherence to the academic honesty and integrity, misrepresentation, fabrication of any idea, data, fact, sources, will constitute sufficient ground for disciplinary action by the university and can also invoke penal action from the sources which have not been properly cited and acknowledged.

Acknowledgement

First of all, I would like to thank the Almighty GOD, the creator of everything, source of wisdom and beginning of Knowledge for giving me the strength, Knowledge, ability and opportunity to undertake this research and to complete it. Without his blessings, this achievement would have not been possible. I thank him he is with me from the beginning to the end of this work. I would like to thank holy virgin St Mary, mother of Lord. Foremost I would like to express my sincere gratitude to my best friend Mezmur Bekalu for his priceless support and consultation on the entire work. Also I would like to extend my heartfelt acknowledgement to my Advisor **Abreham Debasu** for the continuous support on my thesis work.

Abstract

Natural Language Processing, or NLP, is a discipline of computer science that uses computer-based methods to analyze language in text and speech. The purpose and ambition of NLP is to enable computers to converse quickly and easily. Emailing, texting, and cross-language communication are all used in our daily lives. Sentiment analysis, also known as opinion mining, is a branch of natural language processing that mines a text to extract subjective information from the source material, allowing a company to better understand the social sentiment surrounding their brand, product, or service by monitoring online conversations. Nowadays, in the era of social media, firms' service and products have a significant impact on customer input and feedback. Customers freely express their feelings about a company's service and products on the company's social media pages. Companies face a significant and tough problem in extracting meaningful information from this unstructured, unorganized, massive, and fragmented data. Amhara Media Corporation is one victim of this and sentiment analysis is used for resolving such issues, it allows businesses to quickly learn what customers think and feel about their service or product. Sentiment analysis can be performed in Document level, Sentence level and Aspect level. Document level sentiment classifies the whole document into positive negative or neutral but the customer opinion need may not in that manner. Hence the negativity or positivity of a single sentence does not measure the entire document as a whole. Sentence level sentiment analysis also categorizes the entire sentence as negative, good, or neutral, although consumer feedback may not be in those categories. Many publications have been written on Amharic sentiment analysis, however none of them have looked at the aspect level or used a deep learning approach. This research focuses on sentiment analysis utilizing aspect level with a hybrid deep learning approach. The Dataset were collected from in Excel format from the Amhara media corporation's official Facebook page. To improve the quality of the data, the researcher applied tokenization, stop word removal, Emoji, punctuation, eliminating non-Amharic texts, word embedding, and normalization and converting English language comments to Amharic and utilize comment exporter software to collect 10,000. The researcher used CNN, LSTM, Hybrid CNN with GRU and hybrid CNN with LSTM. Finally, Hybrid CNN with GRU is selected as best approach due to the better accuracy it registered.

Keywords: Sentiment Analysis, Aspect Level, Amharic, Comment

List of Acronyms

AAU	Addis Ababa University
AMC	Amhara Media Corporation
ANN	Artificial Neural Network
BDU	Bahir Dar University
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DBU	Debre Birhan University
GRU	Gated Recurrent Unit
HTML	Hyper Text Markup Language
LSTM	Long Short-Term Memory
MNB	Multinominal Naïve Bayes
NB	Naïve Bayes
NLP	Natural Language Processing
POS	Part of Speech
RAM	Random Access Memory
RNN	Recurrent Neural Network
SA	Sentiment Analysis
SVM	Support Vector Machine
TF-IDF	Term Frequency- Inverted Document Frequency
UOG	University of Gondar
URL	Uniform resource Locater

Chapter One

1. Introduction

As discussed by Devika et al. (2016) Natural Language Processing, or NLP, is a branch of computer science that analyzes language in text and speech using computer-based methods. It is utilized for practical purposes that aid us in our daily lives, such as texting, e-mail, and cross-language communication. Natural Language Processing's main goal and ambition is to let computers converse with people more readily. NLP allows computers to read text, hear voice, analyze it, gauge sentiment, and identify which bits are significant. People can simply order machines and computers to grasp human language instructions with the help of NLP by customizing in a way that machines and computers can understand. Today, various NLP techniques are used by companies to analyze social media posts and know what customers think about their products (Grosz, 1982). Companies are also using social media monitoring to understand the issues and problems that their customers are facing by using their products. Using speech recognition, language translation, text summarization, sentiment analysis, relationship extraction, document categorization, and named entity identification, we can extract important information and knowledge (Chilot, 2019).

1.1. Background

Sentiment analysis, often known as opinion mining, is a subset of natural language processing that determines whether a document is positive, negative, or neutral. Sentiment analysis is text mining that finds and extracts subjective information from source material, allowing a company to better understand the social sentiment surrounding their brand, product, or service while monitoring online conversations (Nandi and Agrawal, 2016).There are a lot of social media users nowadays, and a lot of information is shared in fractions of seconds all over the world. People express their emotions through comments and reactions on social media posts. Sentiment analysis is a method of assessing these remarks and categorizing the customer's feelings and emotions as good, negative, or neutral (To, A. T. S, 1991). Polarity classification is a method of identifying or classifying a given sentiment as positive, negative, or neutral depending on the weight assigned to the sentiment and document propagation based on the document's circumstances. Sentiment analysis is being used in a variety of fields, including the film and music industries, sports and recreation, large corporations to maintain staff morale, product vendor companies, and social media. Sentiment analysis in social media plays a critical role in determining the emotions and feelings of individuals who remark on a specific person's or organization's post. Sentiment analysis incorporates emotions and opinions in addition to a simple count of mentions or comments. It entails gathering and evaluating information from people's social media posts about your brand.

According to Chachra et al. (2018) level of Sentiment analysis can be done in one of the following ways:

Document level: The entire document is categorized as good, negative, or neutral at this level.

Sentence level: The entire sentence is classified as good, negative, or neutral at this level.

Aspect level: For each feature present in the document/sentence, the entire document/sentence/phrase or comment is classified as positive, negative or neutral at this level.

The polarity of a document or a sentence level only classifies the entire document or sentence; it does not identify the aspects present in the document/sentence. For example, just because the document's polarity is positive or negative does not mean that the document has a positive or negative opinion for each aspect. Opinion mining at the aspect level is used to determine people's opinions on various topics.

There are two aspects that can be found. Explicit aspects are nouns and noun phrases that are easily detected in the reviews. “The voice quality of this phone is excellent,” for example, is an explicit component of her voice quality that can be seen clearly in the sentence.

Explicit aspects are nouns and noun phrases that are not easily detected in the reviews. Implicit aspects are nouns and noun phrases that are not easily identified in the reviews. For example, in the statement "This phone does not fit in my pockets," the implicit aspect "does not fit in my pockets" denotes the aspect size that cannot be stated recognized in the sentence.

1.2. Statement of the Problem

Companies and organizations are interested in hearing people's thoughts and reactions to their services and products. They have interest to know what peoples are thinking about their product and what reactions and feelings do they have in their overall service and product they delivered. how our product is positive to the people who use our service, what peoples like our product would be. All these things and people’s emotions plus feelings are gathered from the customers by using World Wide Web and Internet. People's daily activities, reactions, and feelings have recently become more widely expressed through social media platforms. Social Media is currently riding the world with huge amount of information and media data. The opinions of the people are much interested to know the level of the company level of acceptance of a certain organization or company by social Media users. Every day, the Amhara Media Corporation updated their Facebook page with a variety of news. As stated above, Amharic has more than 35, million speakers; the feedback from such huge population has a paramount importance. Because of the vast amount and variety of information available at Amhara Media Corporation, it is difficult to easily manage people's views and feelings in order to describe the general public's reaction to a particular news item posted on their Facebook page.

In social media, mining numerous sentiments and extracting relevant knowledge from user opinions, feelings, and emotions, as well as obtaining any input information from customers, is a huge and tedious task. Furthermore, understanding client sentiments and making a choice based on such raw and unstructured data is tough (Aragaw, 2015).Various studies have been conducted in a variety of languages around the world, including Amharic, English, Arabic, Indian, and Chinese (Heikal et al., 2018; Le & Nguyen, 2015; Prabha & Srikanth, n.d.).

Studies on sentiment analysis of Amharic language are conducted on different levels using machine learning and lexicon-based approaches. In Tadesse (2015), combination of lexicon based and machine learning approaches are used to automatically extract opinions from Amharic text. In the case of Amharic language, the vast majority of studies have concentrated on document and sentence level opinion mining systems.

As stated in Katrekar (2013), Sentiment classification can be done at document level, Sentence level and Aspect or Feature level. The comparison between the three levels of sentiment analysis shows that in document level sentiment classification the whole document is classified as positive or negative, but it doesn’t explicitly tell what feature of the object being considered is really negative or positive, additionally being classified as negative, the document may contain features that contain positive opinion making document level sentiment analysis not generalize and do not consider finer details of an object. The other approach sentence level sentiment analysis classifies the whole sentence as positive or negative. In both approaches, the opinionated text does not clearly show what aspect of the objects feature satisfies or dissatisfy customers. But the third approach (aspect level sentiment classification) is fine-grained approach that identifies the entity, features and the aspect of an entities feature is opinionated and takes the lower-level details of the object being reviewed. It first identifies the aspect of the object and finds them in the review then classifies as positive or negative. According to (Katrekar, 2013), aspect level sentiment analysis is better sentiment classification technique but challenging.

When having closer look at sentiment analysis works in Amharic, most of them are done in document level or sentence level. As stated above sentiment analysis will be more representative if it is done at a more fine-grained level that clearly shows what aspect of the object or entity being reviewed customers like or dislike. A favorable document on an item does not imply that the client approves of all of the object's attributes or features. A negative document, on the other hand, does not imply that the buyer disapproves of the entire item. The customer often writes both good and negative features of an object in an evaluation document, such as a product review or service feedback, even if the overall attitude on the object is positive or negative. Therefore, an aspect based Amharic sentiment analysis is needed.

In addition, most of Amharic sentiment analysis works are done using machine learning approach or lexicon-based approach. In (Tadesse, 2015), a combination of lexicon based and machine learning approaches are used for Amharic sentiment classification. According to Gzate (2020), machine learning technique is applied to gain insights into a political sentiment from a given Amharic text. The suitability of machine learning approach for Amharic sentiment analysis is explored in (Yimam et al., 2020). But recently deep learning techniques are applied for sentiment analysis of foreign languages and achieved better results (Ain et al., 2017; Costea et al., n.d.; Dashtipour et al., 2021; Heikal et al., 2018). In Dashtipour et al. (2021) Persian sentiment analysis is proposed using SVM and deep learning algorithm such as CNN and LSTM. The comparison result revealed that deep learning has performed better than SVM. Therefore, it is important to apply to apply deep learning for Amharic sentiment analysis and see whether it achieved better results or not. This is because deep learning techniques (RNN, CNN, LSTM, GRU) are currently becoming flashpoint applied to classify sentiments and performing NLP tasks (Heikal et al., 2018).

Therefore, in this study an aspect-based Amharic sentiment analysis that analyzes sentiments at lower level of granularity is conducted using different deep learning techniques.

Finally, the study would seek to answer the following questions:

- Which hybrid deep learning technique is better to perform sentiment classification of Amharic news comments?
- To what extent Aspect level sentiment analysis on Amharic news works in terms of classification accuracy?

1.3. Objective

1.3.1 General Objective

The main objective of this study is to develop aspect based Amharic sentiment analysis model by using deep learning model on Amharic news comments.

1.3.2 Specific Objective

The following specific tasks were done in order to meet the study's main objective:

- To design hybrid deep learning model for Amharic Sentiment analysis using aspect Level.
- To design and test an Amharic sentiment analysis model.
- To compare and contrast the results of CNN, LSTM, CNN-GRU and CNN-LSTM algorithms on aspect level sentiment analysis of Amharic News Comments.

1.4. Methodology

The solution strategy to be adopted for the research is described as follows. the researcher explained the literature review, what sources of data we used, which tools and algorithms are used, and how the analysis and evaluation of the discovered knowledge done as follows.

1.4.1. Data Collection

For the Data source selection, the researcher used Amhara Media Corporation (AMC) official Facebook page as the primary data source. Since it was legal under Facebook's rules and conditions, and people can freely express themselves on social media. Despite the fact that the researcher focuses solely on the following aspects from the broad idea of sociopolitical domain area, namely immigration data, public relations, and conflict, the researcher scrape the data using the Facepager and software text processing tool and the comment exporter software to download the comments from the post by using post Id. For training and testing the model, the researcher used10000 reviews collected from this issue, of which 4000 contains common vocabulary file.

1.4.2. Design Approach

The model used to develop aspect level sentiment analysis on Amharic news is utilizing deep learning approach. We have used different deep learning approaches for the model, among this Gated Recurrent unit, convolutional neural network and Long Term Short Memory (LSTM) have been applied and the hybrid approach which utilizing CNN and GRU way is efficient and used in this work.

1.4.3. Tools and Techniques

In this research, there are three components these are: preprocessing, feature extraction and sentiment classification. The preprocessing includes removal of unnecessary texts, symbols and characters and also comprises character and short form normalization. This task is performed using Python program. After that the processes comment is embedded using word2vec method. The other component, feature extraction which is done by using CNN and the classification process is performed by using GRU.

1.4.4 Evaluation

For the evaluation of opinion classification using regular expression is usually measured by precision, recall, and F-measure. Precision is the fraction of relevant retrieved instances, while recall is the fraction of retrieved relevant instances. F-measure also used to fairly treat recall and precision which is brought recall and precision into a single measure. Therefore precision, recall, and F-measure had been used based on an understanding and measure of relevance.

1.5 Scope of the study

The study's focus is on sentiment analysis of public news from the supplied media, with an emphasis on opinion mining for polarization into positive, negative, and neutral categories. It excludes both subjective and objective categorization. Only grammatically accurate sentiment texts are used in this study, which analyzes and categorizes sentiment texts written in Amharic. And solely analyze sentiment at the Aspect level. As a result, this study is unconcerned with grammatically poor texts, slang, thoughts represented by an image/picture, audio, video, or other emotional symbols. Furthermore, words or phrases with indirect or concealed meanings, such as an idiom, are not included in this study. One of the most difficult problems in machine learning and classification is domain independence. This approach is only applicable to the Amharic News domain.

1.6 Significance of the study

Knowing what other people think is a deciding aspect for logical decision making in today's commercial and political scenarios. Prior to making decisions, it will be able to examine the emotion of a large amount of data acquired automatically. This study can assist the Amhara Media Corporation organizations in improving their News services in the future, in addition to being an academic exercise to complete the program's requirements. The findings of the study can be used to help construct a full-fledged opinion that can be employed in a mining system for Amharic or any other Ethiopian language. The outcome of the study can also be used as input data for recommender and opinion retrieval/search systems, which is a significant benefit.

1.7 Organization of the Research

The following is the rest of the thesis. The second chapter contains reviews of various types of literature on opinion mining, as well as methodologies and machine learning and deep Learning techniques. The third chapter covers the study's general approach, including corpus preparation and preprocessing, system design, feature selection methods, classification strategies, and performance measurement. The experimental results and discoveries of how these tests and procedures are implemented are discussed in Chapter Four. Finally, Chapter 5 discusses the conclusion as well as the featured work.

Chapter Two

2. Literature Review

In this section, review of existing literature on sentiment analysis, and the application of various techniques for sentiment analysis are done. A description is made employment of machine learning and Deep Learning Techniques to determine sentiment in social media opinion using Amharic news were described in detail.

2.1 Amharic Language

Amharic is the official working language of Federal Government of Ethiopia. Within Semitic language family, it has the greatest number of speakers next to Arabic language. Tefera (2005) stated that Amharic is one of the languages with its own writing system. Many people use it as first language (mother tongue) for communication.

2.2 Nature of Amharic Language

Amharic is a Semitic language spoken across Ethiopia. It has a lot of inflection and dialect diversity. It is the second most spoken Semitic language in the world (after Arabic) and one of the five most spoken languages in the world today. Despite the comparatively significant number of speakers, it is still a minority language on the African continent. Language for which only a few computational linguistic resources have been produced, and for which there are few computational linguistic resources available. Has been accomplished in terms of developing higher-level Internet or computer-based apps that are helpful for those who only speak Amharic have this option.

Ethiopia is Africa's third most populous country, with over 80 languages spoken there. Three of these are the most common: Oromo, a Cushitic language spoken in the country's south and central regions and written in the Latin alphabet; Tigrigna, spoken in the north and neighboring Eritrea; and Amharic, spoken throughout the country but primarily in the eastern, western, and central regions. Semitic languages include Amharic and Tigrinya. In contrast to many other Semitic languages, both languages have their own script that is written horizontally and left-to-right.

Amharic is the second most widely spoken Semitic language in the world (after Arabic). It is now the second in the country. Ethiopia is governed by a 1993 Constitution. Divided into nine mostly autonomous sections, each with its own nationality and language However, Amharic is the official language of Ethiopia and was for a long time the official language of Ethiopia. In primary and secondary schools across the country, the major literal language and medium of instruction is higher education is conducted in English across the country. On his thesis work Gebremeskel (2010) depicted the majority of Amharic and Tigrinya speakers are Orthodox Christians, with the languages sharing roots with the Coptic Church's ecclesiastic Ge‘ez. Written Ge‘ez dates from the 4th century A.D. at the earliest.

The characters in later versions of the script represent consonant-vowel (CV) phoneme pairs, but the initial versions only had consonants. Each syllable pattern in current written Amharic has seven alternative forms (called orders), which correspond to the seven vowel sounds. The first order is the most basic; subsequent orders are formed from it through more or less regular adjustments that indicate the various vowels.

2.3 Amharic Language Word Classification

Yimam et al. (2021) uses the morphology and place of the word in an Amharic phrase as criteria for categorizing Amharic words into five fundamental groups.ስም (noun), ግስ(verb), ቅፅል(adjective), ተውሳከግስ(Adverb), and መስተዋድድ(preposition) are the five categories (Aragaw, 2015).

Noun: If a word can be pluralized by adding the suffix ኦች / ዎች (owch) and used to nominate something like a person or an animal, it is classed as a noun.

In a sentence, it is the subject. Pronouns, which were formerly regarded an autonomous category by linguists, have been grouped with nouns after taking into account the distinctive nature of the language, as previous linguists just adopted the English language structure for Amharic. The following are some of the pronouns in Amharic ይህ ,ያ , እሱ, እስዋ , እኔ , አንተ ,አንች …; quantitative specifiers, which includes አንድ, አንዳንድ, ብ□, ጥቂት, በጣም …; and possession specifiers such as የእኔ , የአንተ , የእሱ.

Verb: any word which can be placed at the end of a sentence and which can accept suffixes as /ህ/,/ሁ/,/ሽ /, etc. which is used to indicate masculine, feminine, and plurality is classified as a verb. For example in ―አበበአንበሳገደሇ‖ ―ገደሇ‖ is a verb since it appears at the end of the sentence.

Adjective: is a word that comes before a noun and add some kind of qualification to the noun. But every word that comes before a noun is not an adjective. For it to be an adjective it should also satisfy the condition when the word ―በጣም‖ is added to it, it should be meaningful. For example ―ትሌቅበግ‖ in this example ―ትሌቅ‖ is an adjective to check it really is an adjective adding the word ―በጣም‖ before the adjective if it is meaningful it is an adjective if not it isn‘t an adjective. In this case it is meaningful and ―ትሌቅ‖ is an adjective. Adverb: a word that qualifies the verb by adding extra idea from time, place and situations point of view. The following are adverbs in Amharic ትናንት, ገና ,ዙሬ , ቶል , ምንኛ , ክፉኛ , እንደገና , ጅሌኛ and ግምኛ .

Preposition: a word that doesn't take any kind of suffix and prefix, that can‘t be used to create other words and which doesn‘t have meaning by itself but can represent different adverbial roles when used with nouns. The different prepositions include ከ፣ሇ፣ወደ፣ስሇ፣እንደ፣ወዘተ.

2.4 Punctuation Marks and Numerals in Amharic Language

Distinct punctuation signs are used for different purposes in Amharic. A colon (two dots) was used to separate two words in ancient scripture. The two dots are now connected. Whitespace is used in their place. The end of a statement is indicated by four dots (አራትነጥብ።).While ነጠሊሰረዛ (፣ or ፥) is used to divide lists or ideas in the same way that a comma is in English. The punctuation marks are used to separate sentences and used to separate words. Numbers in Amharic can be expressed using the Arabic number system's symbols, the Ethiopic number system's symbols, or the Arabic number system's words and symbols. The Arabic, Amharic, and alphanumeric representations of numerals are shown in the table below.

Amharic Numeral Representation

Table 1: Alphanumeric representation of Ethiopic Numbers

Arabic	Ethiopic	Alphanumeric	Arabic	Ethiopic	Alphanumeric
1	፩	አንድ	20	፳	ሃያ
2	፪	ሁለት	30	፴	ሰላሳ
3	፫	ስስት	40	፵	አርባ
4	፬	አራት	50	፶	ሃምሳ
5	፭	አምስት	60	፰	ስልሳ
6	፮	ስድስት	70	፳	ሰባ
7	፯	ሰባት	80	፹	ሰማኒያ
8	፰	ስምንት	90	፺	ዘጠና
9	፱	ዘጠኝ	100	፻	መቶ
10	፲	አስር	1000	፼	አንድሺ

2.4.1 Characteristics of Amharic Writing

The properties of Amharic writing system are presented as follow to present comments from collected Facebook page to required data format.

Character Redundancy: Amharic uses the entire Geez alphabet in its writing system. It then added some additional symbols to represent certain other sounds that were not represented by the Geez alphabet's characters. Gebremeskel (2010) wrote that the Amharic letter has duplicate characters as a result of this haphazard borrowing from Geez.

Spelling variants of a term would increase the number of words describing a text unnecessarily, reducing the efficiency and accuracy of sentiment analysis classifiers, as well as any NLP system classifier. As a result, while processing Amharic documents for sentiment classification, word variations (spelling differences) caused by inconsistent use of superfluous letters should be normalized. The multiple forms of a character that have the same sound are transformed to one common form during the pre-processing step of Amharic documents in this work.

2.4.2 Amharic Text Sentiment Analysis

For impoverished languages such as Amharic, there are two fundamental impediments to growth in language processing. To begin with, the diversity of the languages may need the invention of new strategies. Second, the limited availability of existing resources and tools makes building and testing new ones more complex and time-consuming. Amharic does not have a large corpus. For Natural Language Processing experiments, the availability of labeled language resources, such as annotated corpora and domain-specific labeled language resources is crucial. Due to a shortage of resources, manual verification and annotation of electronic text content is typically a requirement for the development of NLP systems. Another concern with the Amharic language is its lack of linguistic encoding capabilities. Currently, the language does not enable language encoding (or does so insufficiently). Finally, the morphology of Amharic is quite variable. Nonetheless, it is understudied linguistically. Even linguistic research on a language like Arabic, which is spoken or understood by over a billion people, is dwarfed by the amount of linguistic analysis available for English. Amharic has a far more difficult situation. According to Worku (2009) sentiment analysis of Amharic text continues to be a challenge.

2.5 Types of Opinions

According to Patel et al. (2015) opinions can be classified in to two major categories. Those are regular opinions and comparative opinions.

2.5.1 Regular Opinions

Regular opinions can be direct or indirect. For instance, the display quality is excellent (የማሳያ ጥራት ጥርት ያለ ነው). It directly refers to a facet of "image quality"(የስዕል ጥራት)here, implying a direct judgment. "After using the cream, my skin completely broke out." In this case, the entity "cream" is used to indicate a negative impression on the aspect "skin." Direct opinions are easier to comprehend. To determine the polarity of indirect opinions, one must first understand the data source domain. For instance, in the preceding example of a cream review, the skin broke out and expressed a bad opinion Patel et al. (2015).

2.5.2 Comparative Opinions

Unlike Regular opinions, they can have differing viewpoints about the same thing. For example, the S6's CPU speed and screen quality are superior to the iPhone 6, yet the iPhone 6's metal frame is more appealing than the S6's(የ S6 ሲፒዩ ፍጥነት እና የማያ ገጽ ጥራት ከ iPhone 6 ይበልጣል፣

ሆኖም ግንየiPhone6 የብረት ክፈፍ ከS6ዎቹ የበለጠ የሚስብ ነው::). In this sentence, S6 has two favorable and one negative view. There are two types of comparative opinions: explicit and implicit. Explicit comparisons are easier to assess because they communicate a single, positive or negative opinion. One Plus, for example, has a faster processor than Yureka. The aspect of CPU speed is directly compared here, and the entity One Plus comes out on top. The term "implicit comparison" refers to an objective statement in which opinions are expressed in an indirect and oblique manner. For example, program x takes longer to execute than program y. As Patel et al. (2015) discussed, a longer execution time indicates that program x's performance is inferior to program y's. As a result, a bad view about program x is stated.

2.6 Sentiment Analysis

Sentiment analysis, often known as opinion mining, is a natural language processing technique for sorting data into positive, negative, and neutral categories. As Aragaw (2015) stated sentiment analysis is a type of text mining that identifies and extracts subjective information from source material, allowing a firm to better understand the social sentiment surrounding its brand, product, or service.

There are a lot of social media platforms these days, and a lot of content is produced in fractions of seconds all over the world, provoking a wide range of reactions from people. Sentiment analysis divides a client's feelings and emotions into three categories: positive, negative, and neutral. Polarity classification is a way of recognizing or classifying a given sentiment as positive, negative, or neutral based on the weight attributed to the sentiment and the document's circumstances (Leetaru, 2020).

The film and music industries, sports and recreation, huge enterprises to maintain staff morale, product vendor companies, and social media are all using sentiment analysis. In social media, sentiment analysis is important for detecting the sentiments and feelings of people who comment on a given person's or organization's post. In addition to a simple count of mentions or comments, sentiment analysis integrates emotions and opinions. It comprises obtaining and analyzing data from your brand's social media posts (Steinberger et al., 2017).

Leetaru (2020) presented that sentiment analysis can be done on one of the following.

• **Document level:** The entire document is classified as good, negative, or neutral at this level.

• **Sentence level:** The entire sentence is classified as good, negative, or neutral at this level.

• **Aspect level:** For each feature present in the document/sentence, the entire document/sentence is classified as positive, negative, or neutral at this level.

The polarity of a document or a sentence simply classifies the document or sentence as a whole; it does not distinguish the features included within it. For example, simply because a document's polarity is positive or negative doesn't indicate it has a positive or negative perspective on each of its aspects. At the aspect level, opinion mining is utilized to find out what individuals think about certain topics.

Two type's aspects are found in user reviews explicit and implicit. Explicit aspects are nouns and noun phrases that are easily detected in the reviews. For example, "This phone has excellent speech quality," when voice quality is an explicit characteristic that can be seen clearly in the statement. Explicit aspects are nouns and noun phrases that are not easily detected in the reviews. Implicit aspects are nouns and noun phrases that are not easily identified in the reviews. For example, in the statement "This phone does not fit in my pockets ይህ ስልክ በኪሴ ውስጥ አይገጥምም," the implicit aspect "does not fit in my pockets, በኪሴ ውስጥ አይገጥምም" denotes the aspect size that cannot be stated recognized in the sentence.

In this study, aspect level sentiment analysis is proposed. An Aspect based Opinion Mining system that extracts characteristics and opinions from phrases and determines whether the provided sentences are positive, negative, or neutral for each feature. Negativity is likewise dealt with by the mechanism. As Aragaw (2015) The semantic orientation of the sentences is determined using an supervised approach using a Hybrid deep Learning technique.

2.7 Approaches to Sentiment Classification

Sentiment Classification techniques can be roughly divided into machine learning approach,

lexicon based approach and hybrid approach (D'Andrea et al., (2015). The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The Lexicon-based Approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. It is divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The hybrid Approach combines both approaches and is very common with sentiment lexicons playing a key role in the majority of methods. The various approaches and the most popular algorithms of are illustrated in Figure below.

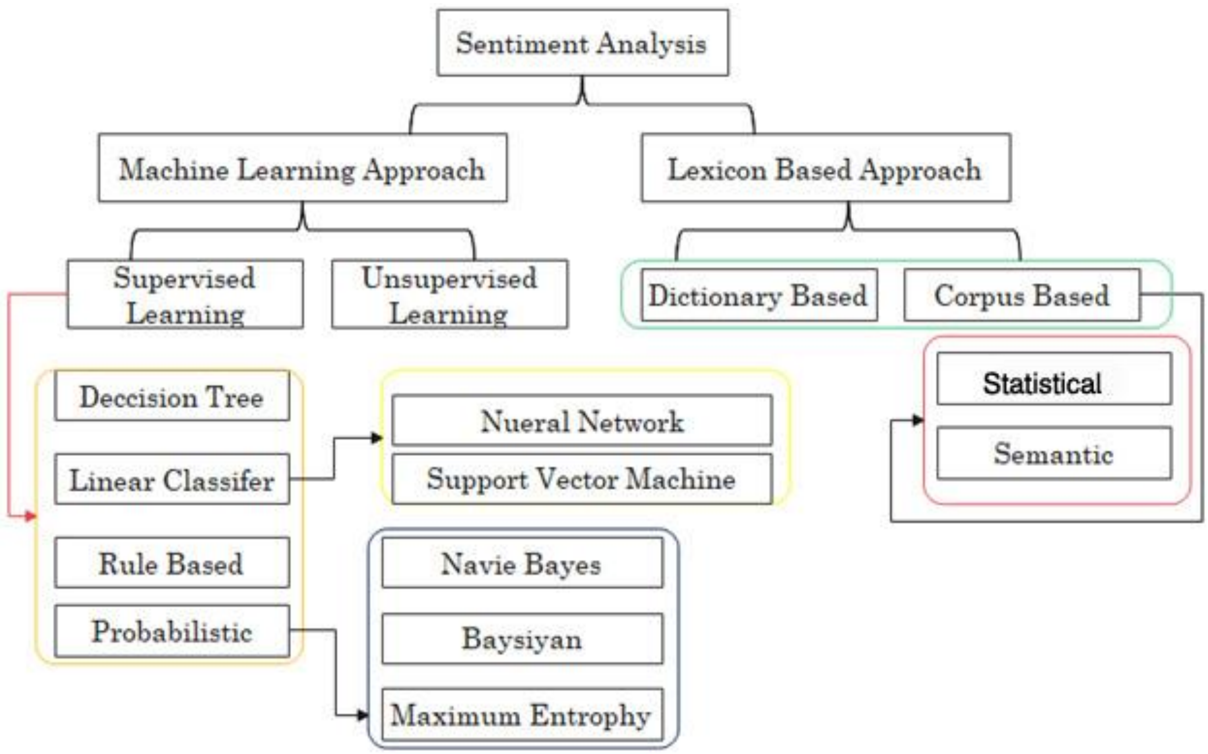

Figure 1: Sentiment Classification Techniques (Sharma et al., p.79)

Text categorization approaches based on machine learning can be classified into two categories: supervised and unsupervised learning methods. A high number of labeled samples are used in the supervised approaches. Documents for training when finding labeled data is challenging, unsupervised methods are used. Documents for training the lexicon-based method rely on locating the opinion lexicon that will be used to analyze the text. This strategy employs two methods. The dictionary-based strategy relies on obtaining opinion seed words and then looking for synonyms and antonyms in a lexicon. The corpus-based approach begins with a seed list of opinion words, and then finds other opinion words in a large corpus to help in finding opinion words with context specific orientations. This could be done by using statistical or semantic methods. There is a brief explanation of both approaches 'algorithms and related articles in the next subsections.

2.7.1 Machine Learning

2.7.1.1 Artificial Neural Network

In any case, a neural system has an input layer and an output layer. Between the input and output levels, some system topologies may have multiple hidden layers. There must be at least one node in each stratum. Each input neuron is linked to each output neuron in the next layer. In neural networks, two stages of operation, testing and training, are always present. During the training phase, the neural system uses the prepared dataset as input and modifies the association loads to achieve the desired affiliation or characterization. During the testing stage, the neural systems is tried with the testing dataset (unique in relation to preparing dataset) to recover comparing yields dependent on the information found from the training stage (Tamiru, 2018). An input layer, one or more hidden layers, and a single output layer make up a neural network. Each layer might have a varied number of neurons and is fully coupled to the layer above it. The network architecture of neural networks influences their behavior. The architecture of a network can be described in terms of the number of neurons, layers, and types of connections between layers. The raw vector input is used for the input layer. The output (activation) of the previous layer's neurons is the input to neurons in subsequent levels. The connection weights and the activation function type influence data as it flows through the network in a feed forward fashion.

Input layer: They receive an object's data feature first, then process the object (data) before sending their first output to the first hidden layer. The number of neurons in an input layer is usually the same as the network's input feature. One or more hidden layers follow the input layers.

Hidden layer: The output from the input layer is processed and sent to the next hidden layer. The data is processed until it reaches the final output layer, where the output values determine or recognize the object or input data. The learned information collected from the raw training data is encoded by the weight values on the connections between the layers. The ability of neural networks to model nonlinear functions is dependent on hidden layers.

Output layer: output (prediction or classification) of our model is answered from the output layer. The output layer gives us an output based on the input from the input layer. Depending on the setup of the neural network, the final output may be a real valued output (regression) or a set of probabilities (classification). This is controlled by the type of activation function we use on the neurons in the output layer.

Connections between layers: The connections between layers in a fully linked feed-forward network are the outgoing connections from all neurons in the previous layer to all neurons in the next layer. With the back propagation learning process, these weights are gradually modified as the system discovers the optimal solution. Each node is fully connected to all input nodes and computes the weighted sum of those nodes. Its structure is one-dimensional. It aids in the classification of input patterns by utilizing high-level features retrieved by the previous layer.

2.7.1.2 Deep Learning Approach

2.7.1.2.1 Convolutional Neural Network (CNN)

Convolutional neural networks are specialized kinds of neural networks designed to process and analyze data. CNN is a class of deep learning with invariance built into a multi-layer network structure which is more suitable for the recognition of data Zhang (2021). It takes an input, assigns weights (and biases) to various words in data to differentiate them from one another. Compared to other classification algorithms CNN requires much lower pre-processing aspects.

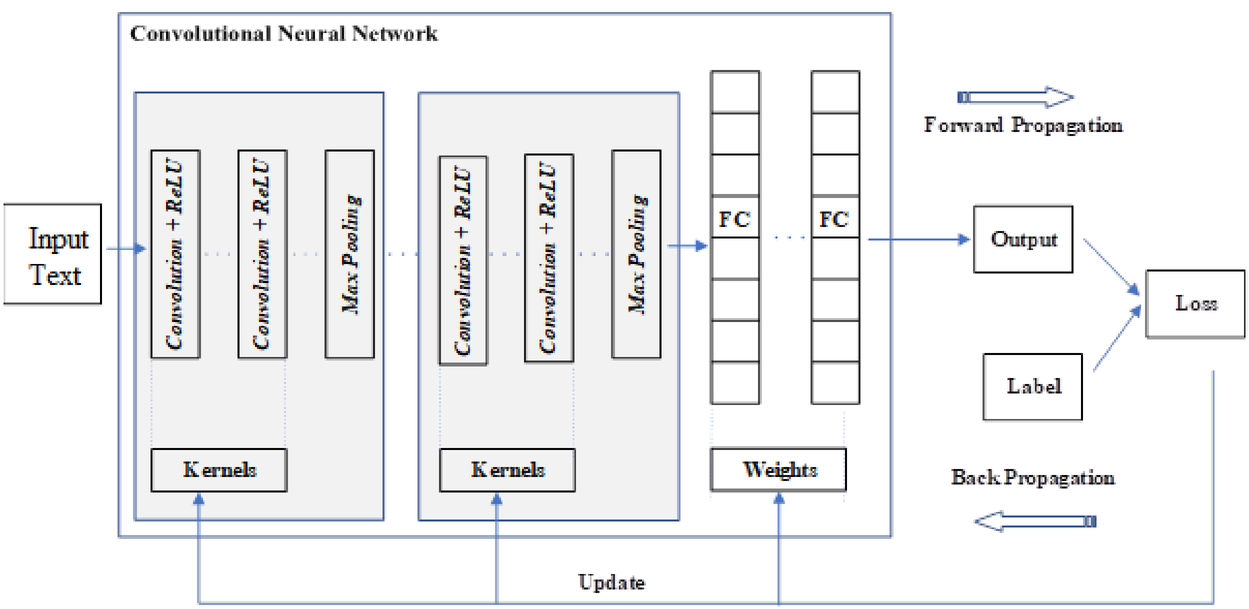

Figure 2: CNN Architecture and Training Process

In traditional neural networks, each output interacts with each input. Therefore, back-propagation learning leads to a gradient that either vanishes or explodes when propagated through many layers which makes deeper network architectures impossible to train to overcome this problem CNN typically has sparse connections by using filters with lower dimensions than the input data. Besides, algorithms using CNN often implement a greedy approach, which means that the network is trained layer-wise and preserves the complexity of each layer by freezing its weights once trained and sending its output as an input to the next layer.

2.7.1.2.2 Building blocks of CNN architecture

A convolutional neural network (CNN) contains one or more convolutional layers, pooling or fully connected, and activation function. Convolutional layers use a convolution operation to the input passing the result to the next layer. This operation allows the network to be deeper with much fewer parameters. Convolutional neural networks show outstanding results in image and speech applications, however, Yoon Kim (2014) in Convolutional Neural Networks for Sentence Classification describes the process and the results of text classification tasks using CNNs. In Text Understanding from Scratch, Xiang Zhang and Yann LeCun (2016) demonstrate that CNNs can achieve outstanding performance without the knowledge of words, phrases, sentences and any other syntactic or semantic structures with regards to a human language. As depicted in Figure 2, typical architecture consists of repetitions of a stack of several convolution layers followed by a pooling layer, and one or more fully connected layers. As the input reaches each layer, the content of the input is transformed in some way and transmitted to the next layer. The procedure which transforms the input data into output through these layers is called forward propagation.

2.7.1.2.3 Convolution Layer

Convolution layer is the core building block of CNN which performs a feature extraction to detect a pattern from a given input volume with the help of a combination of linear and nonlinear operations, i.e., convolution operation and activation function. It consists of filters with their own width and height. These filters are smaller in terms of their spatial dimensions and convolved throughout the entire input.

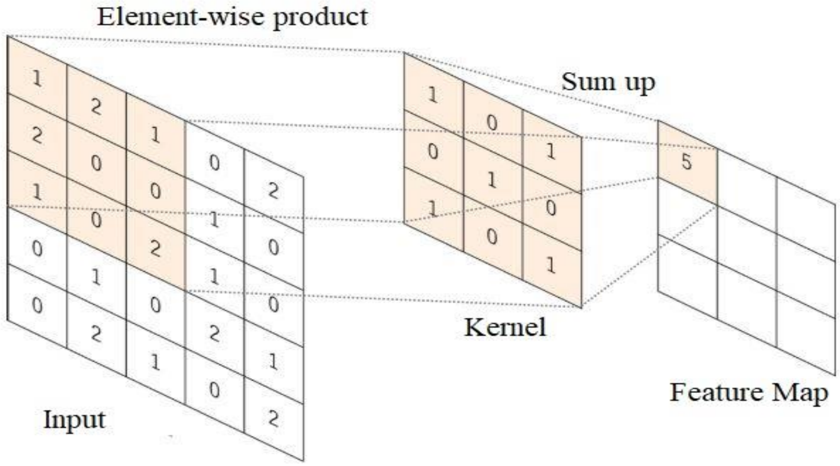

Figure 3: An Example of Convolution Operation as Adapted from (Yamashita et al., 2018)

Convolution is a specialized type of linear operation which involves an element-wise multiplication of a filter and an input matrix followed by a sum at each location of a feature map as exemplified in figure 3.

A 3x3 sized kernel is applied across the input matrix, and an element-wise product between each element of the kernel and the input tensor is calculated at each location and summed to obtain the output value in the corresponding position of the output tensor, called a feature map.

2.7.1.2.4 Hyper-parameters

Hyper-parameters are variables that need to be set during the configuration before the training process starts. These variables are external to the model which cannot be estimated from the given input data. Hyper-parameters determine the network structure (E.g., the number of hidden layers) and the way how the network is trained (E.g., learning rate) (Aszemi and Dominic, 2019). Hyperparameters that determine the network structure include the kernel size and type, stride, padding, hidden layer, and activation functions. Whereas those which determine the network trained are, learning rate, dropout, momentum, number of epochs, and batch size.

Table 2: List of Parameters and Hyper-Parameters in CNN

Feature learning metrics	**Parameters**	**Hyper-parameters**
Convolutional layer	Filters	Filter size, number of filters, stride, padding, activation function
Pooling layer	None	Pooling method, filter size, stride, padding
Fully connected layer	Weights	Number of weights, activation function
Others	…	Model architecture, optimizer, learning rate, loss function, mini-batch size, epochs, regularization, weight initialization, dataset splitting, data augmentation

A quick review of what those parameters mean to the network is provided below in short as described in (Radhakrishnan, 2017).

- Filter size and type – filter size is just the size of the filter, and filter type means to determine the values of the actual filter.
- Hidden layers – are the layers between the input and output layers. Many hidden units within a layer with regularization techniques can increase accuracy. A smaller number of units may cause under fitting.
- Dropout – is a regularization technique to face an over fitting problem by randomly selecting neurons that are ignored (inactivated) during training.
- Learning rate – defines how quickly a network updates its learnable parameters at the end of each batch.
- Momentum – helps to know the direction of the next step with the knowledge of the previous steps.
- Epochs – the iterations of the entire training dataset shown to the network while training.
- Batch size – the number of sub-samples given to the network before the parameters are updated.

2.7.1.2.5 Activation Function

An activation function is a non-linear function that is added into a convolutional neural network to help the network learn complex patterns in the data. It is biologically inspired by activity in our brains, where different neurons fire or are activated by different stimuli. The activation function presents non-linearities to the neural network, which are required for multi-layer networks to detect nonlinear features. Therefore, it increases the expression ability of a neural network model, which can make the deep neural network truly have the significance of artificial intelligence by activating features of neurons and solve non-linear problems (Gu et al., (2018).

An activation function is mainly applied to the neural network by taking the output signal from the previous neuron then converting it into some form and deciding what is to be fired to the next neuron of the network (Wang et al., (2020). It does some kind of operation on the inputted value and transforms the value between some lower limit and upper limit number. Typical activation functions that are widely used are sigmoid, tanh, and ReLu as discussed in Figure 4.

Sigmoid

A sigmoid function which is ranged between 0 and 1, is useful when binary classification is required to be done. It takes an input and transforms a very negative number to 0 and a very positive number to 1.

Tanh

In the case of hyperbolic tangent function (commonly referred to as tanh) the range comes to be between -1 and 1. Tanh is an updated version of the sigmoid function with a better convergence rate.

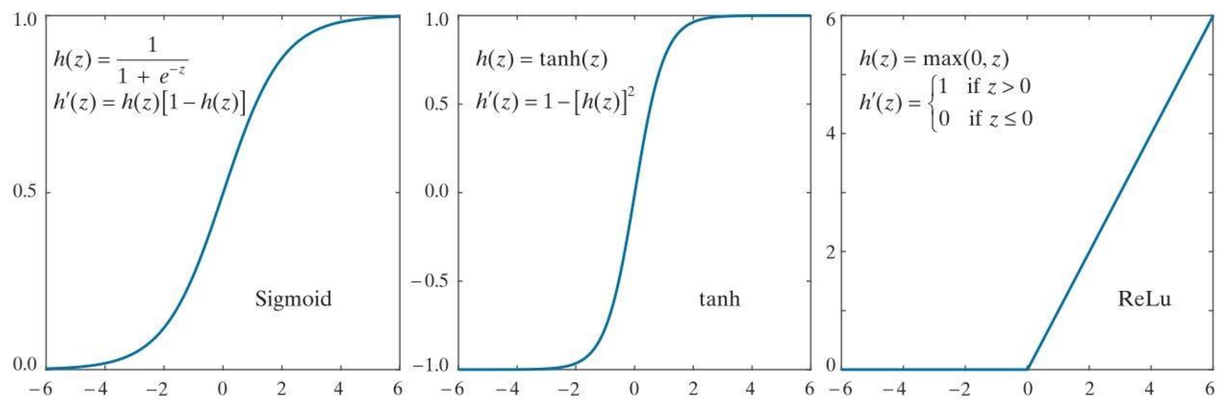

Figure 4: Activation Functions; sigmoid function; Hyperbolic tangent (tanh) function and Rectifier Linear unit (Relu)

ReLU

Rectified linear unit (ReLU) is a widely used activation function in CNN which forces the output to be zero if the input value is less than or equal to zero. Otherwise, it will make the output value equal to the input value. The disadvantage of the ReLu function is that if the value of the input is positive, it does not change anything; simply it forwards the value to the next layer node. However, the major advantage considered is that if the ReLU function receives a negative value, it stops the neuron from stimulating by making it 0 and hence preventing the neuron from firing.

SoftMax Function

The sigmoid function is useful for binary classification whereas SoftMax is used for problems involving a multiclass classification. It produces values in the range of 0 and 1 therefore it is used as the final layer in classification models.

Pooling Layer

The pooling layer is a down sampling layer that is typically added to CNN's following individual convolutional layers. Each convolutional layer has some number of filters associated with them with a specified dimension and these filters convolve on the text input. When the convolution operation is performed the resulting output is generated. This output is a matrix with the values that were computed during the convolutions that occurred on the text. As stated in Gu et al., (2016), there are two types of pooling layers namely max-pooling and average-pooling.

Max-pooling is the most common type of pooling method. It reduces the resolution of a given output of a convolutional layer by looking at larger areas of the text at a time going forward which reduces the number of parameters in the network and consequently reduces computational load. It assumed that the higher valued inputs as being the ones that are the most activated. Therefore, max-pooling picks out the most activated inputs and preserves these high values forward while discarding the lower valued inputs that are not as activated.

In the case of average-pooling, the average value of all the inputs in the batch is selected by considering the specified filter and stride size.

Fully Connected Layers

Fully connected layers, also known as a dense layer, connect every neuron in one layer to every neuron in another layer like a traditional neural network. The flattened matrix which is a one-dimensional array of numbers (or vector), goes through a fully connected layer to classify the texts. Once the features are extracted by the convolution layers and down sampled by the pooling layers, they are mapped by fully connected layers to generate the final outputs of the network. This might be the probabilities for each class in classification tasks. Mostly a fully connected layer typically has the same number of output nodes as the number of classes for a given task. Each fully connected layer is followed by a non-linear function, such as ReLu (Yamashita et al., 2018).

A fully connected layer includes a lot of parameters that need complex computational during a training period. This is considered as a major drawback of this layer as discussed in (Albawi and Mohammed, 2017).

2.8 Long short-term memory (LSTM)

This networks will solve the main shortcoming of RNN, as it has a special cell architecture that helps it to store information for long time (Shewalkar et al., 2019). LSTM contains three gates: an input gate, an output gate and a forget gate. At each iteration, the three gates try to remember when and how much the information in the memory cell should be updated.

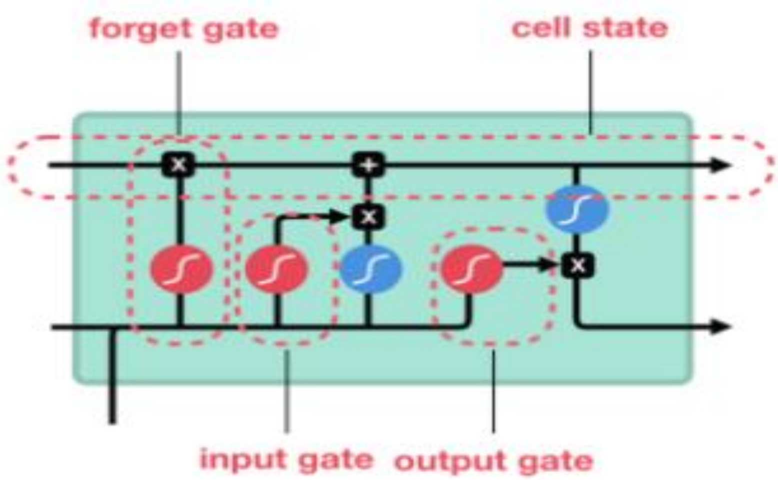

Figure 5: Structure of LSTM

A. Input gate

The input gate is used to update the cell state by passing the previous hidden state and current input into an activation function. The activation function will decide which value will be updated and determine the important value for the next step.

B. Output gate

It is the last gate that decides what the next hidden state should be. The hidden state contains previous input information, pass the previous hidden state and the current input are passed to sigmoid function. Then the newly modified cell state is passed to the tanh function. Finally, the tanh output is multiplied with the sigmoid output to decide what information the hidden state should carry.

C. Forget gate

This gate is used to decide what information should be kept or forgotten from the previous state and thus determines what the next hidden state should be.

2.9 Gated Recurrent Unit (GRU)

These networks are similar to LSTM, and have gated units like LSTM that are used for controlling the flow information inside the units. It removes the cell state found in LSTM networks and used hidden state to transfer information. It has two gates, the reset gate and update gate.

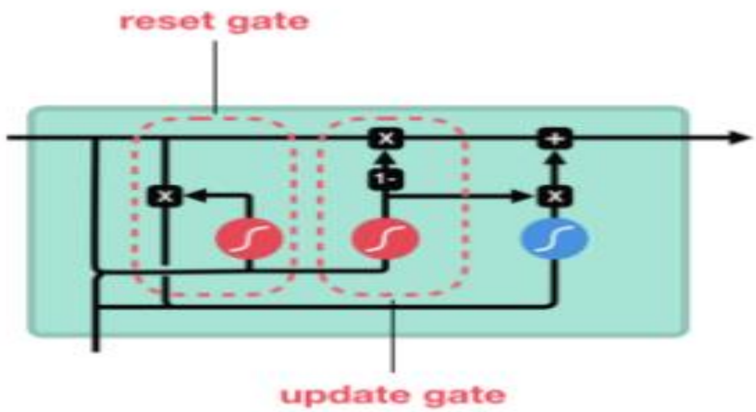

Figure 6:Structure of GRU

A. Update gate

It acts similarly to forget gate and input gate of LSTM network and determines which information to forget and what new information to add to the next step.

B. Reset gate

This gate is used to determine how much information from previous states should be forgotten. Since GRU has fewer gates, it is likely to have fewer operations than LSTM networks which makes it faster to train. Unlike LSTM, GRU networks have fewer parameters and separate memory cell for controlling information flow. Additionally, GRU do not have an output gate thereby it exposes it's full content (Shewalkar et al., 2019). Even if the operation inside GRU is different, the workflow of GRU is similar to RNN. Comparatively, Both LSTM and GRU are proved to be equally efficient and are capable of handling long term dependencies between sequential inputs (Shewalkar et al., 2019).

2.10 Related Work

For more than a decade, a lot of work has been done in the field of opinion mining or sentiment analysis on English texts. Most previous work in sentiment analysis has relied on lexicon-based approaches to identify the sentiment of texts using dictionaries that define the sentiment-polarity of words and simple linguistic patterns to classify the sentiment of texts. In various studies conducted in English and other languages, the procedures outlined in section 2.3 have yielded quite satisfactory results. When it comes to identifying the sentiment of Amharic writing, however, there are several obstacles to overcome.

2.11 Sentiment Analysis in Other Languages

Pang and lee Sharma and Dey (2012) pioneered the field of sentiment analysis. They worked on a sentiment polarity classification problem, employing Naïve Bayes and support vector machines (SVM) to choose between positive and negative labels for each text document, which was represented as a bag of words with weights for word presence. They also explored using negation, word locations, and part of speech (POS) information without much success, and discovered that many strategies that aid with subject categorization actually hurt sentiment polarity accuracy. The trials were conducted on a collection of 2000 movie reviews (1000 positive and 1000 negative) acquired from the internet, with no specific information concerning polarity, such as ranks, scores, or star ratings. The data set was made freely available to the public, and it has since become the de-facto standard for training and evaluation in most following studies.

Kaushik et al. (2014) used lexicon based technique i.e. a dictionary of sentiment bearing words along with their polarities is used to classify the text into positive, negative or neutral opinion. The dictionary is domain specific i.e. the polarities of the words in the dictionary are set according to a specific domain e.g. book reviews, political blogs. The dictionary used in their approach is made for movie review domain. It contains more than 30 different emoticons along with their polarities, the strength of the polarity of every word, and also various negations and blind negation words so that they can be identified in the sentence. In the research, twitter data is categorized according to polarity. Sentiment calculation is done for every tweet and a polarity score is given to it. The polarity score is calculated by using the algorithm _ALGO_SENTICAL‘. They implemented the algorithm using a sandboxed version of Hadoop. Then the data is refined and the algorithm is implemented upon it with the help of HiveQL. The purpose of their research was to device a method that can quickly compute the sentiments of huge data sets without compromising too much with accuracy. Their approach has performed very well in terms of speed. It took 14.8 Seconds to analyze the sentiments behind 674412 tweets. The algorithm also performed very well in terms of accuracy. Tweets were analyzed with an accuracy of 73.5 %. However, this work was implemented on a single node sandboxed configuration while it could perform much better in a multimode enterprise level configuration.

Alaa El-Halees (2011) have proposed a combined approach which aims at mining opinions from Arabic documents. In their approach they used three methods at sequence: First, lexicon-based method is used which classifies some documents. The classified documents used as training set for maximum entropy model which subsequently classified some other documents. After that, k-nearest model is used to classify the rest of the documents. They have done experiments with 1143 posts containing 8793 Arabic Statements. Their system achieved an accuracy of 80.29%. The accuracy almost went from 50% using one method, 60% using two method and 80% using three methods. They also claimed that the experimental results further show recall and precision of positive documents are better than the negative one.

As Zhang et al. (2006) discussed Combining Lexicon based and Learning-based Methods for Twitter Sentiment Analysis propose a new entity level sentiment analysis method for Twitter. The method first adopts a lexicon based approach to perform entity-level sentiment analysis. But they agree that this method can give high precision, but low recall. To improve recall, they used additional tweets that are likely to be opinionated and identified automatically by exploiting the information in the result of the lexicon-based method. A classifier is then trained to assign polarities to the entities in the newly identified tweets. Instead of being labeled manually, the training examples are given by the lexicon-based approach. The whole process has no manual labeling. The approach they propose is an unsupervised method except for the initial opinion lexicon, which is publicly available. The researchers used accuracy to evaluate the whole classification performance of each method with three classes, positive, negative and neutral. For positive and negative sentiments on entities, they employ the standard evaluation measures of precision, recall and F-score. Their experimental results show that the proposed method dramatically improves the recall and the F1-score, and outperforms the state-of-the-art baselines.

Lawrence Palanisamy et al. (2013), considers only the case of two sources of information, i.e., labeled examples in the target domains and background lexical knowledge. However, Pooling Multinomials is a general framework that can be used to combine multiple multinomial models these could be derived from training data from different related domains or different sources of background knowledge. Their empirical results show that their approach of Pooling

Multinomials performs better than various baselines that use lexical knowledge and labeled data in isolation, and an alternative approach to using background knowledge in a semi supervised setting.

2.12 Sentiment Analysis in Amharic language

Most research efforts in the area of opinion mining deal with English texts and little work with Amharic text. In publications, we could not find any work that mentioned the idea of Amharic opinion mining (as per the attempts of the researcher to find publication on the area). However, we found three post graduate works done on Amharic text sentiment analysis and revised them as follows. As to the researcher ‘s knowledge, pioneering the work on sentiment analysis of Amharic texts is Gebremeskel‘s (2010) sentiment mining model for opinionated Amharic texts. Basically, the study uses sentiment and subjective lexicon of terms for classifying reviews based on how many positive and negative terms are present in the subjective textual document. This is based on a rule-based classifier where if there are more positive than negative terms then it is considered as positive; else, if there are more negative than positive terms then it is considered as negative. If there are equal numbers of positive and negative terms then the opinion is neutral.

But the research done by Gebremeskel has some limitation; first the system uses a user defined dictionary which contains only l164 opinion terms. Since the system considers only terms which is found in the dictionary, some important opinionated terms may be ignored which result in a wrong classification. The other limitation is that, the size of the dictionary is very small that does not incorporate all the subjective terms of the domain. Preparing the dictionary manually by itself is difficult and time consuming. And since the dictionary was prepared manually, human generated error may occur and may let the performance of the system to decline. Another previous contribution in the area include the work by Alemu (2015) who attempt to apply opinion mining for discovering hidden knowledge from the opinionated Amharic textual documents in the movie review domain. They used a binary classification model using Naïve Bayes and decision tree algorithms to classify movie reviews as positive and negative. In their experiment Naïve Bayes showed a better performance relative to decision tree. Unigram features and a simple bag of words are used as a feature extractor for binary classification of the data. Before applying feature reduction, they come up with a performance of 81% accuracy for Naïve Bayes and 72% accuracy for decision tree. On their second experiment by using information gain as a feature reduction technique to choose the most informative words, they improved the performance of the algorithms by 11% and 9% for NB and decision tree respectively. From their experiment it is clear that the results are promising. However, because of the limited corpus of opinionated Amharic documents they collect they failed to cover the entire movie review domain and to adequately train the machine learning algorithms. Besides this, context valence shifters are not handled in the study.

Another work by Mengistu (2013) focuses on document level sentiment classification using supervised machine learning techniques on Amharic opinionated text. The study used a total amount of 576 (192 positive, 192 negative and 192 neutral sentences) training and testing data. The result shows the performance of the three machine learning algorithms (NB, MNB and SVM) using three feature selection methods. The feature selections are n-grams presence, ngrams-frequency and n-grams TF-IDF. In their report Naïve Bayes achieved 77.6% performance with uni-grams for each representation, Multinomial Naïve Bayes achieved 74.7% accuracy using uni-grams term frequency for each representation and SVM achieved 78.8% using n-grams term frequency representations. Based on their relative performance, of all the three algorithms and feature selection methods, Support Vector machine achieved the best result using uni-grams term frequency.

2.13 Summary of Related Work

Table 4: Summery of related Work

Author and Year	Research Title	Research methodology	Research Gap	Remark
Pang and Lee (2002)	Using very simple statistics for review search:	Naïve Bayes and support vector machines (SVM)	The data is de-facto standard for training and evaluation	Sentiment polarity classification
Kaushik et al. (2014)	A Scalable, Lexicon Based Technique for Sentiment Analysis	lexicon based technique	Analyzed tweet approve 73% accuracy	674,412 number of collected tweets and 74% accuracy
Gebremeskel (2010)	sentiment mining model for opinionated Amharic texts	rule-based lexicon classifier	Limited list of opinions and use user defined Dictionary	Use 306 dataset for the classification of sentiment values.
Alemu (2019)	Deep Learning Approach for Amharic Sentiment Analysis	Naïve Bayes and decision tree algorithms	Can't convert emojis to into numerical rather than use their mechanisms and difficult on the classification	1500 comments were collected and 1000 were used to train the model which is not enough to train deep learning Model
Mengistu (2013)	Document level sentiment Classification	Use NB.MNB and SVM	Use SVM with unigram approach	Classified documents with acceptable accuracy
Chilot (2019)	Public sentiment analysis on Amharic News	SVM and Naïve Bayes approaches were applied	Have collected only 1200 Dataset and apply only SVM and Bayes	SVM achieves87%,82%,84% F-measure precision and recall Metrices.

Wondwosen (2008)	A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts	Supervised Machine Learning Approach Naïve Bayes	The classification algorithm is selected due to Simplicity	Use 608 dataset collected from social media, product marketing and news.
Bestinat (2015)	Sentence level Amharic Sentiment Analysis Model: A Combined Approach	Machine learning and lexicon based combined approach	Indirect opinions have not been considered	Proposed hybrid approach performs 82.5% classification rate
Worku (2013)	Automatic Amharic text news classification A neural networks approach	Using Neural Network Approach	Complete stop word removal have not been applied and have effect in recognition Rate.	71.96 % average accuracy were achieved
Teshome (2019)	Sentence level opinion mining for Amharic language	Using machine Learning approach	The dataset have no well-defined categorization	Achieves 93 % recognition accuracy

2.14 Conclusion

Most of the research works were explored related to sentiment analysis using sentence level and document level. Among the related literatures reviewed for this thesis none are done on sentiment analysis at aspect level. There were also few works which also focus aspect level sentiment analysis but all still there is no work which uses Different deep learning Techniques. But still there are research gaps which are not included in their work and mainly our work is focused on this point. A scholar Chilot (2019) was show aspect level Amharic sentiment analysis using machine learning approach but still there are gaps in this work and recommends as future work. According to the above research reviews, in other languages like Chinese, India, etc. different researchers used Deep learning for sentiment analysis and achieves better classification rate. But in the above studies on Amharic sentiment analysis, lexicon based and traditional machine learning approaches are applied. But recently deep learning techniques are applied in sentiment analysis specially in foreign languages and achieved astonishing results (Ain et al., 2017; Dashtipour et al., 2021; Heikal et al., 2018; Ramadhani, 2017; L. Zhang et al., n.d.). Therefore, applying such deep learning techniques in Amharic sentiment analysis will improve the performance of sentiment classification models. In this study, different deep learning techniques are applied to develop Amharic sentiment analysis model.

Chapter Three

3. Designing Aspect Level Sentiment Analysis Model

3.1. Introduction

In this section, we will discuss the overall design of sentiment analysis using monitored Amharic news, provide the basic architecture and brief description of different tasks involved in the process are discussed in detail. The overall activity consists of data preprocessing, feature extraction and sentiment classification into a predefined class and classifier combination. Therefore, the basic activities that will be performed in the development process like dataset preparation, preprocessing, feature extraction, and classification will be described in the next consecutive sections.

3.2. Proposed Architecture

The proposed architecture shown in figure 8 below has three components that include preprocessing, feature extraction, classification. Each component is integrated and work together to discover sentiment polarity from a given user comment. The first component in the architecture is preprocessing that includes tasks such as removing punctuation, numbers, URLs and stop words from a given sentiment. Comments from users is first tokenized by using space between words as a separator, then patterns that match URL that represent web address of resources is removed. The next task is removing numbers and special symbols that appear together with comments of users. The other task performed as part of preprocessing is stop word removal. Stop words are common words and do not add importance in the meaning of the user comment. Therefore, it needs to be removed from the user comment. The next task after preprocessing is feature extraction for extracting relevant features from processed text. CNN in this case is used for learning important local features, and then fed to GRU so that their sequence is learned and passed to GRU for classification combined with Softmax layer. But before features are extracted from user comments, the textual comments should be converted into numerical format that is useful for neural networks. To do so, neural word embedding layer from keras is used to convert the text into fixed integer representation. Then the embedded text sequence is fed to the first layer of the convolutional neural network. This layer extracts locally relevant features from the input integer representation of user's comment. Useful features are extracted from the user comment by using sequence of convolutional filters. Features extracted by using convolutional filters are mapped into reduced representation by using MaxPooling operation. Then the pooled feature map is fed to GRU layer for classification. GRU layer learns the sequential relation that exist between words in the text sequence and then classifies the text sequence into corresponding sentiment polarity as ‘positive’ (‘አዎንታዊ’) or Negative (‘አሉታዊ’).

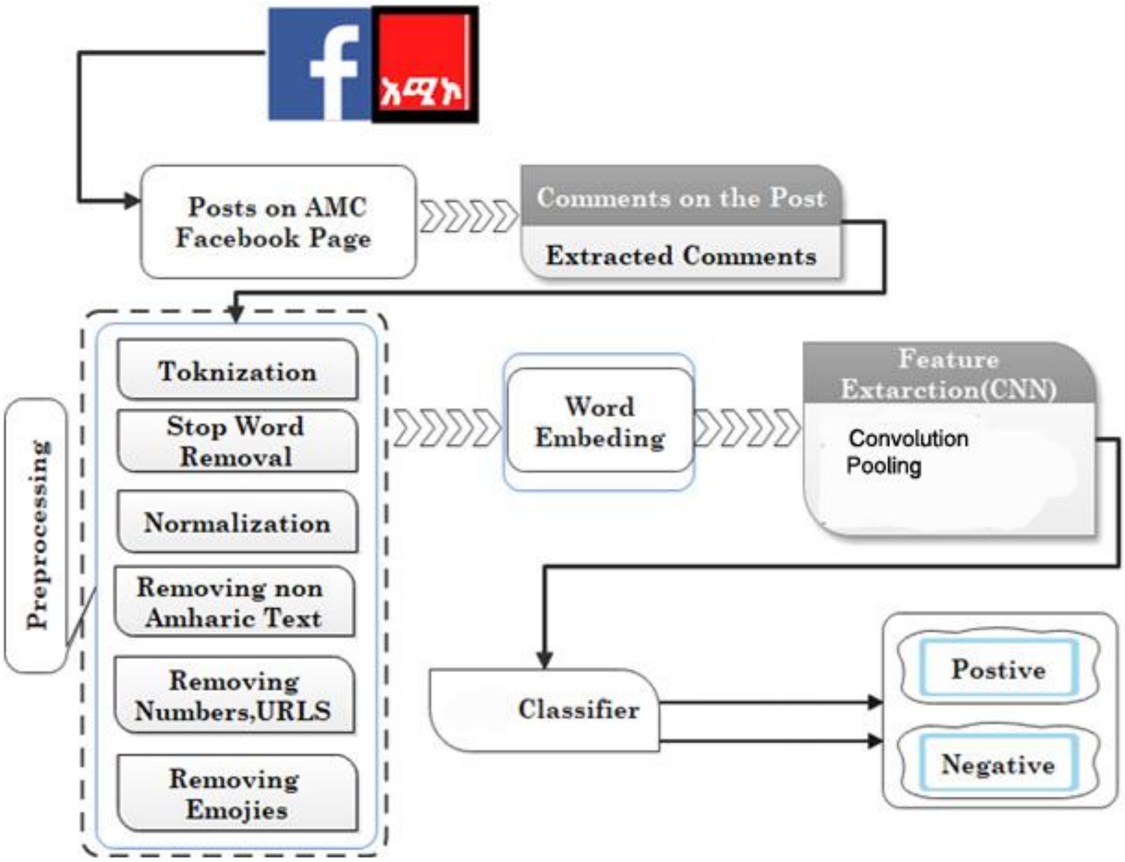

Figure 7: Proposed System Architecture

As shown in the architecture above, the textual data is processed by the preprocessing component and fed to the embedding layer. The embedding layer performs conversion of text data into numeric representation. Then the embedded sequence is fed to CNN layer for feature extraction. CNN then extracts locally relevant features that are important for identifying the individual sentiment polarity by applying sequence of convolution and pooling operations. Then convoluted feature sequences are re-sampled by pooling layer found next to the convolution operation.

Generally, a pooling layer will be added after the convolution, pooling layer is also arranged into feature maps of a number equal to those of the convolution layer, but with smaller maps. In the pooling layer, a sub-sampling will be done on the activations of the convolution layer to obtain new representations with a reduced feature representation. This is because pooling layer leads to more efficient training by reducing the total number of trainable parameters.

The final pooled features are supplied to GRU to be learned for classification. Finally, in the output layer, SoftMax function is used to calculate the predicted probabilities of each class label in a given textual data representing user comment.

3.3. Data Collection

Nowadays many users use social media outlets to post online contents that should be shared among many peoples. Amhara Media Corporation has Facebook page used to share the daily news broadcasted to wider social media users, users who visited the news shared forwarded their feeling towards the news. The feelings expressed as comments on the corporations Facebook page are used as an important sentiment for identifying viewers attitude and act accordingly for a corrective measure. The data used in this study therefore includes Amharic sentiments i.e., the sentiment text collected from the official Facebook page of Amhara Media Corporation. The sentiments of users given as comments are scraped from the Facebook page by using Comment Exporter and downloaded into an excel file. Total of 10,000 Amharic comments are collected from the official Facebook page of Amhara Media Corporation.

3.4. Preprocessing

Social media data and text usually contain unnecessary symbols and texts, thus need to be removed from the text to make them relevant for further analysis. The unnecessary symbols and texts include URLs, special characters (@, #, $, & and other symbols), HTML scripts that are not relevant for the analysis task. To make the text suitable for the next task, unnecessary symbols and texts must be removed and therefore preprocessed. The preprocessing component performs the task of cleaning and removing unwanted symbols and text. Preprocessing in this case involves removing stop words, emoji icons, other special characters, numbers, punctuation marks and other white spaces found between tokens.

3.5. Tokenization

In this study, Tokenization is one of the tasks that is performed as a part of text preprocessing. This involves converting the sequence of text into discrete tokens that are relevant for further processing. Sequence of text may consist of stop words, numbers and other special symbols that are not important for the next task. Therefore, this sequence of text is reduced to discrete tokens and any special symbol or number appended to tokens of the text can be easily removed once the text is tokenized and reduced to individual tokens.

On the other hand, continuous Amharic text extracted from social media websites such as Facebook pages may contain stop words that do not transmit meaning for the text analysis task. Such type of text should be removed as they only complicate the text analysis and consume the computational resource needed for the text analysis task. Therefore, list of Amharic stop words are prepared and removed from the continuous text through a word lookup strategy.

3.6. Normalization

Normalization is the process of handling different writing system to come up to uniform writing format. Amharic writing system contains different characters but have the same sound during reading which are "ሀ, ሐ, ኀ and ሃ, ሓ, ኃ", "አ, ዐ and ኣ, ዓ", "ሰ and ሠ" and "ጸ and ፀ" respectively. Due to this reason words which have same meaning can be written in different spelling structure. For example, Amharic word "ሀሰት" can be written in different spelling structure such as "ሀሠት" "ሐሰት", "ሐሠት", "ኀሠት" and "ኀሰት" which means false. All these words have the same meaning and sound but have different lexical structure. Also, there are different words which have the same meaning but having different writing structure which means some words are written in short term. for example, 'ዓ.ም', and 'ዓ/ም' have the same meaning with 'ዓመተምህረት' which means Ethiopian calendar. In this study, to solve such types of problem we normalize each similar character and word by changing the character and the word to their common standard form. Therefore, short term and character normalization is performed to have similar word structure. The following algorithm is applied to normalize character and word.

Input: - Amharic and Ge'ez tokens step 1: load tokens step 2: load file which contain list of character and words step 3: for each character and word=>check step 4: if character or word get from list step 5: change each character and word into standard form output: - normalized Amharic tokens

Table 3:Algorithim for Amharic Text Normalization

3.7. Embedding

Once stop words, numbers, punctuation marks and special symbols are removed and normalized, the next task is converting the text sequence into integer sequences. Neural networks work on numeric data, so, textual data should be converted to integer representation. One of the mechanisms to convert textual data into integer representation is word embedding.

As explained by Will Koehrsen (2018), Word embedding is the process of mapping categorical or discrete variables into vector of continuous numbers. It reduces the dimensionality of categorical variables when represented as dense vector in numeric form.

Word embedding for textual data in this study is learned through the embedding layer provided by keras which is used by neural networks Brownlee (N.D). Keras requires the input textual data to be integer encoded and represented by unique integer. This task is performed by the tokenizer class available in keras; that is input textual data is converted into integer encoding by using tokenizer class. Then integer encoded sequence is padded so that it can have fixed length that makes it relevant for neural networks.

3.8. Feature Extraction and Classification

Feature extraction is the process of identifying distinguishing features from either an input image or text. Once an input text is converted into integer representation and padded into fixed sized vectors using word embedding technique, the next step is extracting distinguishing features needed to uniquely identify each text. Relevant local features are extracted from the input text by using series of convolutions available in convolutional neural network. According to Kastrati et al. (2021) one dimensional CNN is useful for extracting local features from given user comments. On the other hand, recurrent neural networks like LSTM and GRU are good in capturing long range dependencies between input text and contextual information. This step consists of inputs obtained from the embedding layer. It consists of three convolution layers, one dropout layer and an activation layer. The convolution layer consists of sequence of convolutional filters used to capture the important local features that represent an input and a convolutional kernel used to scan the useful features from the input. It takes textual input and padded to a fixed length of 40 words as an input, followed by an embedding layer comprising word embedding of size 300D. A Conv1D layer with 512 1D convolution filters of size 3 and a ReLU activation function is appliedon top of the dropout layer. Finally, a fully-connected dense layer composed of a Softmaxfunction and 2 units is used to compute the probability distribution over three sentiment orientations (positive, negative).

`

Chapter Four

4. Experiment Result and Discussion

4.1. Introduction

In the previous chapter, we have discussed the overall design of the proposed model and its detailed architecture. In this chapter we will discuss the implementation and experimental evaluation of the proposed model for Aspect level sentiment analysis of Amharic news. Furthermore, the experimental details such as dataset used for training and testing the proposed model, implementation tools used, evaluation and test results will be presented in detail.

4.2. Dataset Description

Sentiment analysis is performed by using user comments towards some product or issue as input for the analysis task. In this study, our main aim is determining the sentiment polarity of users comment towards Amharic news released from Amhara Media Corporation. Therefore, we collect users comment given to Amharic news posted from the official Facebook page of Amhara media Corporation. In general, for data acquisition, two alternatives are available; the first alternative is directly using data collected by someone else for another research if there is any and the second alternative is collecting data by ourselves and using it for our study. But, as part of our investigation, there is no already available data collected and used for research purpose. Thus, we choose to collect data from the official Facebook page of Amhara Media Corporation by ourselves.

User comments from Amhara Media Corporation are exported into excel file by using comment exporter available at (http://www.commentexporter.com). Totally 10,000 user comments are collected from the official Facebook page of Amhara media corporation.

4.3. Experimental environment setup

To develop the model, we have used Intel® Core™ i5-8250U HP computer with processor speed of 1.8GHz and 8GB RAM with windows 10 operating system. The implementation is done on Python environment with Keras front end and TensorFlow backend. For writing the code for implementation of the proposed model, we used Pycharm and Jupyter Notebook editors.

Table 5: Experiment Tool and Specification

Tools used	Specification
Computer	8GB RAM, 1600 MHz DDR3, Intel® Core™ i5-8250U CPU
Programming language	Python
Libraries	Keras (Tensor Flow 2.3.0 as a backend) and Open CV are used, Moviepy, Librosa
Epoch	100,100,100,100,100 epochs
Batch Size	32
Dataset partition	60% Training, 20% Validation, 20% Testing set
Editors	Jupyter notebook, Pycharm editor

4.4. Experimental result

To conduct the experiment the available data is partitioned into training and testing samples by following a train-test split of 80%, 20% of training, and testing sets respectively. 80% of the data is used to train and develop the actual model.The rest 20% of the data (testing set) is used to test the model on how well it predicts on unseen data. To analyze and evaluate the performance of the proposed model, we have used accuracy, precision, recall and loss as measures of performance metrics.

The experiment for Amharic sentiment analysis model is performed in four ways. First raw user comments are taken and preprocessed to remove unwanted symbols, numbers URLs and stop words that are not necessary for the next process. Then text sequence is converted into vectors of integer representations by using keras embedding layer. After converting the text sequence into word embedding different architectures are used for extracting features from the word embedding and classifying user comments into their corresponding sentiment polarity. These include CNN, LSTM, CNN-LSTM and CNN-GRU architectures, then the performance of all these architectures is compared and the model with high performance is selected as a sentiment analysis model for monitored Amharic news.

4.4.1. CNN Experiment on Amharic News Sentiment Analysis

Training of the model in this case involves series of convolution layers for extracting locally relevant features from the word embedding of Amharic user comments and pooling layers for down-sampling the information. In this method, user Amharic comment is converted into one-hot encoded integer sequences and then padded into fixed length vectors by using keras neural embedding layer. All the sentence sequences have a maximum length of 40 words and sentence sequences with length less than 40words will be filled with zeros from the left side. After word embedding, the padded sequence is given to CNN for feature extraction and recognition. CNN model by itself extracts relevant features using sequence of convolution layers. The training consists of series of convolution, pooling, activation and fully connected layers arranged in a sequential order. The experiment is conducted by using a total text dataset consisting of 10,000 sentences that belong to two classes such as positive (“አውንታዊ”) and negative (“አሉታዊ”). Then the dataset is divided into 80%, 20% of training and testing set.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_2 (Embedding)      (None, 100, 64)           640000
_________________________________________________________________
conv1d_2 (Conv1D)            (None, 93, 32)            16416
_________________________________________________________________
global_max_pooling1d_2 (Glob (None, 32)                0
_________________________________________________________________
dropout_2 (Dropout)          (None, 32)                0
_________________________________________________________________
flatten_2 (Flatten)          (None, 32)                0
_________________________________________________________________
dense_4 (Dense)              (None, 10)                330
_________________________________________________________________
dense_5 (Dense)              (None, 2)                 22
=================================================================
Total params: 656,768
Trainable params: 656,768
Non-trainable params: 0
```

Figure 8:CNN model Summery for Amharic Sentiment Analysis

```
Epoch 00097: val_accuracy did not improve from 0.90923
Epoch 98/100
719/719 [==============================] - 37s 51ms/step - loss: 0.0173 - accuracy: 0.9889 - val_loss: 1.7713 - val_accuracy:
0.8748

Epoch 00098: val_accuracy did not improve from 0.90923
Epoch 99/100
719/719 [==============================] - 33s 46ms/step - loss: 0.0225 - accuracy: 0.9889 - val_loss: 1.6023 - val_accuracy:
0.8764

Epoch 00099: val_accuracy did not improve from 0.90923
Epoch 100/100
719/719 [==============================] - 34s 47ms/step - loss: 0.0198 - accuracy: 0.9883 - val_loss: 1.4011 - val_accuracy:
0.8748

Epoch 00100: val_accuracy did not improve from 0.90923
```

Figure 9:Training Process of CNN Model

As shown in the figure above, the CNN model is trained and validated on 8000 and 2000text sentence sequences respectively.

Model Evaluation

The model is evaluated by using 2000 testing samples. The classification report for the testing samples is presented in figure 11 below. The model achieved test accuracy of 95% which is considered to be good and close to the training accuracy which is 98.8%. Precision, recall and F1-score are recorded as shown below.

	precision	recall	f1-score	support
አዎንታ	0.94	0.97	0.96	1000
አሉታ	0.97	0.94	0.95	1000
accuracy			0.95	2000
macro avg	0.95	0.95	0.95	2000
weighted avg	0.95	0.95	0.95	2000

Figure 10:Model performance Using CNN

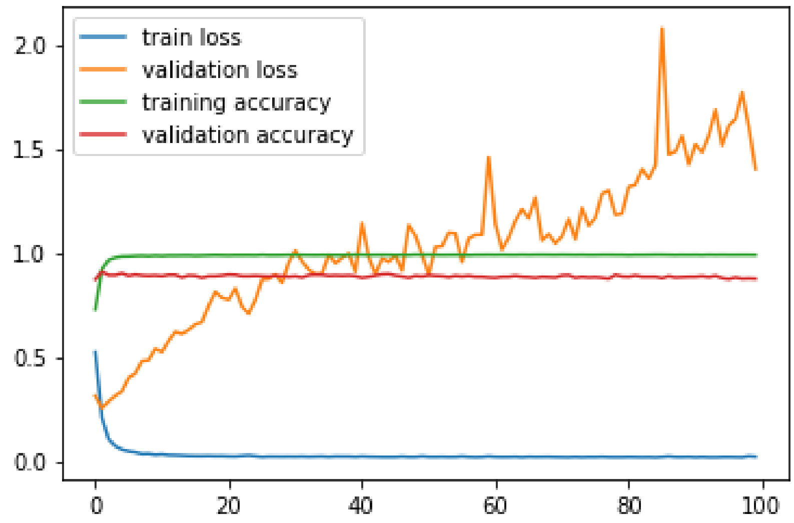

Figure 11: Graph of CNN Model

As shown in the figure 12 above we have used 100 epochs to train our CNN model and the graph of train loss, validation loss, training accuracy and validation accuracy is shown in blue, yellow, green and red color respectively.

4.4.2. LSTM model for Amharic news sentiment analysis

In this experiment, a total of 10,000 comments are used with a train-test split of 80%, 20% training, validation and testing samples are used to train the model, validate the model for overfitting and under fitting, evaluate the performance of the model respectively. In this case Amharic sentence sequences that represent users' comment is converted into a fixed length integer vectors through keras word embedding layer, to make them relevant for the neural network. Then padded sequence of integer vectors representing the sentence sequence is given to LSTM for recognition and classification.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 40, 64)            640000
_________________________________________________________________
lstm_2 (LSTM)                (None, 40, 64)            33024
_________________________________________________________________
lstm_3 (LSTM)                (None, 64)                33024
_________________________________________________________________
dense_3 (Dense)              (None, 60)                3900
_________________________________________________________________
dense_4 (Dense)              (None, 40)                2440
_________________________________________________________________
dense_5 (Dense)              (None, 2)                 82
=================================================================
Total params: 712,470
Trainable params: 712,470
Non-trainable params: 0
```

Figure 12:LSTM Model for Amharic Sentiment Analysis

```
0.8396

Epoch 00097: val_accuracy did not improve from 0.86463
Epoch 98/100
639/639 [==============================] - 58s 91ms/step - loss: 0.0133 - accuracy: 0.9920 - val_loss: 1.4748 - val_accuracy:
0.8380

Epoch 00098: val_accuracy did not improve from 0.86463
Epoch 99/100
639/639 [==============================] - 58s 91ms/step - loss: 0.0135 - accuracy: 0.9912 - val_loss: 1.3358 - val_accuracy:
0.8349

Epoch 00099: val_accuracy did not improve from 0.86463
Epoch 100/100
639/639 [==============================] - 54s 84ms/step - loss: 0.0132 - accuracy: 0.9912 - val_loss: 1.3645 - val_accuracy:
0.8372

Epoch 00100: val_accuracy did not improve from 0.86463
```

Figure 13:Training Process of LSTM Model

The figure below shows the result of the performance metrics precision, recall, and f1-score for testing data. The model is evaluated on a test data of 2,000 which is 20% of the total dataset used, and achieved a test accuracy of 94% as shown in the diagram below.

	precision	recall	f1-score	support
አዎንታ	0.94	0.93	0.94	1000
አሉታ	0.93	0.94	0.94	1000
accuracy			0.94	2000
macro avg	0.94	0.94	0.94	2000
weighted avg	0.94	0.94	0.94	2000

Figure 14:LSTM Model Performance on Amharic Sentiment Analysis

As shown in the training and validation accuracy and loss curve below, training accuracy of the model increases with epoch number and the loss decrease uniformly with epoch number. Training accuracy of the model is mostly greater than validation accuracy. Validation loss of the model is higher than training loss. This shows that model is performing well.

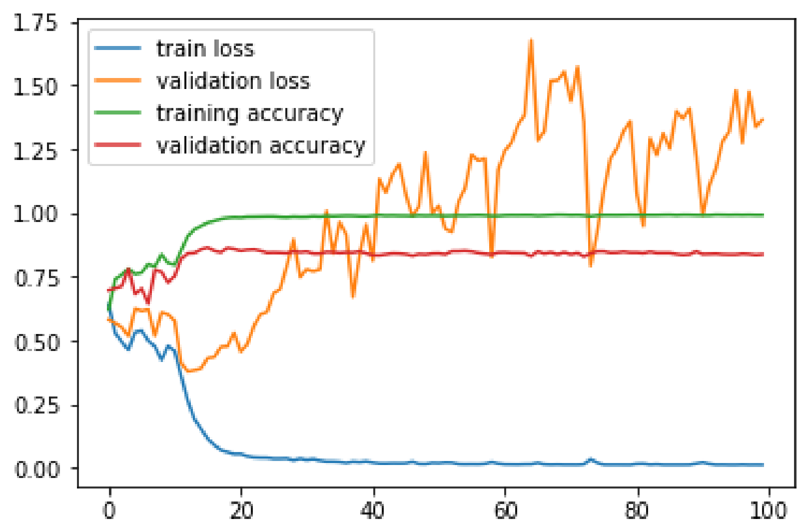

Figure 15:Training and validation accuracy and loss curve of LSTM Model

4.4.3. CNN-GRU model for Amharic news sentiment analysis

In this experiment, a total of 10,000 comments are used with a train-test split of 80%, 20% training and testing samples are used to train the model, validate the model for overfitting and under fitting, evaluate the performance of the model respectively. In this case Amharic sentence sequences that represent users' comment is converted into a fixed length integer vectors through keras word embedding layer, to make them relevant for the neural network. Then padded sequence of integer vectors representing the sentence sequence is given to CNN, that consist of convolution and pooling layers; consequently, features are extracted from the sentence sequence by using convolutional filters of CNN. After feature extraction phase, feature sets are down sampled by pooling layer (applying MaxPooling operation). This helps the neural network to reduce the information space and improve computational efficiency. For performing feature extraction from the sentiment side only Conv1D that makes convolution along the sentiment axis is used. Then features are extracted by three Conv1D layers, and once feature extraction phase is completed, extracted feature sequence is given to GRU for classifying them to their corresponding class label. GRU used temporal dependencies between feature sequences to map them into corresponding class label by using SoftMax layer. The model achieved a training accuracy of 99.99%and validation accuracy 88%. There is no large variation between training and validation accuracy that implies the model is performing well. The overall trend is shown in the figure below.

```
719/719 [==============================] - 77s 107ms/step - loss: 0.0188 - accuracy: 0.9887 - val_loss: 0.7236 - val_accurac
y: 0.8951

Epoch 00097: val_accuracy did not improve from 0.89828
Epoch 98/100
719/719 [==============================] - 74s 103ms/step - loss: 0.0201 - accuracy: 0.9885 - val_loss: 0.7780 - val_accurac
y: 0.8967

Epoch 00098: val_accuracy did not improve from 0.89828
Epoch 99/100
719/719 [==============================] - 76s 106ms/step - loss: 0.0214 - accuracy: 0.9906 - val_loss: 0.7733 - val_accurac
y: 0.8779

Epoch 00099: val_accuracy did not improve from 0.89828
Epoch 100/100
719/719 [==============================] - 78s 109ms/step - loss: 0.0195 - accuracy: 0.9899 - val_loss: 0.9095 - val_accurac
y: 0.8779

Epoch 00100: val_accuracy did not improve from 0.89828
```

Figure 16:Traning Process of CNN-GRU hybrid Approach

To achieve the above performance result, different techniques like dropout and batch normalization are applied, that results in a model free from overfitting. As shown in the above figure 17, the model improves the performance of the previous models with low gap between training and validation accuracy and loss that clearly shows model does not suffered from either under fitting or overfitting.

The figure below shows the result of the performance metrics precision, recall, and f1-score for testing data. The model is evaluated on a test data of 2,000 which is 20% of the total dataset used, and achieved a test accuracy of 94% as shown in the diagram below.

	precision	recall	f1-score	support
አዎንታ	0.93	0.95	0.94	1000
አሉታ	0.95	0.93	0.94	1000
accuracy			0.94	2000
macro avg	0.94	0.94	0.94	2000
weighted avg	0.94	0.94	0.94	2000

Figure 17:Performance of CNN-GRU Model

As shown in the training and validation accuracy and loss curve below, training accuracy of the model increases with epoch number and the loss decrease uniformly with epoch number. Training accuracy of the model is mostly greater than validation accuracy. Validation loss of the model is higher than training loss. This shows that model is performing well.

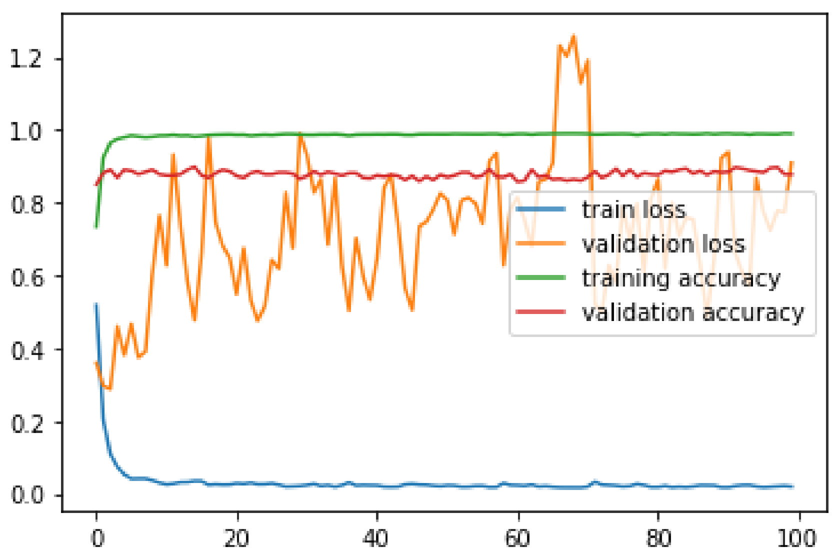

Figure 18:Training and validation accuracy and loss curve of CNN-GRU Model

4.4.4. CNN-LSTM Amharic sentiment analysis recognition model

In this experiment, a total of 10,000 audio samples are used with a train-test split of 80%, 20%, training and testing samples are used to train the model, validate the model for overfitting and under fitting, evaluating the performance of the model respectively. As done in other experiments, important features are extracted by using CNN with sequence of convolutional filters, then extracted feature sequence is pooled to reduce the dimensionality of data so that to improve the computational efficiency and LSTM is applied to deal with sequential feature sequence extracted from sequence of Amharic sentences. Then, LSTM will map the sequential feature vector into corresponding class label. The model achieved a training accuracy of 98.92%, validation accuracy of 86% and testing accuracy of 93%. The network structure of the model is shown in the figure 20 below.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 100, 64)           640000
_________________________________________________________________
batch_normalization_1 (Batch (None, 100, 64)           256
_________________________________________________________________
conv1d_2 (Conv1D)            (None, 96, 64)            20544
_________________________________________________________________
max_pooling1d_2 (MaxPooling1 (None, 48, 64)            0
_________________________________________________________________
conv1d_3 (Conv1D)            (None, 47, 64)            8256
_________________________________________________________________
max_pooling1d_3 (MaxPooling1 (None, 23, 64)            0
_________________________________________________________________
lstm_2 (LSTM)                (None, 23, 64)            33024
_________________________________________________________________
lstm_3 (LSTM)                (None, 64)                33024
```

Figure 19:Model Summery of CNN-LSTM Sentiment Analysis for Amharic News

There is no large variation between training and validation accuracy that implies the model is performing well. The overall trend is shown in the figure below.

```
Epoch 96/100
719/719 [==============================] - 72s 100ms/step - loss: 0.0244 - accuracy: 0.9876 - val_loss: 0.6404 - val_accurac
y: 0.8701

Epoch 00097: val_accuracy did not improve from 0.89515
Epoch 98/100
719/719 [==============================] - 76s 106ms/step - loss: 0.0245 - accuracy: 0.9876 - val_loss: 0.7669 - val_accurac
y: 0.8811

Epoch 00098: val_accuracy did not improve from 0.89515
Epoch 99/100
719/719 [==============================] - 56s 78ms/step - loss: 0.0266 - accuracy: 0.9871 - val_loss: 0.6664 - val_accuracy:
0.8779

Epoch 00099: val_accuracy did not improve from 0.89515
Epoch 100/100
719/719 [==============================] - 49s 68ms/step - loss: 0.0193 - accuracy: 0.9892 - val_loss: 0.9808 - val_accuracy:
0.8623

Epoch 00100: val_accuracy did not improve from 0.89515
```

Figure 20:Training Process of CNN-LSTM Model

The model is tested with 20% of the overall dataset and the accuracy recorded for 2000 of the data is shown in the figure below.

	precision	recall	f1-score	support
አዎንታ	0.95	0.91	0.93	1000
አሉታ	0.92	0.95	0.93	1000
accuracy			0.93	2000
macro avg	0.93	0.93	0.93	2000
weighted avg	0.93	0.93	0.93	2000

Figure 21:Accuracy Model of CNN-LSTM Model

To test to what extent the model is working on testing data, the model is tested on 2000 data and the classification report for the model is shown below. Accuracy, Precision, recall and F-Measure are used as a performance metrics to measure the performance of the model. The overall trend is shown in the figure below.

As shown in the figure below representing accuracy to loss curve, training and validation accuracy of the model is increasing with epoch number and the loss is decreasing except in some epochs. There is no wide variation in training and validation accuracy and loss. This is because, different techniques such as dropout and batch normalization are applied to mitigate overfitting and under fitting problems.

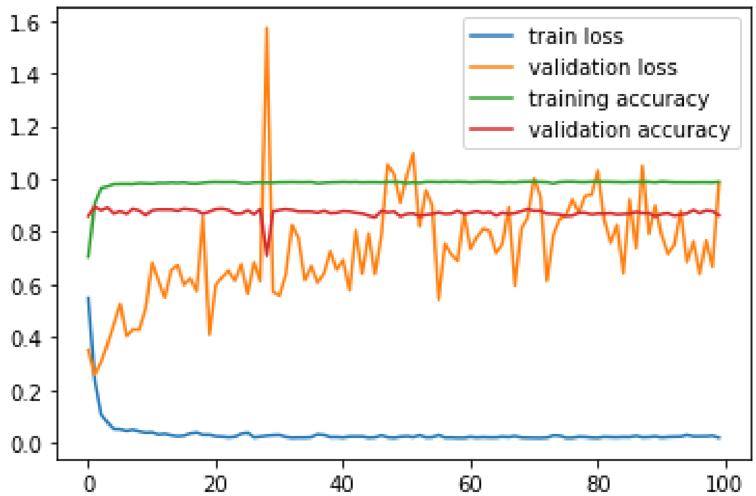

Figure 22:Accuracy to loss curve of CNN-LSTM model

4.5. Summary of comparison of Models

Summary of all experiments conducted is summarized in the table below. The results obtained including training and testing accuracy, time taken for training the model and number of parameters of the model is summarized. As shown in the Table 5.

Table 4: Model Comparison on Amharic Sentiment Analysis

Model	Training Accuracy	Testing Accuracy	Training time/epoch	Validation Accuracy	Validation Loss	Number of parameters
CNN model	97%	95%	7 Sec	87%	40%	656,768
LSTM Model	96%	93%	34 Sec	78%	52%	712,470
CNN-LSTM model	98%	93%	11 Sec	87%	83%	741,526
CNN-GRU model	98.99%	94%	12 Sec	87%	72%	725,398

4.6. Discussion

In this study Amharic news sentiment analysis model is proposed and its performance is evaluated on test data. The system consists of basic components that perform different tasks that are intermediate to the final result, that includes the preprocessing component for removing unnecessary symbols, numbers and URLs and the word embedding component for converting the sentence sequence into integer sequence and then fixed length vectors through Keras word embedding layers. After converting the input sentence into vectors of integers, then the padded sequence is given to the CNN that constitute the feature extraction component for extracting relevant and distinguishing components from the sentence. Four experiments are conducted independently for developing CNN, LSTM, CNN-LSTM and CNN-GRU models for Amharic sentiment analysis.

The results obtained in the four experiments were presented in the previous section. As discussed in the previous section, an CNN based sentiment analysis model achieved an accuracy of 98.8%, LSTM based model achieved an accuracy of 99.1%, CNN-LSTM model achieved an accuracy of 98.92% and CNN-GRU sentiment analysis model achieved an accuracy of 98.99%. This shows that CNN-GRU model for Amharic sentiment analysis outperform all other approaches with better classification accuracy. This is because CNN has good ability to learn features from raw data. As stated in Fanta (2010), neural networks are good in predicting isolated data units when compared with other methods. Applying CNN, a neural network technique comparable recognition accuracy to state-of-the-art techniques is obtained. As stated in Dridi and Ouni (2020), CNN is able to model limited temporal dependency and it is not good to model sequence of sentences which are sequential in nature. But recurrent neural networks can handle temporal dependency effectively and are more effective than CNN.

Therefore, when we think of sentence sequences, it consists of sequences of words that are treated in sequence to form recognizable and meaningful language unit. Therefore, architectures that can process sequential data such as LSTM and GRU are applied to the raw text data. By using these methods an accuracy of 96% and 97% respectively is achieved. By using such techniques this shows that the overall accuracy of the model is improved. Next combinations of CNN-LSTM and CNN-GRU techniques are applied and show remarkable improvement in recognition accuracy. Therefore, based on the experimental results presented above CNN-GRU model outperform all others methods with accuracy.

Therefore, in this study a comparison is made between the different techniques for Amharic sentiment analysis and the best model that has achieved better recognition accuracy is selected. As part of our investigation, there are no research works that has applied combination of CNN, CNN-LSTM and CNN-GRU techniques and make comparative analysis on the performance of these techniques. The main contribution of the study is localization of CNN-GRU and CNN-LSTM architectures for aspect level sentiment analysis of Amharic news comments for which no research has never been done using hybrid approach and compare the performance of this approaches for Amharic sentiment analysis by using comments of users on monitored Amharic News from Amhara Media Corporation.

Chapter Five

5. Conclusion and Recommendation

5.1. Conclusion

Nowadays, huge number of feedbacks is collected from social media users on services and products delivered by organizations and institutions. If analyzed correctly and used by the organization as an input to improve their operation, these comments will have big contribution in enhancing their performance. But analyzing such kind of comments and users' sentiment is very challenging due to the time required and its complexity. But recently, sentiment analysis is making great contribution by enabling automatic analysis of user's comments with in short time and less complexity by using different deep learning techniques.

Therefore, the goal of this study is to compare different deep learning architectures for Amharic sentiment analysis by using social media comments of users on monitored Amharic news from Amhara mass media agency and evaluate them, apply the architecture that perform better for Amharic news sentiment analysis. The performance of different architectures such as CNN, LSTM, CNN-LSTM and CNN-GRU architectures are explored to see their performance. But CNN-GRU architecture out perform all other methods and it is selected for the Amharic news sentiment analysis task owing to its accuracy.

In this study we proposed a model that is capable of recognizing sentiments of users on Amharic news into their corresponding sentiment polarity level as positive or negative. The model has three components; the first component is the preprocessing for processing the text sequence and remove URLs, numbers, symbols, and emojis. The second component comprises of embedding layer for converting sentences into fixed length integer sequences and the convolution and pooling layers of CNN for feature extraction. The third component comprises LSTM layer for sentiment classification. This makes up the classification component.

To conduct this study, user comments consisting of 10,000 Amharic sentences from social media users of Amhara media corporation Facebook page is collected. The comment consists of sentiments categorized in two classes such as 'አውንታዊ'(positive) and አሉታዊ(Negative) polarities. By using these comments CNN, LSTM, CNN-GRU and CNN-LSTM models for sentiment analysis are developed and their performance is compared. The comparison result is presented in the previous section and CNN-GRU model outperformed all other techniques. Regularization techniques such as dropout and batch normalization are applied during model development to mitigate overfitting problem and yield better accuracy. The CNN-GRU model achieved training accuracy of 98% and test accuracy of 98%.

5.2. Contribution

The main contributions of this study are described as follows:

- ✓ We have proposed CNN-GRU Amharic sentiment analysis model for recognition of the polarity of Amharic comments from social media users.
- ✓ We made a comparison among different deep learning architectures such as CNN, CNN-LSTM, and CNN-GRU for Amharic sentiment analysis based on monitored comments on monitored Amharic news from social media pages.

5.3. Recommendation

In conducting this thesis the researcher wanted to consider comments with figurative expression which have surface and undersurface meaning. However, it is becomes ambiguous to use expressions having more than one meaning. Figurative words will have same weight with literal words having one and direct meaning. Due to this challenge such expressions are omitted. Apart from that the study has achieved its purpose and based on the finding and results of the study, the following are recommended by the researcher for future work.

- ✓ Integrating word sense disambiguation and other techniques such part of speech tagging will improve the recognition accuracy of the sentiment analysis task. Therefore, future researchers can work on integrating such tasks to the sentiment analysis task.
- ✓ The current model is trained on limited data. Future researches can focus on improving recognition rate with more diverse data from other social media platforms.

Reference

Abbas, M., Ali Memon, K., & Aleem Jamali, A. (2019). Multinomial Naive Bayes Classification Model for Sentiment Analysis. *IJCSNS International Journal of Computer Science and Network Security*, *19*(3), 62. https://doi.org/10.13140/RG.2.2.30021.40169

ADDIS ABABA INSTITUTE OF TECHNOLOGY SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING EFFECT OF PREPROCESSING ON LONG SHORT TERM MEMORY BASED SENTIMENT ANALYSIS FOR AMHARIC LANGUAGE By : Turegn Fikre A Thesis Submitted to the School of Graduate Studies of Add. (2020).

Albawi, S., & Mohammed, T. A. (2017). *Understanding of a Convolutional Neural Network*. *April 2018*. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Aragaw, B. T. (2015). Sentence level Amharic Sentiment Analysis Model: A Combined Approach. *University of Gondar Msc Thesis*, 32.

Aszemi, N. M., & Dominic, P. D. D. (2019). Hyperparameter optimization in convolutional neural network using genetic algorithms. *International Journal of Advanced Computer Science and Applications*, *10*(6), 269–278. https://doi.org/10.14569/ijacsa.2019.0100638

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., & Belyaeva, J. (2010). Sentiment analysis in the news. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, *September*, 2216–2220.

Brownlee, J. (n.d.). *How to Use Word Embedding Layers for Deep Learning with Keras*.

Chachra, A., Mehndiratta, P., & Gupta, M. (2018). Sentiment analysis of text using deep convolution neural networks. *2017 10th International Conference on Contemporary Computing, IC3 2017*, *2018-Janua*(August), 1–6. https://doi.org/10.1109/IC3.2017.8284327

Chilot, D. (2019). *Public sentiment analysis for Amharic news*.

College of Natural Sciences School of Information Science Addis Ababa University College of Natural Sciences School of Information Science. (2013).

D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*, *125*(3), 26–33. https://doi.org/10.5120/ijca2015905866

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, *87*, 44–49. https://doi.org/10.1016/j.procs.2016.05.124

Dridi, H., & Ouni, K. (2020). Towards Robust Combined Deep Architecture for Speech Recognition : Experiments on TIMIT. *International Journal of Advanced Computer Science and Applications*, *January*. https://doi.org/10.14569/IJACSA.2020.0110469

El-halees, A. (2011). Arabic opinion mining using combined classification approach. *Proceeding The International Arab Conference On Information Technology, Azrqa, Jordan.*, 264–271.

Fanta, H. (2010). *Speaker Dependent Speech Recognition for Wolaytta Language*. Addis Abeba.

Gebremeskel, S. (2010). *SCHOOL OF GRADUATE STUDIES SENTIMENT MINING MODEL FOR OPINIONATED By : Selama Gebremeskel FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES SENTIMENT MINING MODEL FOR OPINIONATED AMHARIC TEXTS*. Angeles, L., Advocacy, S., Location, O. (2002).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. *Cambridge: MIT Press*, *1*(2).

Grosz, B. J. (1982). Natural language processing. *Artificial Intelligence*, *19*(2), 131–136. https://doi.org/10.1016/0004-3702(82)90032-7

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in Convolutional Neural Networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48. https://doi.org/10.1016/j.neucom.2015.09.116

Kastrati, Z., Ahmedi, L., Kurti, A., Kadriu, F., Murtezaj, D., & Gashi, F. (2021). A deep learning sentiment analyser for social media comments in low-resource languages. *Electronics (Switzerland)*, *10*(10), 1–19. https://doi.org/10.3390/electronics10101133

Kaushik, C., & Mishra, A. (2014). A Scalable, Lexicon Based Technique for Sentiment Analysis. *International Journal in Foundations of Computer Science & Technology*, *4*(5), 35–56. https://doi.org/10.5121/ijfcst.2014.4504

Koehrsen, W. (2018). *Neural Network Embeddings Explained _ by Will Koehrsen _ Towards Data Science*.

Leetaru, K. (2020). Sentiment Analysis. *Data Mining Methods for the Content Analyst*, 79–84. https://doi.org/10.4324/9780203149386-11

Nandi, V., & Agrawal, S. (2016). Sentiment Analysis using Hybrid Approach. *International Research Journal of Engineering and Technology*, 1621–1627.

Palanisamy, P., Yadav, V., & Elchuri, H. (2013). Serendio: Simple and practical lexicon based approach to sentiment analysis. **SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, *2*(SemEval), 543–548.

Patel, V., Prabhu, G., & Bhowmick, K. (2015). A Survey of Opinion Mining and Sentiment Analysis. *International Journal of Computer Applications*, *131*(1), 24–27. https://doi.org/10.5120/ijca2015907218

Radhakrishnan, P. (2017). *What are Hyperparameters _ and How to tune the Hyperparameters in a Deep Neural Network*. Towards Data Science; Towards Data Science.

Sharma, A., & Dey, S. (2012). Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis. *International Journal of Computer Applications*, *June*, 15–20. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Performance+Investigation+of+Feature+Selection+Methods+and+Sentiment+Lexicons+for+Sentiment+Analysis#0

Shewalkar, A., Nyavanandi, D., & Ludwig, S. A. (2019). *PERFORMANCE EVALUATION OF DEEP NEURAL NETWORKS APPLIED TO SPEECH RECOGNITION : RNN , LSTM AND GRU*. *9*(4), 235–245.

Shirsat, V. S., Jagdale, R. S., & Deshmukh, S. N. (2018). Document Level Sentiment Analysis from News Articles. *2017 International Conference on Computing, Communication, Control and Automation, ICCUBEA 2017*, 1–4. https://doi.org/10.1109/ICCUBEA.2017.8463638

Steinberger, R., Hegele, S., Tanev, H., & Rocca, L. Della. (2017). Large-scale news entity sentiment analysis. *International Conference Recent Advances in Natural Language Processing, RANLP*, *2017-Septe*, 707–715. https://doi.org/10.26615/978-954-452-049-6-091

Tamiru, N. K. (2018). Amharic Sign Language Recognition based on Amharic Alphabet Signs. In *Addis Ababa University Master Thesis*. Addis Ababa University.

Tefera, S. (2005). *Automatic speech Recognition for Amharic* (Vol. 7). Hamburg University.

Teshome, Y. (2019). *Debre berhan university college of computing*. *June*.

Thakkar, H., & Patel, D. (2015). *Approaches for Sentiment Analysis on Twitter: A State-of-Art study*. http://arxiv.org/abs/1512.01043

To, A. T. S. (1991). Department of Computer Science. *Journal of Computational and Applied Mathematics*, *34*(2), N14. https://doi.org/10.1016/s0377-0427(91)90073-s

Wang, Y., Li, Y., Song, Y., & Rong, X. (2020). The influence of the activation function in a convolution neural network model of facial expression recognition. *Applied Sciences (Switzerland)*, *10*(5). https://doi.org/10.3390/app10051897

Worku, K. (2009). *School of Graduate Studies Faculty of Informatics Department of Information Science Automatic Amharic Text News Classification : a Neural Networks Approach School of Graduate Studies Faculty of Informatics Department of Information Science Automatic Amhar*. *September*.

Yamashita, R., Nishio, M., Kinh, R., Do, G., & Togashi, K. (2018). *Convolutional neural networks : an overview and application in radiology*. 611–629.

Yimam, S. M., Alemayehu, H. M., Ayele, A., & Biemann, C. (2021). *Exploring Amharic Sentiment Analysis from Social Media Texts: Building Annotation Tools and Classification*

Models. 1048–1060. https://doi.org/10.18653/v1/2020.coling-main.91

Zhang, Y. J. (2021). Handbook of Image Engineering. In *Handbook of Image Engineering*. https://doi.org/10.1007/978-981-15-5873-3