JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATICS

MULTI-CLASS SUBJECTIVITY DETECTION AND SENTIMENT ANALYSIS USING MACHINE LEARNING APPROACH: A Case Study on AMHARIC Social Media Posts on COVID-19

By: Meti Bekele Tufa

A thesis submitted to Faculty of

Computing and Informatics Jimma institute of

Technology in partial fulfillment for

The degree of master of science in

Information Technology

June, 2022 GC

Jimma, Ethiopia

MULTI-CLASS SUBJECTIVITY DETECTION AND SENTIMENT ANALYSIS USING MACHINE LEARNING APPROACH: A Case Study on AMHARIC Social Media Posts on COVID-19

BY

Meti Bekele Tufa

A Thesis Submitted to the Faculty of Computing and Informatics Jimma Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Information Technology in the faculty of computing.

Advisor

Dr. Kula Kekeba (PhD)[1]

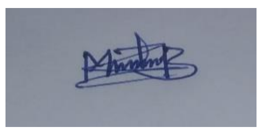
June, 2022GC

Jimma, Ethiopia

[1]Addis Ababa Science and Technology University

Declaration

I hereby declare that this Master thesis entitled **"Multi-class subjectivity detection and sentiment analysis using machine learning approach: a case study on Amharic social media posts on covid-19"** was Submitted to the Faculty of Computing Jimma Institute of Technology in partial fulfillment of the requirements for the degree of Masters of Science in Information Technology is a record of an original work done by me under the guidance of Dr. Kula Kekeba that it has not been submitted for the award of any academic degree, diploma, or certificate in any other university. I have not used any sources other than those listed in the bibliography that is identified as reference.



Meti Bekele Tufa

June 20, 2022

Jimma University

Jimma Institute of Technology

Faculty of Computing and Informatics

A Thesis Submitted to the School of Graduate Studies of Jimma University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology

Approval sheet

Thesis Title: Multi-Class Subjectivity Detection and Sentiment Analysis Using Machine Learning Approach: A Case Study on Amharic Social Media Posts on Covid-19.

Submitted by:

Meti Bekele Tufa	______________	June 20, 2022
	Signature	Date

Approved by:

Name and Signature of advisor

Kula Kekeba (PhD)	______________	June 20, 2022
Principal advisor	Signature	Date

Name and Signature of the Examining Board

______________________	______________	______________
Chairman, Examining Board	Signature	Date
______________________	______________	______________
Examiner	Signature	Date
Mamo Fideno	______________	June 20, 2022
Mock Examiner	Signature	Date

June 2022

Jimma, Ethiopia

Acknowledgment

First and for most, praises and thanks to the **God**, the Almighty, and his mother **Blessed Virgin Mary** for showers of blessings throughout my research work. I would extend my gratitude to my advisor **Dr. Kula Kekeba** for his guidance, feedback, and constructive comments during this work.

I must express my very profound gratitude to **MY PARENTS** and **MY HUSBAND** for providing me with unfailing support and continuous encouragement throughout my years of study, through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Finally, I thanked all of **MY FRIENDS** who have helped and encouraged me throughout my career.

Dedication

I would like to dedicate this thesis to my beloved Father.

Contents

List of Abbreviations

API: Application Program Interface

BOW: Bag of Words

COVID: COrona VIrus Disease

CSV: Comma-Separated Value

FN: False Negative

FP: True Positive

IR: Information Retrieval

MDS: Multi Domain Sentiment

ME: Maximum Entropy

NB: Naïve Bayes

NLP: Natural Language Processing

OM: Opinion Mining

POS: Part of Speech Tagging

RFC: Random Forest Classifier

SVC: Support Vector Classifier

SVM: Support Vector Machine

TF- IDF: Term Frequency Inverse Document Frequency

TN: True Positive

TP: True Negative

WHO: World Health Organization

List of Figure

List of Table

Abstract

Nowadays the growth of internet technology and smartphones help people to liberally express their idea about health issues, governmental policies and services, political campaigns, business and product companies, and public services issues through different social media platforms. Due to the huge amount of data produced on the internet, identifying them into factual statements and opinions for sentiment analysis classification is a difficult task. Sentiment analysis, which is a branch of data mining and natural language processing, examines people's feelings, opinions, sentiments, evaluations, appraisals, attitudes, and emotions toward entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. In this study, we developed a Multi-class Amharic Subjectivity Detection and Sentiment analysis using Machine Learning Approach with natural language processing techniques. For this study, we have collected data from Ethiopian Ministry of Health, official Facebook page by using a face pager tool. We have used a total of 4988 annotated sentences for training and testing purposes. Our current work focused on first classifying statements as Objective, Subjective, and Non-Related. As we know for sentiment analysis only subjective statements are required and therefore, here we take only the subjective statements for the classification. Then after taking the subjective statement, we classify them into six classes: Hope, Fear, Sadness, Anger, Confusion, and Others. TF-IDF is used for the feature extraction tasks. Then on extracted feature representation, we have employed two machine-learning algorithms, Random Forest and Support Vector Machines for classification. We have undertaken several experiments to determine the best performing model. The Support vector machine model is best performing with 93.78% accuracy for Subjectivity Detection and the Random Forest model is best performing with 94.3% accuracy for Multi-Class Sentiment Analysis.

Keywords: *COVID-19, Natural Language Processing, Sentiment Analysis, Subjectivity Detection, TF-IDF, SVM, RFC.*

Chapter One: Introduction

1.1. Background

Coronavirus Disease 2019 (COVID-19) pandemic remains a continuing serious global problem. The outbreak was first noted in December 2019 in Wuhan, China[1]. The World Health Organization (WHO) declared this virus as a pandemic in March of 2020. Since then, this pandemic has affected the lives of millions of people and the global economy with a serious impact on developing countries of the world[2][3]. Ethiopia is one of the 227 countries that have reported COVID-19 cases since June 20, 2022. Several cases and deaths have been identified. According to the reports of the Ethiopian Ministry of Health (MOH), a total of 484,138 cases, 7,523 deaths, and 458,280 recoveries were reported, as of 20 June 2022. In the country, the COVID-19 mortality and morbidity rate are expected to rise since the case reports have been coming from all parts of the country[2].

The WHO and Health Ministry of Ethiopia have disseminated several information related to the burden, modes of transmission, prevention, and controlling techniques of the disease using several media of which, social media is becoming the most commonly used in areas where internet access is available[2][3].

Social media have played essential roles during the outbreak and have continued to do so as COVID-19 spreads, globally[4]. Several people use different social media platforms to share and express their feelings and opinions during the lockdown period and after the lockdown during vaccination[2][4]. Opinions are the heart of all human activity, and have a significant impact on the actions of human beings. Beliefs, perceptions of reality, and the decisions humans make are all influenced by how others perceive and interpret the world. As a result, when humans need to make a decision, they frequently seek the advice of others.

Currently, the internet has a massive amount of text-based information, and the problem of identifying people's expressed opinions in written language is a relatively new and active area of research[5][6].

The growth of internet technology and smartphones help people liberally express their opinion about health issues, governmental policies and services, political campaigns, business and product companies, and public services issues through social media platforms[4]. But due to the unstructured format of data available on the internet, it is tedious task to determine the polarity of texts, extract useful information, find relevant information, summarize, and make efficient decisions. Moreover, it is a time-consuming task to identify the subjectivity of ideas that are rotating on the internet. Therefore, there is a need for automated sentiment analysis of these reviews.

Sentiment analysis is a branch of data mining and natural language processing that examines people's feelings, opinions, sentiments, evaluations, appraisals, attitudes, and emotions toward entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. Sentiment analysis can be used mainly in fields such as health, marketing, politics, and sociology[5][6].

Sentiment analysis can be done at three different levels of abstraction[5][7]:

- ✓ Document-level classification: this determines whether an opinion document expresses a positive or negative opinion or sentiment. It regards the entire document as a basic information unit (talking about one or a single topic)[7][8].
- ✓ Sentence-level classification: this categorizes the sentiment expressed in each sentence. If the sentence is subjective, it categorizes as positive or negative[9].
- ✓ Aspect-level classification: this categorizes sentiment about specific aspects of entities. Users perspectives on different aspects of the same entity can be different[10][11].

The main goal of sentiment analysis is to identify and extract subjective information from text files, such as a reviewer's feelings, thoughts, judgments, or assessments about a specific topic, event, or, as previously mentioned, a company and its activities. This type of analysis is also referred to as opinion mining (with a focus on extraction) or effective rating. Some experts also use the terms sentiment classification and extraction. The goal of sentiment analysis, regardless of its name, is the same to determine a user's or audience's opinion on a target object by analyzing a large amount of text from various sources[12][13].

There are two types of perception about given information: subjective and objective. Subjective refers to an individual's personal opinions or feelings about a particular subject. Subjective perspectives or opinions are not founded on truth or fact[13]. They are one person's interpretation of an idea, as well as their thoughts, feelings, and background. The term Objective refers to factual, data-driven information that is free of bias. When someone gives you an objective assessment of a topic, it is based on data, verifiable facts, or other guaranteed evidence, and does not take the speaker's personal feelings into account[12]. In the case of sentiment analysis, as we are depending mainly on the opinion of people about a specific issue, we only rely on the subjective perspective of people. For performing this task, the subtask subjectivity analysis is desired. Which identify a given piece of information as subjective or objective.

Only a limited number of research works performed yet on the Amharic sentiment analysis task [14][15][10][16][17]. The reasons include morphological complexity, spelling variation, in-availability of labeled data, less existence of researched resources, lack of texts available on the web, inaccessibility and inconvenience to typing, character redundancy, and absence of abbreviation rules are few reasons for the limitation. A few researches on Amharic sentiment analysis use approaches such as lexicon approach[17], hybrid approach[10][14], and machine learning-based[15]. These works have several limitations: (a) they don’t include subjectivity analysis as a part of their work, (b) they only depend on binary classification of sentiments even if [15] try to perform multi-class its low in its accuracy, (c) they only use a small number of dataset which affect the accuracy, (d) they only perform simple classification as positive or negative. Having motivation on the above limitations this research attempts to perform multi-class subjectivity detection and sentiment analysis using machine learning approach: a case study on Amharic social media posts on covid-19.

In general, we proposed a model for extracting subjective information from Amharic texts based on natural language processing tasks and then detecting and classifying that subjective information into predefined classes such as hope, anger, sadness, fear, confusion, and others.

1.2. Statement of the problem

Sentiment analysis is one way to understand the feeling (emotion) of a person about some topic, issue, or idea and currently, a number of people use different social media platforms such as Facebook, YouTube, Twitter, personal blog, etc to freely express or share their opinions and be able to see the opinions of others as well. Moreover, we all have experiences of using social media to get recommendations about different issues such as items, movies, hospitals, and so on; we also give our suggestions on several concepts that include politics, economy, and social concerns.

Due to the easy access to social media and the availability of several applications on smartphones, there is a wider opportunity for the creation of voluminous data. Basically, opinions on the web are categorized into two groups implicit opinions (objective statements that express a factual opinion) and explicit opinions (subjective statements that express one's own opinion). Distinguishing objective statements from the subjective ones is needed for sentiment analysis to depend solely on subjective statements as sentiment classification tasks proceed further. But due to the availability of numerous unstructured data on the web, it is tedious and time-consuming activity to classify the two statements manually. Therefore, performing a subjectivity analysis is required to identify the statements so that the resulting subjective statement can be used for the next sentiment classification task.

The researcher[15] try to perform multi-class sentiment analysis which classify a given sentence into five classes by using machine learning approach on 608 online posts. They use small number of data so that they achieve very low accuracy and they don’t include subjectivity analysis in their work.

Another work on sentiment analysis performed by[14] on sentence level sentiment analysis using 800 Facebook posts by employing lexicon and dictionary based approach to classify into positive, negative, and neutral, however when we talk about sentence level sentiment analysis it has two main tasks: first applying subjectivity analysis then applying sentiment classification. But in their work, they ignored subjectivity analysis which is the main task in sentence level sentiment analysis. They only collect subjective statements for sentiment classification.

Generally, even if few researchers played their role in the growth of Amharic sentiment, analysis in binary classification, it is believed that there is no enough research conducted in the area of multi-class sentiment analysis. The researchers tried their best to bring the idea into the context of Amharic text, however; it is low in its accuracy. Moreover, the researchers did not include subjectivity analysis, which classify the statements first into subjective and objective that can reduce the effort to be devoted to building sentiment corpus by avoiding manual identification of texts. And in general, research of sentiment analysis in the area of health domain is not performed yet in Amharic language. Additionally, people don't only give positive or negative opinion about given topic rather they show their emotion in their comment in a various way. Due to this classifying the sentiments into different emotion are required and performed in this study.

In this research, multi-class subjectivity detection and sentiment analysis using machine learning approach: a case study on Amharic social media posts on covid-19 by first applying subjectivity analysis (classifying as subjective and objective) and then classifying the subjective statement into six classes named as Fear, Sadness, Hope, Confusion, Anger and Other are proposed.

At the end of the study, the research will answer the following research questions accordingly:

- ✓ Which preprocessing technique need to be applied to prepare Quality Amharic sentiment analysis dataset?
- ✓ Which feature extraction technique is better for Multi-class Amharic Subjectivity Detection and Sentiment Analysis for COVID-19 disease?
- ✓ Which Machine learning Algorithm performs better classification?
- ✓ Dose subjectivity analysis have impact in sentiment analysis?

1.3. Objectives of the study

1.3.1. General objective

The general objective of this research work is to design a model for multi-class subjectivity detection and sentiment analysis using machine learning approach: a case study on Amharic social media posts on covid-19.

1.3.2. Specific Objectives

- ✓ To perform extensive reviews on previous research work on sentiment analysis to get a deep understanding of the area.
- ✓ To prepare Amharic text corpus for training and testing from Facebook.

- ✓ To design appropriate Amharic text preprocessing algorithm.
- ✓ To develop subjectivity detection model using machine learning approach.
- ✓ To classify subjective statements into their appropriate predefined category based on appropriate machine learning approach.
- ✓ To evaluate the performance of the designed model using evaluation metrics.
- ✓ Giving conclusion and recommendation.

1.4. Methodology

1.4.1. Research design

Research design can be thought of as the "glue" that connects all of the components of a research project; in other words, it is a blueprint for the proposed research effort[18]. To achieve multi-class sentence-level emotion detection for Amharic text we followed experimental research design methodology. The reason behind to select this methodology is it follows a clear procedure that helps us to achieve our research objectives such as identifying and defining the problem, reviewing relevant literature, formulating hypotheses and deducing their consequences, compiling raw data and condensing to usable form, constructing an experimental design, conduct the experiment, present findings, and conclusions.

1.4.2. Literature review

A literature review is a survey of scholarly sources on a specific topic so we performed a deep review on sentiment analysis and its technique and tools from different books, published papers, journals, articles, and other materials to understand how sentiment analysis deals with a different language and different level. In addition, we make an extensive review to understand the linguistic behavior of the Amharic language.

1.4.3. Data collection

Data (comments and posts) that are used to train and test the model were collected from Ministry of Health, Ethiopia's official Facebook page. The sentences are collected using the Face pager from the topics posted on the Facebook page due to the large amount of data available on Face book. The collected sentences are annotated manually into subjective, objective, and non-related and the subjective statements are classified into fear, hope, anger, sadness, confusion, and others so that the label data help to learn supervised machine learning algorithms. Then in order to clean and make those collected statements understandable by machine learning algorithms we use different natural language preprocessing tasks such as text cleaning, tokenization, normalization, and stop word removal.

1.4.4. Approach and tool

Approaches like Machine Learning, Lexicon-based, and hybrid are used to classify the emotion of the sentiment of a given text given in social media but in our case, a supervised machine learning approach SVM and RFC was employed, which uses labeled datasets to train algorithms. The particular reason for the circumstance is that lexicon-based approaches, fail to recognize the domain-specific word but machine learning have an advantage in its ability to adapt and create trained models for specific purposes (domain) and its achieve high accuracy if the data is excellent in quality and quantity[5]. We used the python programming language as a preprocessing, and analyzing tool for the collected data. And we use different python machine learning library like Pandas, Numpy, Nltk, Matplotib, and Sklearn to read, write and manipulate data frames, to work with arrays, for building Python programs that work with human language data, for data visualization and graphical plotting, and tools for classification, regression, clustering, and dimensionality reduction respectively.

1.4.5. Evaluation

The total data collected by using Face pager is split into two parts for training and testing purpose. The training data help us to build our model and the testing data help us to examine or to measure the performance of our model.

To evaluate the performance of the model built, we used precision, recall, and F-measure. Precision is the fraction of relevant retrieved instances, while recall is the fraction of retrieved relevant instances. F-measure is 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[19].

1.5. Scope and Limitation

1.5.1. Scope of the study

The current study focused on multi-class subjectivity detection and sentiment analysis using a machine learning approach: a case study on Amharic social media posts on covid-19. To achieve the objectives of the study, we have collected data from the Ministry of Health, Ethiopia's official Facebook page by using a face pager. The collected data are annotated properly into a pre-defined label. Our newly built dataset was preprocessed in machine-understandable form by feature extraction technique. The extracted feature representation would be passed to SVM, and RFC machine-learning algorithms to first detect subjective statements and classify the detected subjective statements into predefined classes. The developed multi-class subjectivity detection and sentiment analysis trained models are tested and evaluated with performance evaluation metrics. Finally, a better-performed model with feature combinations was selected.

1.5.2. Limitations of the study

This research was only focus on sentence-level sentiment analysis and it will not include a document or aspect-level sentiment analysis. Even resourced languages like English and Mandarin Chinese moved to multimodal sentiment analysis (image, video). Due to different obstacles, we only limit our work to Amharic text sentiment analysis. According to[20], a sentence expressing a single opinion or emotion there for our research does not apply to compound or complex sentences because they often express more than one sentiment in a sentence. In addition, this work focusses only on comments and posts written for COVID-19 to classify them into six groups Fear, Anger, Confusion, Hope, Sadness and Other.

1.6. Significance of the study

Every day, millions of online users express their opinions about product features, benefits, and value through a variety of channels. This opinion or sentiment data which is sometimes generated invisibly frequently contains critical data points that can be quite useful for organizations looking to improve their customer experience, products, or services. To begin with, sentiment analysis is important because emotions and attitudes toward a topic can be transformed into meaningful data in a variety of fields, including business and research. Also, it saves time and effort because the sentiment extraction process is completely automated the algorithm analyzes the sentiment datasets, requiring little human intervention.

In our current world, knowing the opinion of others about some issue is a key way for making efficient decisions. The significance behind this paper is to identify the given text as subjective and objective and then classify subjective statements into Fear, Anger, Confusion, Hope, Sadness, and Others by collecting data from the Ministry of Health, Ethiopia's official Facebook page by using a face pager. And will be able to automatically analyze the sentiment of a huge amount of collected data prior to making decisions. Therefore, by using the analysis output produced from the system, government bodies including the ministry of health can be able to take appropriate action.

The results of the research can be used as an input to the development of a full-fledged opinion mining system for Amharic language or any other Ethiopian languages. Another important of the study is that the output can be used as input data for recommendation and opinion retrieval/search systems.

1.7. Organization of the Thesis

The following is the remainder of this thesis. The second chapter contains reviews of Amharic language and various types of literature on sentiment analysis, as well as its approaches and various machine learning techniques also review different literature about preprocessing techniques. The third chapter discusses the study's overall methodology, illustrating sentiment analysis techniques and algorithms such as corpus preparation and preprocessing, system architecture, feature selection methods, classification techniques, and performance measurement. The fourth chapter discusses the experimental results as well as the findings of how these experiments and methodologies are implemented. Finally, Chapter five discusses the conclusion and the future work.

Chapter Two: Literature Review

2.1. Overview of literature review

A literature review is a survey of scholarly articles, books, and other sources relevant to a specific issue, field of study, or theory, to provide an explanation, summary, and critical evaluation of these works. Literature reviews are intended to provide an overview of the sources we used while researching a specific topic and to show our readers how our research fits into the larger field of study. A literature review is a summary of the literature on a particular subject or field. It summarizes what has been said, who the key writers are, what theories and hypotheses are prevalent, what questions are being raised, and what methods and methodologies are appropriate and useful. As such, it is not primary research in and of itself but rather reports on other findings [21]. In the literature review section, we are going to see some introduction to natural language processing, an overview of the Amharic language. And also, we discover some kinds of literature about sentiment analysis including the definition of sentiment analysis, type of opinion, components of opinion mining, levels of sentiment analysis, steps to analyze sentiment data, approach to sentiment analysis, subjectivity analysis, challenges in sentiment analysis, and some sentiment analysis related works performed in Amharic language.

2.2. Introduction to Natural language processing

Natural Language Processing (NLP) is a subfield of artificial intelligence that entails the automatic manipulation of natural languages such as speech, image, video, and text in the same way that humans do. Over the last few decades, it has aided businesses in data analysis and information discovery by automating the time-consuming process and allowing machines to understand human languages. Every day, a vast amount of information is reported using various media such as television, radio, social media, and web blogs in the form of video, audio, or text due to the rapid development of information and communication technology[22][23]. Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken and written. We can say it is related to the field of computer-human interaction. There are various challenges in this field, such as understanding natural language, which allows machines to understand human natural language. The most common natural language processing tasks are discourse analysis, morphological separation, machine translation, natural language generation and understanding, recognition of named entities, part of speech tagging, recognition of optical characters, recognition of speech, and sentiment analysis, automatic summarization, relationship extraction, audio recognition, and topic segmentation are some examples.[23][24].

Natural language processing (NLP) is an application area in computer science, heavily supported by the industry with news applications emerging constantly. This method's purpose is to take a different approach and examine natural language. It explores the fundamental concepts of computer science, machine learning, and statistics that make natural language processing a promising study subject. It investigates the application of general methodologies to specific issues that require the use of natural language. As a result, we take a method-oriented approach to NLP rather than an application-oriented approach for computers to analyze, recognize, and derive meaning from human language efficiently and effectively[22][25].

2.3. Overview of Amharic Language

The Amharic Language is a Semantic language spoken by many parts of Ethiopia and it is the official working Language for the Federal Democratic Republic of Ethiopia and thus has official status[26]. Amharic (አማርኛ) is the world's second most spoken Semitic language, after Arabic, and the official working language of the Ethiopian government[27]. Amharic, unlike Arabic and Hebrew, is written from left to right. Most documents and News in the country are produced in the Amharic language. However, still Amharic Language has few electronic resources on the web and little work has been done for different computer-based applications It has a writing system called Fidel or abugida (ፊደል), which was adapted from the now-extinct Ge'ez language. It has only an African-origin script named Ethiopic Fidel, the script Fidel has 33 orders with their seven forms and it gives 7*33=231 unique Fidel’s (ፊደሎች). Amharic Characters are represented by Unicode in a computer, and Unicode provides a unique number for every single character using the program. Since Amharic is one of the world's "low-resource" languages, it lacks the tools and resources needed for NLP approaches.

2.3.1. Amharic Morphology (ስነ-ምዕላድ)

Morphology studies the pattern of word formations including inflection, derivation, or compound word formation using different affixes to create derivational and inflectional morphemes[28]. A phoneme or collection of phonemes forms a morpheme, which is the smallest meaningful unit in a word[29]. For example, from the noun 'ልጅ' -' child' another noun 'ልጅነት'-'childhood', from the adjective 'ቅርብ' -' close' the noun 'ቅርበት' - 'proximity', from the stem 'ውርድ', the noun 'ውርደት'- 'humiliation', from the root 'ዝም' - 'Shut up', the noun 'ዝምታ' - 'Silence' go; can be derived.

2.3.2. Amharic Word Classes

The Amharic Language has different word-class such as noun, verb, adverb, pronoun, adjective, etc.

❖ Noun (ስም)

A noun is the name of a person, place, thing, or idea. According to[29], Amharic nouns are of two types: basic and derived nouns. (e.g., ወንበር |Chair|, መኪና |Car|, ቤት |Home|). Nouns can be derived from Compound Words (sometimes by affixing the vowels ኧ and ኦ):

✓ Noun + Noun => ቤተ + መንግስት=> ቤተ መንግስት

✓ Noun + Verbal Stems => ልብስ + ማጠቢያ=> ልብስ ማጠቢያ

Nouns can be derived from Nouns by suffixing bound morphemes

✓ Noun + morpheme => ሰበብ+ ኧኛ => ሰበብኧኛ => ሰበበኛ]

❖ Verb (ግስ)

Any word which can be placed at the end of a sentence and which can accept suffixes as /ህ/, /ሁ/, / ሽ/, etc in the Amharic language called ግስ (verb).

Example: አንተ ወተት ጠጣ - ህ |You drank milk|

እኔ ልብስ ገዛ - ሁ |I bought clothe|

❖ **Adjective (ቅፅል)**

Another Amharic word class is ቅፅል (Adjective) any words that qualify a noun or an adverb that comes before a noun e.g., ቆንጆ ልጅ |Pretty girl|, and after an adverb በጣም ቆንጆ |Very pretty|. Another specific property of adjectives is when pluralizing the adjectives is it will repeat the previous letter of the last letter for the word. e.g., ረዥም => ረዝዥም, አጭር => አጫጭር, etc.

❖ **Adverb (ተውሳከ ግስ)**

In Amharic Language, adverbs are used to qualify the verb by adding extra ideas to the sentence. Example በጣም, ሁልጊዜ, ገና, ዛሬ, ቶሎ, ምንኛ, ክፉኛ, and እንደገና etc.

- **Preposition (መስተዋድድ)**

A preposition is a word that can be placed before a noun and perform adverbial operations related to place, time, cause, and so on, which can't accept any suffix or prefix from the beginning of the end of the character and which is never used to create a new word. It includes ከ፣ለ፣ ወደ ፣ ስለ ፣ እንደ etc.

- **Pronoun (ተውላጠ ስም)**

This category further can be divided as a deictic specifier, which includes እሱ, እስዋ, እኔ, አንተ, አንቺ quantitative specifier, which includes አንድ, አንዳንድ,ብዙ, ጥቂት, በጣም and possession specifier such as የእኔ, የአንተ, የእሱ,እነሱ, etc.

2.3.3. Amharic Phrases (ሐረግ)

A phrase (ሐረግ) is a group of words that conveys some meanings but does not make complete sense by itself. It is always a part of a sentence and it is any group of words, often carrying a special idiomatic meaning; in this sense, it is synonymous with expression.

2.3.4. Amharic sentence (አረፍተ ነገር)

A sentence is a group of words that expresses or conveys a complete meaning. The order of words inside Amharic sentences is different from English. Generally, the verb goes at the end of the sentence and the structure is Subject / Object / Verb. It can be a statement that is used to declare, explain, or discuss an event[29]. Example Abebe [subject] eat [verb] his lunch[object], In Amharic አበበ [subject] ምሳዉን[object] በላ[verb] sentence may be incomplete like a phrase to express something. Amharic sentences are classified into two: simple and complex sentences. A simple sentence is a sentence that has only one verb phrase. Whereas a complex sentence is a sentence that is categorized as, a simple sentence and formed by complex phrases.

Generally, Amharic sentence structure is subject-object-verb (SOV). However, sometimes sentences may occur as OSV. Depending on the placement of words in a sentence, the sentence meaning can be changed unless the word (object of the sentence) has the object marker 'ን'/'-n'.

For instance, ‘ጅብ ውሻ ይበላል’|The hyena eats the dog| and ‘ውሻ ጅብ ይበላል’ |A dog eats a hyena| the words used in both sentences are the same but different meaning. ‘ጅብ’ |Hyena| and ‘ውሻ’ |Dog| are the subjects of the sentences in the first and second sentence respectively[29]. Amharic nouns do not have different subject markers or morphemes (affix). However, a subject can be identified from its place in a sentence. ‘-ን’ /’-n’ is a suffix that is used as a sign for Amharic objects.

2.4. Sentiment Analysis

One of the most important things to transferring and interchanging ideas between peoples is through sharing their idea, which says opinion. People can express their ideas or opinions in the form of posting on social media like Facebook, Twitter, and Forum discussions concerning their day-to-day life. The growth of the information from social media makes sentiment analysis a relevant field to find the opinions of others. Sentiment analysis is the field of study that examines human opinions, sentiments, appraisals, evaluations, emotions, and attitudes towards entities for example services, products, individuals, organizations, issues, topics, events, and there attributes [5].

Sentiment analysis is a text classification technique that works with subjective statements. Because it processes opinions to learn about public perception, it is also known as opinion mining [30]. Opinion mining, opinion extraction, sentiment mining, subjectivity analysis, effect analysis, emotion analysis, review mining, are all synonyms for sentiment analysis with slightly different task. They are currently all included together under the heading of sentiment analysis or opinion mining [31].

The terms sentiment analysis and opinion mining are interchangeable throughout our document. Sentiment analysis combines various research areas such as natural language processing, data mining, and text mining, and is quickly becoming of critical importance to organizations as they strive to integrate computational intelligence methods into their operations and attempt to shed more light on, and improve, their products and services. The goal of sentiment analysis, also known as opinion mining (SAOM), is to discover people's written opinions (text). The sentiment is defined as "how one feels about something", "personal experience, one's feeling", "attitude toward something", or "an opinion" [32].

In general, SA is concerned with the extraction of data relating to a group's thoughts or opinions on a specific topic. Sentiment classification, feature-based sentiment classification, and opinion summarization are the main areas of research in the field of Sentiment Analysis and opinion mining. Sentiment classification is the process of categorizing entire documents, texts, or reviews based on how people feel about certain objects. Feature-based Sentiment classification, on the other hand, takes into account people's feelings about specific objects' features. Opinion summarization differs from traditional text summarization in that it mines only the product features on which customers have expressed their opinions [33].

2.4.1. Types of opinion

According to[13][34] opinion can be classified as regular and comparative and we can classify opinions based on how they are expressed in text as implicitly and explicitly.

Regular and Comparative opinion

- **Regular opinion:** In the literature, a regular opinion is often referred to simply as an opinion, and it has two major sub-types.
  - ***Direct opinion*:** A direct opinion is expressed directly about an entity or an aspect of an entity[13], such as ኮሮና ቫይረስ ነው | It is a corona virus |.
  - ***Indirect opinion*:** An indirect opinion is expressed indirectly about an entity or aspect of an entity based on how it affects other entities[35]. For example, the sentence ይህን መድሃኒት ከወሰድኩ በኋላ, ራስ ምታቴ ይጨምራል | After taking this drug, my headache rises | the above sentence describes an undesirable effect of the drug on "headache", which indirectly gives a negative opinion or sentiment to the drug. In this case, the entity is the *drug* and the aspect is the *rise in headache*.
- **Comparative opinion**: A comparative opinion indicates a connection of similarities or differences between two or more objects, as well as the opinion holder's preference based on some of the entities common characteristics[34]. For example, the sentences, የ ኮሮና ራስ ምታት ከጉንፋን አንጻር በጣም ከባድ ነው | Corona headache is more severe than flu| and ኮሮና ከባድ ራስ ምታት አለው | Corona has a severe headache| express two comparative opinions.

Explicitly and Implicitly

- **Explicit opinion**: An *explicit opinion is* a subjective statement (statements that express one's own opinion about something) that gives a regular or comparative opinion[13], for example, የ ኮሮና ራስ ምታት ከጉንፋን አንጻር በጣም ከባድ ነው | Corona headache is more severe than flu| and ኮሮና ከባድ ራስ ምታት አለው | Corona has a severe headache|.
- **Implicit (or implied) opinion**: An *implicit opinion* is a statement that implies a regular or comparative opinion based on an objective statement (statements that express a factual opinion)[13]. Such an objective statement usually expresses a desirable or unfavorable fact, such as ኮሮና ቫይረስ ዓለም አቀፍ ወረርሽኝ ነው | *Coronaviruses is a global pandemic* |.

2.4.2. Component of opinion mining

Opinion mining is composed of three main components as suggested by [5].

I. **Opinion orientation:** Any positive, negative, or neutral opinion expressed on a specific object is referred to as opinion orientation[14]. Example: እንደ አቶ አለሙ ገለጻ ከአፕል ኩባንያ የገዙት ስልክ አስደናቂ ነው| According to Alemu, the phone that he bought from Apple Company is amazing. The above statement stated above describes that the apple phone is amazing. Hence, the word "amazing" in this statement shows sentiment orientation, which is positive.

II. **Opinion Holder:** an opinion holder is a person that is a writer or an author of the opinion that express its feeling towards some particular object, company, product, or idea. From the above example, Alemu is an opinion holder or author of the opinion[13][5].

III. **Opinion Object:** is an object on which writers express an opinion, an opinion object may be a product or service. In the above example, a phone is a product on which an opinion is expressed[13].

2.4.3. Sentiment classification levels

Depending on the level of detail required, sentiment analysis can be conducted at three different levels[34]:

I. Document-level classification

This is the simplest of the other classification approach in which the whole document is considered as a single unit of information (talking about one topic). This level aims to classify the entire document depending on classes, which can be positive or negative i.e. given a product review, the task is to determine whether it expresses positive or negative opinions about the product. In this level of sentiment classification, the document must be able to contain the opinion about a single object only such as book, film, hotel thus it’s not applicable for a document having more than one entity[34][8][5].

II. Sentence level classification

This level of analysis is very close to subjectivity classification, and the task is limited to the sentences and their expressed opinions at this level. This level specifically determines whether each sentence expresses a positive, negative, or neutral opinion. Sentiment classification polarity is calculated at the sentence level for each sentence because each sentence is considered a separate unit with different opinions. Sentence level sentiment classification, have two main tasks: subjectivity analysis (subjectivity detection) and sentiment classification. Statements can be one of the two types either fact (objective statement) or opinion (subjective statement). For sentiment analysis, we only need the subjective statement so in sentence-level classification before we classify the document into positive, negative, or neutral classes we require to perform subjectivity analysis or detect a given sentence whether it is subjective or objective. Then after having the subjective statements, we proceed to the second step, which is sentiment classification. In sentiment classification, we classify the subjective statements into positive, negative or neutral classes[8][34][14].

III. Aspect level classification

Aspect-based sentiment analysis is also known as a feature or attribute-based sentence analysis. It is used to analyze different features/attributes/aspects of the product. In aspect level classification, we classify the sentiment concerning the given aspect of entities. Different people may have different opinions for a different aspect of the same entity. Aspect level classification is not only focused on classifying the sentence into positive or negative it also states the aspect of the entity in which the opinion is about. For example, የጁፒተር ሆቴል አልጋ በጣም ምቾት አለው | The Jupiter Hotel bed is very comfortable |. In the above example, the entity is ጁፒተር ሆቴል |Jupiter Hotel |the aspect for which opinion given is አልጋ | Bed |and the positive sentiment is በጣም ምቹት | Very comfortable|. In many applications, entities and/or their different aspects describe opinion targets. Thus, the goal of this level of analysis is to notice sentiments on entities and/or their aspects[5][10].

2.4.4. Steps to analyze sentiment data

According to[5] sentiment analysis is a complex task, which is implemented through five steps.

1. Data collection

Data collection is the first and core task of sentiment analysis in which data used for the analysis task are collected from different sources such as blogs, forums, social networks, and the like. The data gated from the above sources are user-generated content, which is disorganized, has different writing styles, and so on, therefore, we need to extract and classify the data using text analytics and natural language processing. Everything that follows will be determined by the quality of the data gathered and how it has been annotated or labeled[13].

Data sources for sentiment data

The data required for sentiment analysis should be specialized and in large quantities. The most difficult aspect of the sentiment analysis training process is not finding large amounts of data, but rather finding relevant datasets. These data sets should cover a broad range of sentiment analysis applications and use cases[36].

- **Blogs**

Blog pages and blogging are becoming increasingly popular as Internet usage grows. The names associated with the universe of all blog sites are referred to as the blogosphere. Blogs are used to express one's personal views on any product or topic. On a blog, people like to share their thoughts, ideas, and suggestions with others. Blogging is a common occurrence due to the ease with which blog posts and reviews can be created, the fact that it is free form, and the fact that it is unedited. Many of the studies involving sentiment analysis use blogs as a source of opinion.

- ❖ **Review sites**

The opinion of others is an important factor for any user when making a purchasing decision. The majority of user-generated reviews and suggestions can be found on the Internet. Product or service reviews are available in the form of unstructured opinions. Sentiment classification studies use reviewer data collected from e-commerce websites such as www.yelp.com (restaurant reviews), www.amazon.com (product reviews), and www.flipkart.com (product reviews), which host millions of products reviewed by customers.

- ❖ **Data sets**

Many works in the field classify films based on user reviews. The Multi-Domain Sentiment (MDS) contains various types of product reviews taken from Amazon.com and Flipkart.com, such as Books, dresses, Kitchen Appliances, and Electronics items, with numerous positive and negative suggestions/reviews for each territory.

- ❖ **Micro-blogging**

Micro-blogging is a popular communication tool among internet users. Every day, a large number of messages appear on microblogging websites such as Twitter, Tumblr, and Facebook. Twitter is a popular microblogging service where users send messages known as "tweets." These Tweets are used to express their thoughts/suggestions on various topics. These Twitter messages are sometimes used as a data source for Sentiment Classification.

- ❖ **API Data**

Data for social media can be uploaded using Live APIs. A news API can help to gather information from various news publishers, whereas a Facebook API can provide all of the publicly available data you require from its platform. Open-source repositories such as Kaggle, Amazon reviews, or Yelp can also be used.

- ❖ **Manual**

If we already have data from a CRM (Customer Relationship Management) tool, we can manually upload it as an a.csv file to the sentiment analysis API.

2. **Text preprocessing**

The process of cleaning and preparing the text for classification is known as pre-processing the data. Noise and uninformative sections such as HTML tags, scripts, and ads are common in online writings. Furthermore, many words in the text have little bearing on the overall direction of the document. Keeping those words increases the problem's dimensionality, making classification more difficult because each word in the text is treated as a separate dimension. The hypothesis of properly pre-processing the data is as follows: reducing noise in the text should help improve the performance of the classifier and speed up the classification process, allowing for real-time sentiment analysis[37].

Pre-processing techniques

- **Sentence segmentation:** The process of breaking down a text document or corpus into individual sentences is known as sentence segmentation, also known as sentence boundary detection. This aids in identifying word boundaries, allowing for further processing of each sentence. Using a sentence tokenizer, segmentation is performed whenever a full stop or punctuation occurs[38][39].

- **Change to lower case:** Text is typically made up of abbreviations and all capital letters. This step is frequently overlooked, but it is one of the simplest and most effective steps of text preprocessing, especially when the dataset is significantly sparse. It has been discovered that variations in capitalization produce different results. This means that the computer treats the same words with one in capitals and one in lowercase as two distinct worlds, and two distinct word vectors are formed in the later stages of word embeddings. Thus, in-text pre-processing, making all words lowercase has been the best practice[40][39].

- **Tokenization:** Tokenization is the process of separating a stream of text into words, phrases, symbols, or other meaningful elements known as tokens. The goal of tokenization is to explore the words in a sentence. The token list is used as input for further processing such as parsing or text mining. Tokenization is useful in both linguistics (as a type of text segmentation) and computer science (as part of the lexical analysis)[39][38].

- **Stop Word removal:** Except in a few cases, words like "the," "our," "is," "and," and so on have no meaning in Natural Language Processing. For example, the Text or document classification use case does not give these extra words any weightage. Only the keywords used to create the topics are extracted. As a result, the more stop words that are identified and cleaned up, the better the results of classification algorithms. It is also worth noting that in certain use cases, such as conversational models, the use of negation words such as "No," "can't," "won't," and "not" is critical in determining the context of the sentence and its intent[38].

- **Stemming:** Stemming is the process of combining a word's variant forms into a single representation, the stem. For instance, the words "presentation," "presented," and "presenting" can all be reduced to a single representation "present." This is common information retrieval (IR) text processing procedure based on the assumption that posing a query with the term presenting implies an interest in documents containing the words presentation and presented[39][40].

- **Lemmatization:** Lemmatization is the process of removing or replacing a word's suffix to bring it back to its base, which is known as the lemma. Unlike stemmed words, the lemma is always a meaningful word. Lemmatization is a popular text preprocessing step in Natural Language Processing that has proven to produce excellent results[38].

3. Sentiment detection (subjectivity analysis)

Any text in this world can befall into either of the two group's facts (objective statement) or opinions (subjective statement). Unfortunately, in the case of sentiment analysis, we only need the opinion part so sentiment detection can identify facts from opinion. Additionally, we also identify texts that are not related to sentiment analysis[13][12].

4. Sentiment Classification

Sentiment classification is an automated technique that recognizes thoughts in text and categorizes them as positive, negative, or neutral based on the emotions expressed**.** Then after having valuable benchmarks, those subjective statements identified in step three are further classified into their respective class. Different sentiment analysis approaches such as machine learning, lexicon-based, and hybrid approaches are used to perform the sentiment classification task[5].

5. Output description

After the classification step over the test results are displayed on graphs such as pie charts, bar charts, or line graphs once the analysis is completed.

2.4.5. Subjectivity analysis

Subjectivity analysis is an important subtask of sentiment analysis because it ensures that only opinionated information is passed on to the polarity classifier and that factual information is filtered out[13][41]. The factual (objective) statement usually contains a fact about given information thus there is no opinion included in it. And opinionated (subjective) statements most of the time contain personal opinions or feelings, views, judgments, or beliefs about given information. A subjective sentence can express may types of information e.g., opinion, evaluation, emotion, beliefs, speculations, judgments, allegation, standalone problem.

Subjectivity analysis is divided into two parts: extracting subjective clues such as terms, phrases, or expressions and then using them to classify the corresponding sentence (or document) as subjective or objective. The second task is known as subjectivity classification, and it is typically the first process in most sentiment analysis applications. It determines whether a given text expresses an opinion or not. Subjectivity analysis is performed at the document, sentence, phrase, and word levels. At the word or phrase level, it is commonly referred to as sentiment extraction. Because a document is a collection of sentences, its subjectivity is heavily reliant on those sentences. As a result, subjectivity analysis at the sentence level is unavoidable in some applications. For example, in applications such as sentiment summarization, opinion question answering, and information extraction, it is necessary to distinguish between subjective and objective sentences. While subjective sentences are important in sentiment summarization and opinion question answering, objective sentences containing facts are important in information extraction[42].

To express their opinions, people use subjectivity indicators such as adjectives, nouns, adverbs, and sentimental verbs. These are known as opinion terms. Statistical analysis reveals that the usage patterns of these opinion terms vary across domains and contexts. While adjectives, nouns, and adverbs are more descriptive for expressing opinions about products, verbs are more useful for expressing feelings about social issues[41].

2.4.6. Challenges in sentiment analysis

As the author of[14][43] tries to describe sentiment analysis face several challenges and these challenges are the reason why the accuracy of most sentiment analysis work decrease. The challenges are:

Identify comparison words

Identifying comparison words and whether they are giving positive or negative feedback depends on their context. Therefore, it is not an easy job as sometimes good are bad and bad are good.

Different People Different Writing Style

The fact that comments or views are entered by people who are different from each other in the way they write, their use of language, abbreviations, and their knowledge is a challenge on its own. People also do not express an opinion in the same way. One might use certain negative terms in a sentence text that appears in an online newspaper and that which appears in an online forum is widely different. The mining of online forums and discussions is a challenge on its own. Some possible reasons include the use of abbreviations, the entry of comments by different people, who differ in the way they write, or in the knowledge of the language they use[22].

Opinions Change with Time

Another challenge occur in the issue is being able to monitor opinions changing with time. This helps us to observe if a certain product gets improved with time, or people change their opinion about a product and get convinced for it with time[33].

Strength of Opinions

Identification of the strength of an opinion is another challenge faced in opinion mining. The strength of an opinion can change as the discussion progresses in a forum, i.e. arguments used during the discussion are strong enough to change the strength of opinions[22].

Misleading Opinions due to sarcastic and ironic statements

There are sarcastic and ironic sentences in the text. Positive words can metaphorically have a negative sense of usage in such a situation. As a result, text in a statement can be difficult to identify as sarcastic or ironic, leading to incorrect orientation and misleading opinion mining[22][43].

Sentences with a mixed view

When people express both positive and negative opinions in the same sentence, opinion mining becomes more difficult. This is especially true when people communicate through informal channels such as blogs and forums. People are more likely to combine opposing viewpoints in the same sentence. Such sentences can be difficult to parse for opinion mining[43].

2.4.7. Approach to sentiment analysis

Sentiment analysis is an automated way to study opinions given by people about any world entity or any target. The target may be organization, product, movies, politics, social issues, and the likes. Therefore, due to the vast number of users of internet opinion given to this target is an increase in such a way that manual analysis is impossible. For this reason, an automated way of analyzing this opinion is demand [5][34].

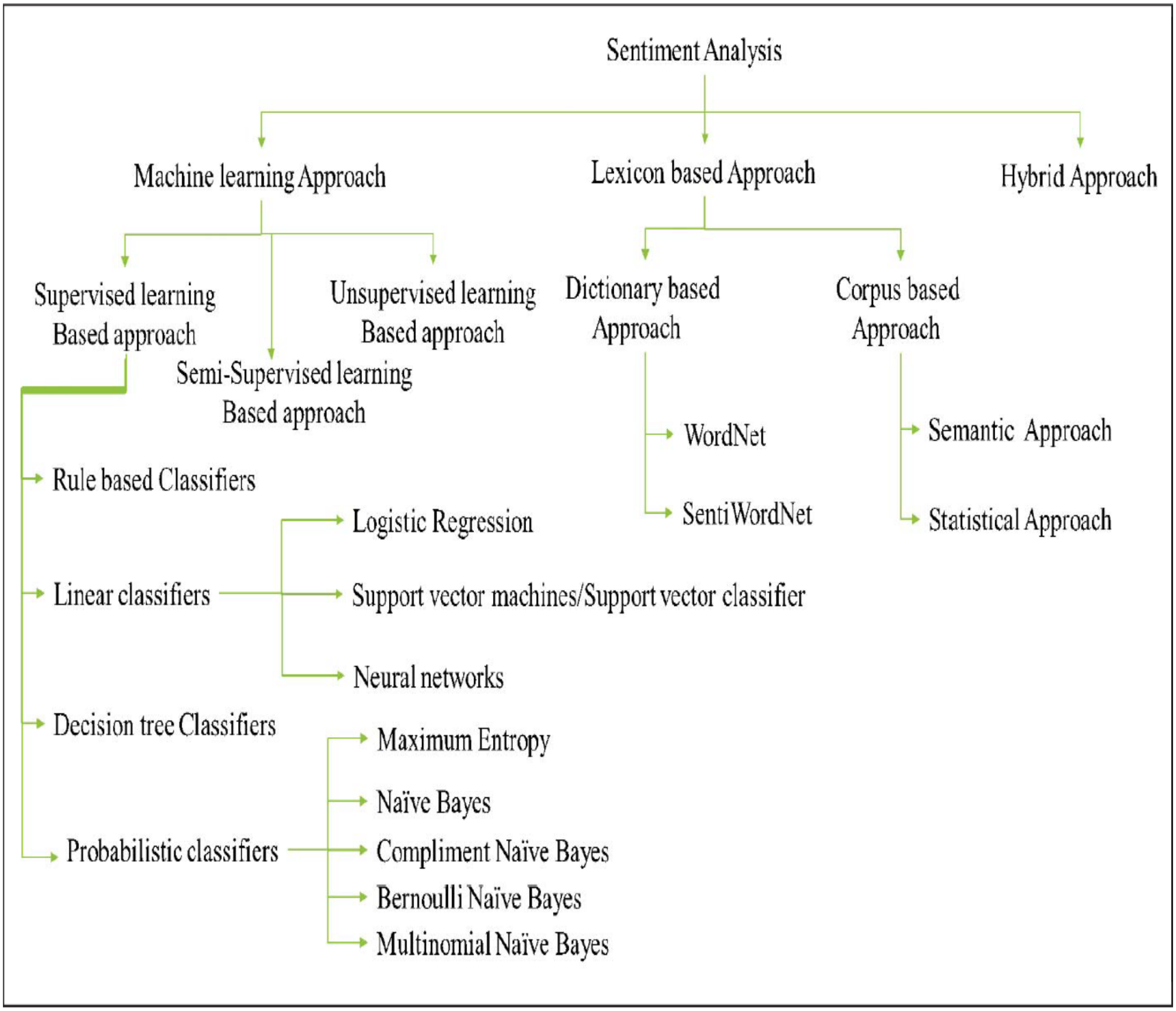


Figure 1 Classification approach to sentiment analysis[44]

As the author of[35] describes in the figure above the lexicon-based approach, machine learning-based approach, and hybrid approach combining both lexicon-based and machine learning approaches are the three main approaches for categorizing sentiments or opinions into negative, positive, and neutral categories.

Machine learning-based Approach

Text features are extracted and used to perform classification in the machine learning approach. Because of its ability to handle a large amount of data online and because it is automatic, it is more popular than other techniques. In the machine learning approach, there is a prior training of our dataset we need two datasets of document: training dataset and test dataset. The training dataset is used to build the model, which learns the different characteristics of a given document. Then we use the test data set to validate the built model. The machine learning techniques give more accuracy to sentiment classification. Unfortunately, the performance of the machine learning approach is heavily dependent on the features chosen, the quality and quantity of training data, and the domain of the dataset, and it generally necessitates the creation of large expert annotated training corpora from scratch, specifically for the application at hand, and may fail if training data are insufficient. There are three types of machine learning approaches: supervised, unsupervised, and semi-supervised[35].

- ❖ **The supervised machine learning approach**

Supervised machine learning techniques make up a large portion of the machine learning techniques used for sentiment classification. They are commonly used for classification. The data is divided into two sets in this case: one is a labeled training set and the other is an unlabeled test set data. The model is trained on the training set before being validated using the test set of data. Support Vector Machines, Nave Bayes, Random Forest, Logistic Regression, Decision Trees, and other important algorithms fall into this category. Because these techniques are highly dependent on training data, the accuracy may be greatly influenced if there is a case of incorrectly labeled data in the training set or fewer data taken in the training set. The main goal of this method is for the algorithm to "learn" by comparing its actual output to the "taught" outputs to determine polarity. The patterns approach is used in supervised learning to predict label values on additional unlabeled data[45] [44][46].

- ✓ **Naive Bayes**

The Naive Bayes algorithm employs a probabilistic approach. In supervised machine learning, these algorithms are used. The primary goal of the Naive Bayes algorithm is to train a classifier. This train classifier is then used to classify test data. This algorithm's operating principles are shared by all processes. These working principles are as follows: classifiers trained for classification problems simply assume that the value of a single feature is independent of the value of any other feature, given the class variable. The Naive Bayes algorithm is the most basic and widely used algorithm for train classifiers. The Naive Bayes classification model calculates the probability of a class based on how frequently words appear in documents. The model employs BOWs feature extraction, which disregards the position of the word in the document. It uses the Bayes theorem to predict the probability that a given feature set belongs to a particular label [44][47].

Bayes theorem is defined as:

$$P(C/D) = \frac{P(D/C)\,P(C)}{P(D)}$$

D is a review and C is class. For a given textual review D and a class C (positive, negative), the conditional probability for each class given a review is P(C/D)[36].

- **Maximum Entropy Classifier (ME)**

The Maxent Classifier (also known as a conditional exponential classifier) uses encoding to convert labeled feature sets to vectors. This encoded vector is then used to compute weights for each feature, which are then added together to determine the most likely label for a feature set. This classifier is parameterized by a set of X{weights}, which are used to combine the joint features generated by a X{encoding} from a feature set. The encoding, in particular, maps each C {(feature set, label)} pair to a vector. The probability of each label is then calculated using the equation below[44]:

$$P(fs/label) = \frac{\text{dotprod (weights, encode (fs, label))}}{\text{Sum (dotprod (weights, encode(fs,l)) forlinlabels)}}$$

Kaufmann[40] used the ME classifier to detect parallel sentences between any language pairs with limited training data. Other tools for automatically extracting parallel data from non-parallel corpora use language-specific techniques or necessitate large amounts of training data. Their findings demonstrated that ME classifiers can produce useful results for nearly any language pair. This opens the door to the development of parallel corpora for many new languages.

Support vector machine (SVM)

The linear properties of probability are used by the support vector machine. The main idea behind SVMs is to find linear separators in the search space that can best separate the various classes. Because of the sparse nature of the text, few features are irrelevant; however, because they are correlated with one another and generally organized into linearly separable categories, text data are well suited for SVM classification. SVM can create a nonlinear decision surface in the original feature space by nonlinearly mapping the data instances to an inner product space where the classes can be separated linearly with a hyperplane. Put, SVM signifies cases as factors in the area which are planned to a high-dimensional area where the planned cases of individual sessions are separated by as large as a possible peripheral range to the hyperplane. News cases are planned into that same area, and based on which part of the hyperplane they are placed in, they are expected to fit in with a certain category. SVM hyperplanes are completely established by a relatively small portion of the training conditions known as support vectors. The qualified classifier is unaffected by the relaxations of the exercising data. SVM has been applied efficiently in text classification and a large range of series handling programs[48][44].

The idea for SVM is to find a boundary (known as a hyperplane) or boundaries that separate classes of data. SVM does this by taking a set of points and separating those points using mathematical formulas as it is shown in Figure 2 there are 2 classes x, o and there are 3 hyperplanes A, B, and C. Because the normal distance between any of the data points is the greatest, hyperplane provides the best separation between the classes, representing the maximum margin of separation[44].

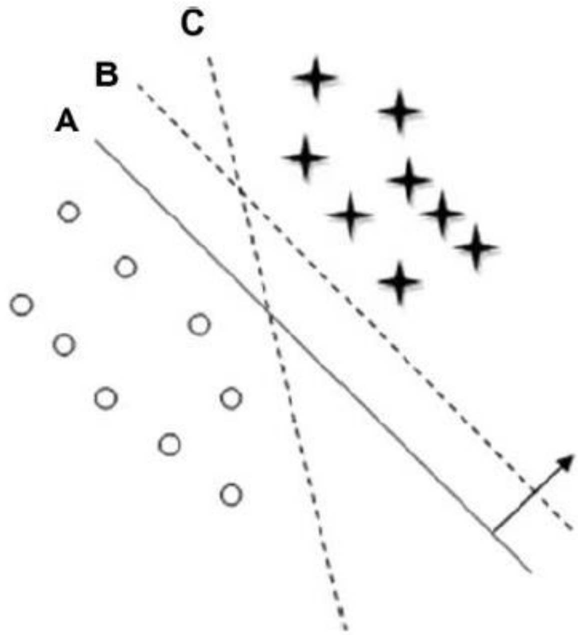


Figure 2 Using support vector machine for classification [44]

SVMs are used in many applications, among these applications are classifying reviews according to their quality is the one. Chen and Tseng[49] have used two multiclass SVM-based approaches: One-versus-All SVM and Single-Machine Multiclass SVM to categorize reviews. They proposed a method for evaluating the quality of the information in product reviews considering it as a classification problem. They also adopted an information quality (IQ) framework to find an information-oriented feature set. They worked on digital cameras and MP3 reviews. Their results showed that their method can accurately classify reviews in terms of their quality. It significantly outperforms state-of-the-art methods.

SVMs were used by Li and Li[50] as a sentiment polarity classifier. Unlike the binary classification problem, they argued that opinion subjectivity and expresser credibility should also be taken into consideration. They proposed a framework that provides a compact numeric summarization of opinions on micro-blogs platforms. They identified and extracted the topics mentioned in the opinions associated with the queries of users, and then classified the opinions using SVM. They worked on twitter posts for their experiment. They found out that the consideration of user credibility and opinion subjectivity is essential for aggregating micro-blog opinions. They proved that their mechanism can effectively discover market intelligence (MI) for supporting decisionmakers by establishing a monitoring system to track external opinions on different aspects of a business in real-time.

- **The unsupervised machine learning approach**

Unsupervised learning techniques, unlike supervised learning approaches, do not use a labeled set or training set of data to train the model, so they are useful when labeling the input data is difficult. However, for the model to learn, a large amount of data is required; otherwise, the model may produce incoherent results. Unsupervised learning is the process of learning patterns in input when no specific output values are provided, implying that the learner receives only an unlabeled set of examples. Unsupervised methods can also be used to label a corpus that will be used for supervised learning later on. K-means clustering, KNN (k-nearest neighbors), Hierarchal clustering, Apriori algorithm, are some of the popular examples of unsupervised machine learning algorithms [51][35].

- **Semi-supervised machine learning**

Semi-supervised learning methods combine the benefits of both supervised and unsupervised learning methods. In these, the model learns using both labeled and unlabeled data. These techniques were developed in response to a lack of labeled data and to address a shortcoming of the unsupervised learning approach by incorporating some prior knowledge into the unsupervised model[44],[52].

Lexicon based Approach

In this approach, we use a sentiment lexicon, which contains sentiment words with their sentiment polarity. This approach determines the sentiment of a given word in a document by matching the word with the sentiment lexicon. The lexicon-based approach does not need any prior training of the dataset. The semantic orientation of these words, whether positive or negative, is already determined. It computes the sentiment score for each document by aggregating the sentiment polarity of all opinion words in the document. The documents are then categorized as either positive or negative. The positive document has a larger positive word lexicon, whereas the negative document has a larger negative word lexicon. Lexicon-based approaches do not require time for training data and most domains lack labeled training data, in this case, lexicon-based approaches are very useful for developing applications. However, the performance of the lexicon-based approach is dependent on opinion words included in the dictionary, the polarity of many words is domain and context-dependent. Dictionary-based lexicon construction, and corpus-based lexicon construction are methods for creating a sentiment lexicon[53], [54].

- **Dictionary-based lexicon construction**

This method is based on extracting the opinion seed from the text data and searching the dictionary for its antonyms and synonyms.

Initially, a seed list is created by manually selecting opinion words that are difficult to find due to limited context-oriented text, and thesaurus and dictionaries are then searched to determine their antonyms and synonyms, and the synonyms are then included in the list of seed words, and the process is repeated. The basic policy of dictionary-based techniques is first it identifies the polarity of the word after that finds it in the dictionary and its synonym and antonym and known as collection WorldNet or word finder. It follows the iterative process after finding a new word then adding it in stem words. Repeat this iterative process until not search for new words. Then manual inspections continue for error correction[35], [54] .

- **Corpus-based lexicon construction**

The corpus-based approach is concerned with the construction of a list of seed opinion words, which is then expanded using information from the corpus text. There is no problem with limited context-oriented text information because the corpus contains a pool of text information that is mostly present in the specific domain. This can be accomplished through the use of a semantic approach as well as statistical approaches. The corpus-based approach performs relatively well in terms of accuracy by addressing the problem of finding specific opinion words with specific orientations; this is a significant advantage of the corpus-based approach that the dictionary-based approach lacks[54], [55], [35].

Hybrid approach

To improve overall performance, the hybrid approach combines methods from both the lexicon-based approach and the machine learning-based approach. According to some research techniques, a hybrid approach improves sentiment classification performance[35]. This author combined these two techniques in concept-level sentiment analysis and achieved high accuracy[56].

2.5. Related work

Many works have been done in the area of sentiment analysis in different languages like English, Chinese, Arabic, and in local languages like Amharic, Afaan Oromo, Tigrigna and so on. To perform the sentiment analysis task, different researchers, used one of the three approaches: lexicon, Machine learning, or hybrid for making analysis. In the related work section, we are going to see the work of different scholars based on the approach they used.

Sentiment analysis based on lexicon Based approach

As far as we know, the first work on sentiment analysis of Amharic texts is done by[57] they propose a sentiment mining model for opinionated Amharic texts. 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. They achieved 86% and 58% accuracy for positive and negative respectively. However, the research has some limitations; first, the system uses a user-defined dictionary, which contains only 303 opinion terms. Since the system considers only terms which are found in the dictionary, some important opinionated terms might be ignored which results 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.

According to[14], they propose sentiment classification in the Amharic language at the sentence level. They apply lexicon (dictionary-based) approaches. Their research is conducted based on 800 data gathered from social media Facebook without the restriction of the domain they classify the sentiments as positive, negative, and neutral. The results of the experiment show the performance of the method has 92% precision and 82% recall for positive class, for negative class 94% precision and 96% recall whereas 90 % recall and 90 % precision for neutral class achieved.

According to[17] feature level opinion mining model for Amharic texts have been performed. The objective of the study was to determine an opinion on features of the domains. In this study, the author first extracted features of the domain and then determined the opinions in the extracted features by employing some rules. An opinion word in the sentence was detected from Amharic general-purpose opinion word lexicons that contain 1001 sentiment words, which are 578 negative and 423 positive words. The author collected 484 Amharic reviews manually from the hotel, university, and hospital for experimental activities. In this study the researchers performed two experiments: the first experiment is based on the general linguistic rule regarding position of the adjective words. Unlike adverbs, which often seem capable of popping up almost anywhere in a sentence, adjectives nearly always appear immediately before the noun that they modify. The second experiment is based on the stated general rule and an adjective which comes after a noun to modify it. From experiment one, the author got the result of an average precision of 95.2%, an average recall of 26.1% for feature extraction and an average precision of 78.1%, average recall of 66.8% for opinion word determination. From experiment two, an average precision of 79.8%, an average recall of 34% for feature extraction, an average precision of 80%, an average recall of 92.7% for opinion word determination. The strength of this study was to determine an opinion on features of the domain in the review sentence. However, the author used only adjectives, as sentiment words to determine the opinions of the review sentences, but sentiment words are not only adjectives but also include adverbs, verbs, and nouns. In addition to this, the sentiment words are not sufficient.

Sentiment analysis based on machine learning approach

The researchers of[15] propose document-level multi-scale opinion mining in the Amharic language in social media, Product marketing & News domain they collect 608 reviews. They employed a Naïve Bayes machine-learning algorithm and used unigram, bigram, and hybrid variants as features. They achieved accuracy of 43.6%, 44.3%, and 39.5% for unigram, bigram, and hybrid language models, respectively for classifying the reviews into five classes (very positive, positive, neutral, negative, very negative). They recommend that the accuracy can be further improved by building larger data sets. In addition, they recommend including subjective analysis, which classifies the review first into subjective and objective, which can reduce the effort to be devoted to building sentiment corpus by avoiding manual identification of texts.

The researchers in[58] proposed Multi-Class Sentiment Analysis from Afaan Oromo text. In this work, sentence-level sentiment analysis is done to classify the statements into five multiple classes very negative, negative, neutral, positive, and very positive. Here, they proposed two methods Support Vector Machine and Random Forest supervised machine learning approaches to classify sentiment polarity from Oromia broadcasting network (OBN) Twitter page by Ethiopian language Afaan Oromo. They used Tf-IDF for feature extraction and as input to machine learning algorithms. They collected about 1810 data from Oromia broadcasting network (OBN) Twitter page. The performance of the proposed approaches shows that Support Vector Machine and Random Forest achieved an accuracy of 90% and 89% respectively. They recommend aspect-level sentiment analysis on Afaan Oromo and apply the hybrid approach.

Studied by[59] performed on sentiment analysis for classifying Amharic opinionated text into positive, negative, or neutral by using machine learning approach on data collected from ERTV, Fana broadcasting, and diretube.com domains. The author employed three machine learning classification techniques (Naïve Bayes, Multinomial Naive Bayes, and Support Vector Machines). The author used n-grams presence, n-grams frequency, and n-grams-TF-IDF features selection methods. The experiments are conducted using 576 Amharic opinionated texts collected from SERTA, Fana Broadcasting, and diretube.com manually. The Experiment indicates that uni-grams term frequency feature selection methods perform the best for all algorithms (Support Vector Machine, Naive Bayes, and multinomial Naïve Bayes). Based on their relative performance of classification, Support Vector Machine registered 78.8% accuracy outperforming the others, Naïve Bayes with 77.6%, and multinomial Naive Bayes with 74.7%. As shown from their result SVM performed better than NB and MNB algorithms.

A Study by[16] works on the title of opinion mining from Amharic entertainment text using machine learning approaches (Naïve Bayes, Decision Tree, and Maximum Entropy). The experiment was conducted using 616 Amharic opinion texts. The study obtained 90.9 %, 83.1%, and 89.6% using Naïve, Bayes, Decision Tree, and Maximum entropy algorithms respectively. However, the study did not control negation, because the study uses uni-gram as a feature for classification. The result only shows positive and negative polarity but it did not include neutral.

Sentiment analysis based on a hybrid approach

The researcher of[60] proposed a sentence-level Amharic sentiment analysis model using a combined approach of lexicon-based and machine learning that automatically extracts opinions from Amharic text. The data set collected for the experiment consists of 600 movie comments whose lengths are not more than three sentences. The comments are collected from the Ethiopian movie sites and social networks. They used combined approaches consisting of two methods where first, the lexicon-based method is used to classify sentences into positive and negative categories. Second, the resultant classified sentences are used as a training set for the machine learning method, which subsequently classifies some other sentences. They achieved 82.5% accuracy for the combined approach. They recommend building a huge corpus for training purposes and building a lexicon of polarity terms should improve the performance of the proposed approach.

A study by [10] proposed a work to design aspect/feature-level opinion mining from Amharic news texts. They have collected 1200 data from the Amhara Mass Media Agency Facebook page. They employed crafted rules using rule-based for labeling data and a supervised approach to training and testing the data. Support Vector Machine and Naive Bayes classifiers were used to classify text into positive, negative, and neutral sentiment classification. In their experiment, they showed that bag of words module feature extraction methods performs the best in both. The result showed as Naïve Baye's precision, recall, and F-measure evaluation metrics 84%, 80%, and 81 % respectively. For SVM precision, recall and F-measure evaluation metrics are 87%, 82%, and 84 % respectively. They do not include subjective or objective classification in their work. They only focus on data in the text format also; idioms are not included in their work. Their research is limited to the extraction of explicit features. They recommend works on extracting implicit features also.

However different scholar performed different study to make a significant contribution in the development of Amharic sentiment analysis, as far as our knowledge is concerned none of them included subjectivity analysis in their work thus affecting the accuracy of the classification, and increasing computation time. Therefore, in this work, we include subjectivity analysis to identify facts from opinion before the classification task is performed. Table 1 shows us summary of related work.

Table 1 Summary of related work

Author	Title	Methods	Class	Dataset	Accuracy
Selama Gebremeskel [57]	Sentiment mining model for opinionated Amharic texts	Rule based	2 (Positive and negative)	Movie and news (303 corpus size)	86% for Pos 58% for Neg
Yodit Teshome [14]	Sentence level opinion mining for amharic language	Lexicon (Dictionary-based)	3(positive, negative and neutral)	Facebook (800 corpus size)	87% for Pos 94% for Neg 90% for Neu
Tulu Tilahun [17]	Opinion mining from amharic blog	Lexicon	2 (Positive and negative)	Manually collected (101 corpus size)	85%
Wondwossen Philemon[15]	A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts	Naïve Bayes	5(Very Positive, Very Negative, Neutral, Positive, Negative)	Facebook, Twitter, DireTube and Ethiopian Reporter websites (608 corpus size)	43.6% for Unigram 44.3 for Bigram 39.5% for Hybrid
Negessa Wayessa [58]	Multi-Class Sentiment Analysis from Afaan Oromo Text Based On Supervised Machine Learning Approaches	SVM RFC	5(Very Positive, Very Negative, Neutral, Positive, Negative)	OBN Tweeter (1810 corpus size)	90% for SVM 89% for RFC
Mengistu Kassa [59]	Sentiment analysis for classifying Amharic opinionated text	NB MNB SVM	3(Positive, Negative and Neutral)	SERTA (576 corpus)	77.6% for NB 74.7% for MNB

					78.8% for SVM
Chilote Dessalew [10]	Public sentiment analysis for Amharic news	Lexicon and Machine Learning	2 (Positive and Negative)	Movie (600 corpus size)	82.5%

Chapter Three: Methodology

3.1. Introduction

In the field of sentiment analysis, the problem of identifying subjective sentences and extracting sentiment of a given text can be tackled using different approaches. Some of the notable ones are machine learning approach, lexicon approach and hybrid approaches. The core objective of multi-class subjectivity detection and sentiment analysis using machine learning approach is to provide a tool that helps to detect subjectivity of a give text and based on the output of the subjectivity detection performing multi class sentiment analysis. To achieve this objective, the architectural design of sentence level sentiment analysis system, the components of sentiment analysis system and the algorithms used for subjectivity detection and sentiment term detection in sentence level sentiment analysis are discussed in this Chapter.

3.2. System architecture for proposed multi-class subjectivity detection and sentiment analysis

The proposed multi-class subjectivity detection and sentiment analysis system has five main components. These are

- ✓ Data collection
- ✓ Text preprocessor
- ✓ Feature extractor
- ✓ Subjectivity analyzer/ subjectivity detector
- ✓ Sentiment classifier

As we know, for performing machine learning text classification data is the crucial part. For our research we collected our data from Ethiopian Ministry of Health, official Facebook page. We select this page as our main data source because it’s a governmental official page and hence the posts are correct and relevant. After collecting and preparing the data, we moved to the second component which is preprocessing. The preprocessor component contains four subcomponents: text cleaning, tokenization, stop word removal, and normalization of texts. In the feature extraction component, we are going to compare term frequency-inverse document frequency (Tf-idf) and bag of words (BOW) based on the data we collected. In the subjectivity analyzer component, we designed machine learning **ternary** classifier model that identify a given statement in to subjective statement, objective statement and non-related statement. The last component is sentiment classifier that classifies the given subjective statement in to multi-class. Each component of the system with their function are discussed in detail in this Section. The general architecture of multi-class subjectivity detection and sentiment analysis system is shown in Figure3.

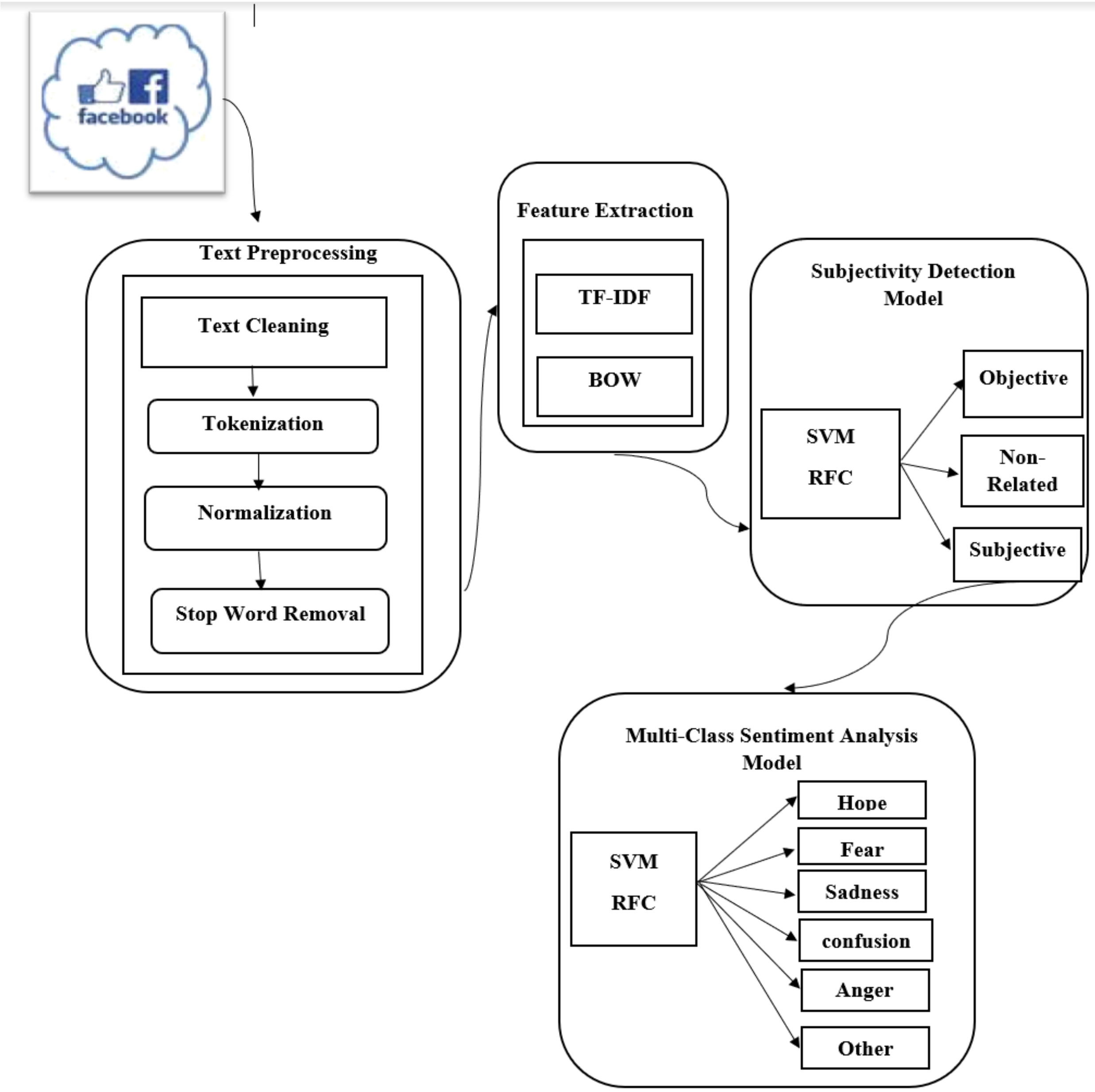


Figure 3 The proposed Multi-class Subjectivity Detection and Sentiment Analysis Model

3.3. Components of the proposed model

3.3.1. Data set preparation

As we said earlier, dataset preparation is crucial task for text-based machine classification. In our context, we are not able to find a prepared Amharic dataset for COVID-19 sentiments. As a result, we had to prepare our own dataset from the scratch. In our case, we selected social media (Facebook) for the data collection purpose because people with different sentiment write their opinion without any fear in this social media platform. As we are going to perform sentiment analysis, Facebook help us to collect those sentiments easily.

Facebook was introduced as a social networking platform in 2004. Since then, it has become a trending social networking service, allowing users to post their personal profiles, share videos, photos and other information. Others can view the profiles, videos, and images on the owner's friend list. As such, social media is the fastest and most preferred source of information on the Web linking users throughout the world. Therefore, people can now make an impact on one another effortlessly and conveniently. Due to the exceptional rise in the amount of information available, the demands for a computerized strategy that reacts to changes in sentiment and to increase tendencies is inevitable[61].

For this study, primary data sources have been taken from Ministry of Health, of Ethiopia official Facebook page because this page was legal under the Facebook company terms and conditions and also, we use additional sources for completeness. We collect about 4988 sentences from posts and comments to prepare our corpus. Face pager, an app designed for fetching publicly available data from YouTube, Twitter, Facebook, and other websites via APIs and webscraping, was used in our work to extract posts and comments from Facebook.

Annotation

Data labeling is an important part of data preprocessing for Machine Learning, particularly for supervised learning, in which both input and output data are labeled for classification to provide a learning basis for future data processing. Labeled datasets assist in training our Machine Learning models to identify and understand recurring patterns in the input fed into them in order to deliver accurate output. After being trained on annotated data, Machine Learning models can begin recognizing the same patterns in new unstructured data[62].

Based on this definition we labeled the data we extracted with Face pager. The manual annotation task for both subjectivity detection and sentiment analysis (to classify the statements in to subjective, objective and non-related class and for further classifying the subjective statement in to Fear, Hope, Confusion, Sadness, Anger and other class) is done by Jimma University psychology and Amharic language MSc Students based on the annotation guidelines we give to them. Finally, the data is saved in (.CSV) format to be read by python programming language for further processing task.

3.3.2. Text preprocessing

Pre-processing routines get the data ready for analysis. Before we begin training, the data must be pre-processed to remove defects. When data is collected as a result of an experiment, the next step is to model the data in order to extract useful information. Figure 4 explain preprocessing steps used in our study.

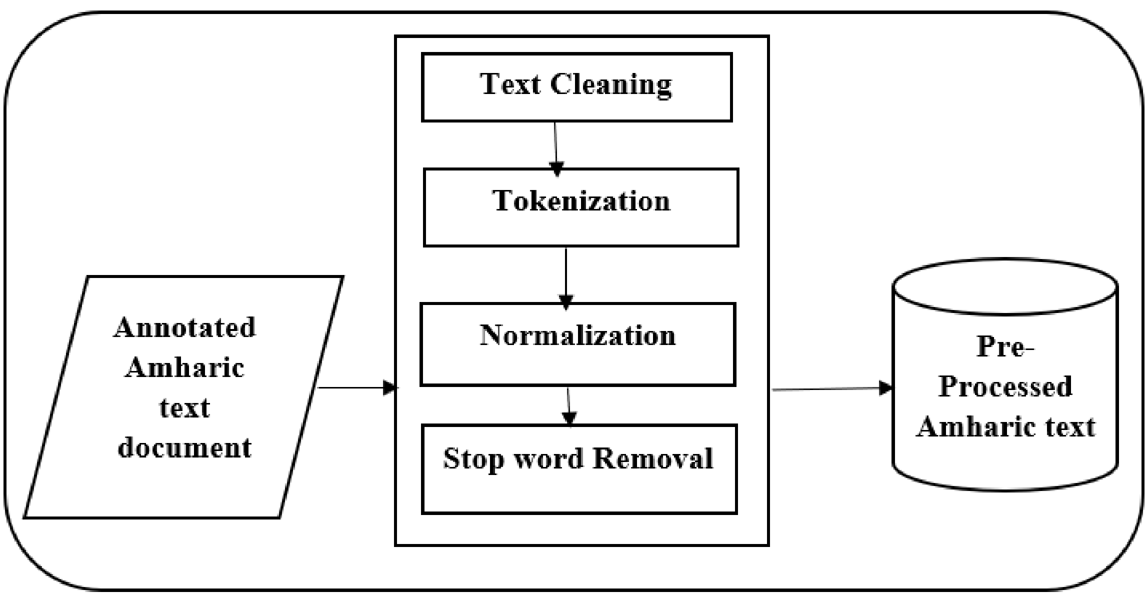


Figure 4 general data cleaning and preprocessing process

3.3.2.1. Text cleaning

The first important step in any machine learning or data analysis task is to clean the data. Cleaning is essential for model construction. Cleaning data is dependent on the type of data, and the task to be performed by the data. The ultimate goal of cleaning and preparing text data is to reduce the text to only the words required for our NLP goals. The following tasks are performed to clean the collected data from Facepager.

- ✓ When a data is exported from Facepager additional information about the post or comment is also added so we get rid of the columns such as: path, id, parent_id, level, object_id, object_type, object_key,query_status, query_type, query_time, created_time, like_count, comment_count, and error_message.
- ✓ Eliminating non-textual, non-Amharic contents and symbols like hashtags, URLs, links and others.
- ✓ Spell correction

3.3.2.2. Tokenization

Tokenization is the next step in pre-processing. Tokenization is the process of separating a stream of text into words, phrases, symbols, or other meaningful elements known as tokens. The goal of tokenization is to find the words in a sentence. The list of tokens is used as input for subsequent processing. The splitting criteria are mainly at the occurrence of a space or punctuation marks.

Historically, in Amharic writing system, words were used to be explicitly delimited with a non-whitespace character (:). However, in recent years, the delimitation used is white space and punctuation marks which include (፤ semicolon (ድርብ ሰረዝ) *deribsereze'* (፣ comma (ነጠላ ሰረዝ), netelaserez" (! exclamation mark (ቃል አጋኖ), and (? Question mark (ጥያቄ ምልክት) question mark at the end of sentences[63]. The Amharic punctuation marks and the space between words are used for the tokenization process.

After we used those punctuation marks for tokenization process, we need to eliminate them because these punctuation marks don't have any relevance in developing sentiment analysis model and they are unnecessary tokens. So, they have to be removed. We created Amharic punctuation collections such as (", ", '/', ';', '?', "), etc. to remove those unwanted punctuations from our data set. Then, after checking the collection, remove the punctuations from the text document. Because those punctuations are unnecessary information in a document, they may cause model performance and system memory usage to decrease.

Algorithm 3.1: Tokenization

```
Input: Line of string
Output: token list without punctuation
Load Amharic csv file from the directory
Loop: check punctuations from list
          read character sequence from a file
       if character is space, ፤, ፣, ፡, ።, tab,
            append token
    else
         token+character
while (end of file)
```

Algorithm 3.2: Remove Amharic punctuations

```
Input: Line of token
Output: token list without punctuation
Load Amharic csv file from the directory
Loop: check punctuations from list
          if Amharic punctuation equals character j:
      remove punctuation and store clean texts
else
         store existing token
     end if
end Loop:
```

3.3.2.3. Normalization

Normalization refers to the consistency of characters**,** after tokenization. The Amharic writing system has homophone characters which mean, characters with the same sound have different symbols for example; it is common that the character ሰ and ሠ are used interchangeably as ሰራ and ሠራ to mean "work". These different symbols must be considered as similar because they do not have an effect on meaning. Such type of inconsistency in writing words will be handled by replacing characters of the same sound by a common symbol. Thus, for example, if the character was one of ሐ ፣ ሀ (all of them with a similar sound h) then it was converted to ሀ. The normalizing process puts the Amharic text in a consistent form, thus converting all the various forms of a word to a common form.

Table 2 Example of Amharic normalized character

Same sound Character	Normalized to
ሀ, ሐ	ሀ
ሰ, ሠ	ሰ
ጸ, ፀ	ጸ
አ, ዐ, ዓ	አ

Algorithm 3.3: Normalization

Input: unnormalized texts
Output: normalized character sequence
Load the Amharic CSV file from the directory
do
    for each character in the corpus
      If the character is ሐ,ሀ or in any order then
          Changed to ሀ
    Else if it is ሠ or any order of it then
          Changed to ሰ
    Else if it is ጸ or any other order of it then
          Changed it to ፀ
     End if
  End for
while (end of file)

3.3.2.4. Stop word removal

The most common words in any natural language are known as stop words. These stop words may not add much value to the meaning of the document when analyzing text data and building NLP models or they have no bearing on how sentiments are classified. Stop words in English include words like "the", "a", "an", "so", and "what" etc which are not used to classify opinions or sentiments. Amharic, like other languages, has its own stop words. "ሆ", "ሁሉ", "ነው", "ነበር", and others like "እና", " ወይም", " ውስጥ", " ላይ" and so on. NLTK, or the Natural Language Toolkit, is a library for text preprocessing. NLTK has a list of stop words stored in 16 different languages.

Unfortunately stop word for Amharic language is not included in this library. So that a list of stop words should be identified and listed in order to remove them for Amharic language. In our context a list of stop words are identified and listed.

Algorithm 3.4: Stop word removal

```
Input: list of tokens from a list of all stop words
Output: list of tokens without stop word
do
  read token from list of all features
   if token in the stop-word list
       Remove a token from the document token list
   else
     append token
  end if
while (end of file)
```

3.3.3. Feature extraction

After the completion of the data collection and preprocessing task the second component of our architecture is feature extraction. Because most machine learning algorithms can't handle straight text, we'll use Bag of Words and TF-IDF to create a matrix of numerical values or a real-valued vector to represent our text. The sklearn.feature extraction module can be used to extract features from datasets that include formats such as text for use in a machine learning algorithm.

3.3.3.1. Bag of word model (BOW)

The bag-of-words model, or BOW, is a method of extracting features from text for use in modeling, such as machine learning algorithms. Like the other feature extraction models BOW model also used to convert text sentence in to numeric vectors. A bag-of-words is a text representation that describes the appearance of words in a document. It consists of two parts: a vocabulary of well-known words and a measure of the presence of well-known words. Because any information about the order or structure of words in the document is discarded, it is referred to as a "bag" of words. The model is only concerned with whether or not known words appear in the document, not.

In the creation of BOW model there are two operations we need to perform tokenization and vector creation. Tokenization is the process of breaking down each sentence into words or smaller parts known as tokens. Following the completion of tokenization, we will extract all of the unique words in the corpora. In this case, corpus represents the tokens obtained from all of the documents and used to create the bag of words. Then the next task is vector creation, here in our case we select countvectorizer for creation of the vectors. In this case, the vector size for a given document is equal to the number of unique words in the corpus. For each document, we will fill each entry of a vector with the corresponding word frequency in that document. As example let us look at these two sentences:

S1: ፈጣሪ ሆይ እባክህ ይቅር በለን ኢትዮጲያን ጠብቅልን

S2: ፈጣሪ ኢትዮጲያን ይጠብቅልን የታመሙትን ይማርልን

So to apply BOW model first we need to convert the sentences in to token which is dividing the sentences into words and generate a list with all unique words in alphabetical order, we will get the following output:

Unique terms: [“ፈጣሪ”, “ሆይ”, “እባክህ”, “ይቅር”, “በለን”, “ኢትዮጲያን”, “ጠብቅልን”, “ይጠብቅልን”, “የታመሙትን”, “ይማርልን”]

Our next step is Creating vectors for each sentence with the frequency of words. Simply we create the vector-matrix by assigning 0 for the absence of the words and, 1 for the presence of the word.

Table 3 Vector matrix for BOW

	ፈጣሪ	ሆይ	እባክህ	ይቅር	በለን	ኢትዮጲያን	ጠብቅልን	ይጠብቅልን	የታመሙትን	ይማርልን
S1	1	1	1	1	1	1	1	0	0	0
S2	1	0	0	0	0	1	0	1	1	1

Vector of S1: [1111111000]

Vector of S2: [1000010111]

Therefore, as we see here each document is represented as an array having a size same as the length of the total number of features. All the values of this array will be zero apart from one position and that position represents words address inside the feature vector. The final BOW representation is the sum of the word’s feature vector. Unfortunately, in BOW model we are retaining no information about the grammar of the sentences or the ordering of the words in the text. Furthermore, the vectors would contain many 0s, resulting in sparse matrix.

3.3.3.2. Term frequency inverse document frequency (TF-IDF)

The BOW method is the simplest and most effective model, but it has the disadvantage of treating all words equally. As a result, it cannot distinguish between common and uncommon words. As a result, TF-IDF arrives to solve the problem. The term frequency-inverse document frequency (TF-IDF) model considers the importance of a word based on how frequently it appears in a document and in a corpus. To get a general image of TF-IDF we need to understand the two terms meaning term frequency and inverse document frequency separately. TF represent the frequency of a given word in a document, for a given specific word it calculates the ratio of the number of times a word appears in a document to the total number of words in the document. The term frequency $TF_{t,d}$ of term t in document d is defined as the number of times that t occurs in d.

$$\text{TF (term)} = \frac{\text{Number of times term t appears in a document}}{\text{Total number of items in the document}}$$

IDF assesses the significance of a word in the corpus. It computes the frequency of occurrence of a specific word across all documents in the corpus. It is the logarithmic ratio of the total number of documents to the number of documents containing a specific word.

$$\text{IDF (term)} = \log \frac{\text{Total number of documents}}{\text{Number of documents with term in it}}$$

The above equation helps in the extraction of rare words because the value of the product is greatest when both terms are greatest.

To illustrate this concept let’s take these two sentences

S1: ፈጣሪ ሆይ እባክህ ይቅር በለን ኢትዮጲያን ጠብቅልን

S2: ፈጣሪ ኢትዮጲያን ይጠብቅልን የታመሙትን ይማርልን

The **Term Frequency** and **IDF** of the given terms is:

Table 4 Term Frequency and IDF of a given term

Term	S1	S2	TF(S1)	TF(S2)	IDF
ፈጣሪ	1	1	1/7	1/5	0
ሆይ	1	0	1/7	0	0.3
እባክህ	1	0	1/7	0	0.3
ይቅር	1	0	1/7	0	0.3
በለን	1	0	1/7	0	0.3
ኢትዮጲያን	1	1	1/7	1/5	0.3
ጠብቅልን	1	0	1/7	0	0.3
ይጠብቅልን	0	1	0	1/5	0.3
የታመሙትን	0	1	0	1/5	0.3
ይማርልን	0	1	0	1/5	0.3

After getting the value of TF and IDF the next task is to compute the TF-IDF of the terms by using

$(\text{TF-IDF}_{t,d}) = \text{TF}_t * \text{IDF}$

Table 5 TF-IDF of terms

Term	S1	S2	TF-IDF(S1)	TF-IDF(S2)
ፈጣሪ	1	1	0.000	0.000
ሆይ	1	0	0.042	0.000
እባክህ	1	0	0.042	0.000
ይቅር	1	0	0.042	0.000
በለን	1	0	0.042	0.000
ኢትዮጲያን	1	1	0.042	0.060
ጠብቅልን	1	0	0.042	0.000
ይጠብቅልን	0	1	0.000	0.060
የታመሙትን	0	1	0.000	0.060
ይማርልን	0	1	0.000	0.060

We now have our TF-IDF scores for our vocabulary. TF-IDF also yields higher values for less frequent words and is high when both IDF and TF values are high, indicating that the word is rare in all documents combined but common in a single document.

3.3.4. Proposed Subjectivity Detection model

Subjectivity analysis or subjectivity detection is one major sub task in sentence level sentiment analysis. Currently most of the researcher use online available data for sentiment analysis task. Data available on internet is most of the time heterogeneous which means it contains information which is not related to our domain (non-related statements), information which are related to our domain but they are factual information (Objective Statements) and also it contains people's opinion about our domain (Subjective statements). So, subjectivity analysis deals with the task of separating these three different statements and pass subjective statements to the next sentiment analysis phase. In our proposed model we used SVM and RFC machine learning algorithms for the classification of subjectivity analysis task. We compare and contrast these two models in the following section.

Support Vector Machine

SVM is a supervised machine learning algorithm that can be used to solve classification and regression problems. It attempts to find the best boundary (also known as a hyperplane) between different classes. In layman's terms, SVM performs complex data transformations based on the kernel function selected, and it aims to maximize the separation boundaries between our data points based on those transformations[44]. SVM does not support multiclass classification natively in its most basic form. It allows for binary classification and the separation of data points into two classes. However, in our case, multi-class sentiment analysis was proposed. The same principle is used for multiclass classification after breaking down the multiclassification problem into multiple binary classification problems.

Support vector machine can solve multiclass problems using one-against-one or one-against-rest methods[64]. In the one-against-one approach, instead of trying to distinguish one class from all the others, they seek to distinguish one class from another one. We need a hyperplane to separate every two classes in the one-against-one approach, ignoring the points of the third class. This means that in the current split, the separation takes only the points of the two classes into account. In the one-against-rest binary of problem, SVM is trained for each class in order to distinguish that class and the rest. A hyperplane is required in the one-against-rest approach to separate a class from all others at once. This means that the separation considers all points and divides them into two groups: one for the class points and one for all other points.

Support vector machine (SVM) has many classes library that is capable of performing binary classification and multi classification such as: SVC, NuSVC and linearSVC library. SVC and NuSVC are similar methods and use kernel parameters. Linear SVC, on the other hand, is another implementation of support vector classification for the case of a linear kernel, but it does not accept keyword kernel. In the multi-class classification process, SVC and NuSVC implement the “one-against one” approach. Linear SVC automatically uses the one-against-all strategy by default. In this work, we used SVC library.

Random Forest

Random Forest is a well-known machine learning algorithm from the supervised learning technique. It can be used in machine learning to solve classification and regression problems. It is based on the concept of ensemble learning, which is the process of combining multiple classifiers to solve a complex problem and improve the performance of the model. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset, as the name implies. Instead of relying on a single decision tree, the random forest takes the predictions from each tree and predicts the final output based on the majority vote of predictions.

A random forest is an ensemble of tree structured classifiers and every tree of the forest gives a unit vote, assigning each input to the most probable class label[65]. Random forest is a tree-based machine learning algorithm that makes decisions by combining the power of multiple decision trees. Random forest is a forest of trees. The main advantage of random forest is that it does not suffer from over fitting, even if more trees are appended to the forest[65]. Random forest classifier uses many of the parameters as Decision tree classifier.

3.3.5. Multi-class sentiment analysis model

Subjective statements which are required for the multi-class sentiment analysis task is found in the subjectivity analysis model using supervised machine learning algorithm. The subjective statements detected in the subjectivity analysis model serve as an input in the multi-class sentiment analysis model. By using this model, we are going to classify the subjective statements in to six classes named as Fear (ፍርሃት), Hope (ተስፋ), Confusion (ግራ መጋባት), Anger (ቁጣ), Sadness (ሀዘን), and other (ሌላ) classes.

As we describe in the subjectivity analysis model SVM and Random Forest algorithms are also used here as they perform well for multi-class classification hence classifying a given data in to three and more class is considered as multi-class.

3.4. Evaluation metrics

There are two independent models in our study: The subjectivity analysis model and the multi-class sentiment analysis model. So, we defined the following evaluation metrics to measure the effectiveness of those models. True Positive (TP) refers to the number of observations that are subjective statements and are predicted to be subjective statements. False Positive (FP) refers to the number of observations that are objective statements but are predicted subjective statements.
True Negative (TN) refers to the number of observations that are objective statements and are predicted to be that are objective statements. False-negative (FN) refers to the number of observations that are subjective statements but are predicted objective statements. The main target from here is to find the proper approach with the highest TP and TN lost FP and FN. Based on those four metrics we calculate precision, recall, and F1 score to select the best algorithm for Subjectivity analysis and sentiment analysis[19][66].

Precision: it measures among all of the Subjective statements identified by the model, how many of them are correct Subjective statements[66].

Mathematically, expressed as follows.

$$\textbf{Precision} = \frac{\textbf{TP}}{\boldsymbol{TP+TN}}$$

Where TP is true Positive, TN is True negative.

Recall: it measures among all known true Subjective statements, how many of them are identified by the subjectivity analysis model. It is the proportion of true positives against the whole true or correct data. It shows how well the model avoids false negatives[19]. Mathematically expressed as followed: -

$$\textbf{Recall} = \frac{\textbf{TP}}{\boldsymbol{TP+FP}}$$

Where TP is True positive, FP false positive.

F1-score: measures the combination of precision and recalls to evaluate the overall accuracy of the subjectivity analysis and sentiment analysis model. It is the weighted average precision and recall[66].

$$\textbf{F1-Score} = \frac{2*\text{precision}*\text{Recall}}{\text{precision} + \text{Recall}}$$

Chapter Four: Experiment and Discussion

4.1. Introduction

In this study, we made an attempt to construct multi-class subjectivity detection and sentiment analysis model. In this work we create two separate models which performs subjectivity analysis and then sentiment analysis. This chapter contain all dataset preparation, implementation procedure and experimental results. Then from the result obtained from the experiment, we evaluate the performance of the proposed model.

4.2. Data sets description

We have collected about 4988 data mainly from Ministry of Health, Ethiopia official Facebook page and from some Facebook social media source to train and test our proposed model. We used Ministry of Health, Ethiopia official Facebook page behind the reason of its accountability we get all the posts and comments related to COVID-19 which help us to train and test our model about subjective statements and objective statements. But as we know information in social media is not only focused on COVID-19 data there are a lot of other domains available therefore we need to collect other data which are not related to our domain to train and test our model in a such case that's why we need some Facebook social media sources to collect data about other domain. After the we finish the data collection using Face pager, it was annotated by psychology and linguistic student since, we are going to create supervised model. We have created two datasets one for subjectivity analysis and the other one for multi-class sentiment analysis.

The total number of data and the distribution of the data for the classes for subjectivity detection datasets shown in Table 6 and Figure 5 and for sentiment analysis in Table 7 and Figure 6.

For Subjectivity analysis:

Table 6 Total data for subjectivity analysis

Class	Total number of data
Subjective	2998
Objective	1000
Non related	990

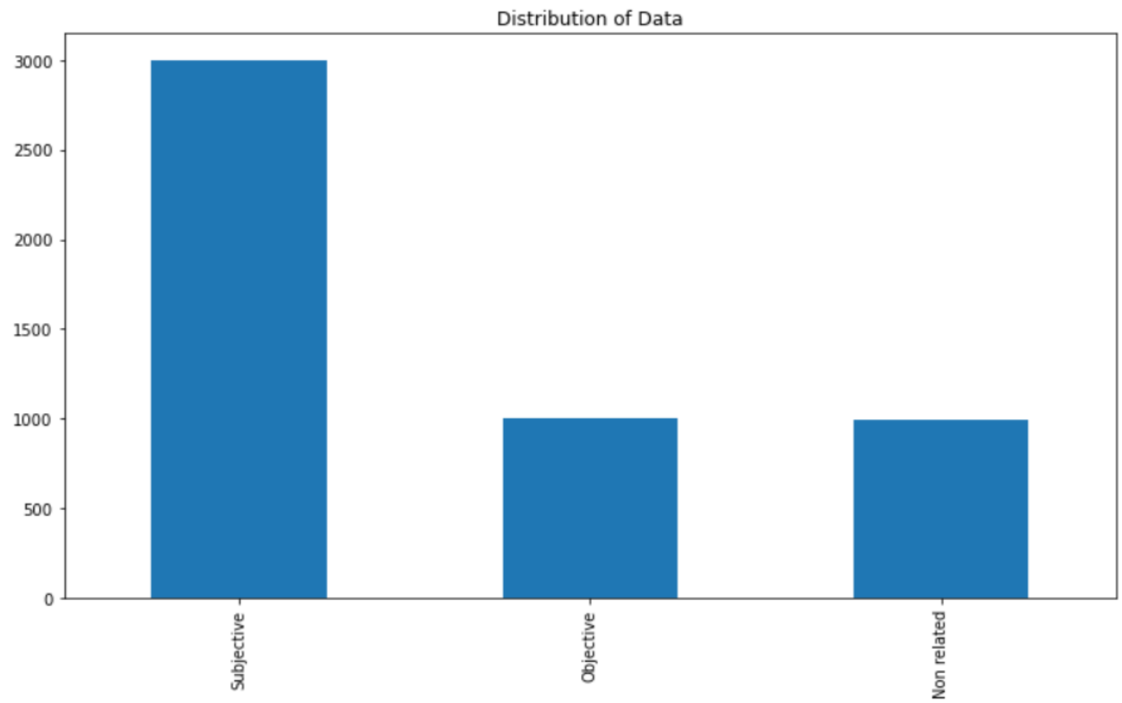


Figure 5 Distribution of data for subjectivity analysis

For sentiment analysis:

Table 7 Total data for multi-class sentiment analysis

Class	Total number of data
Hope	946
Fear	850
Anger	246
Confusion	338
Sadness	292
Other	326

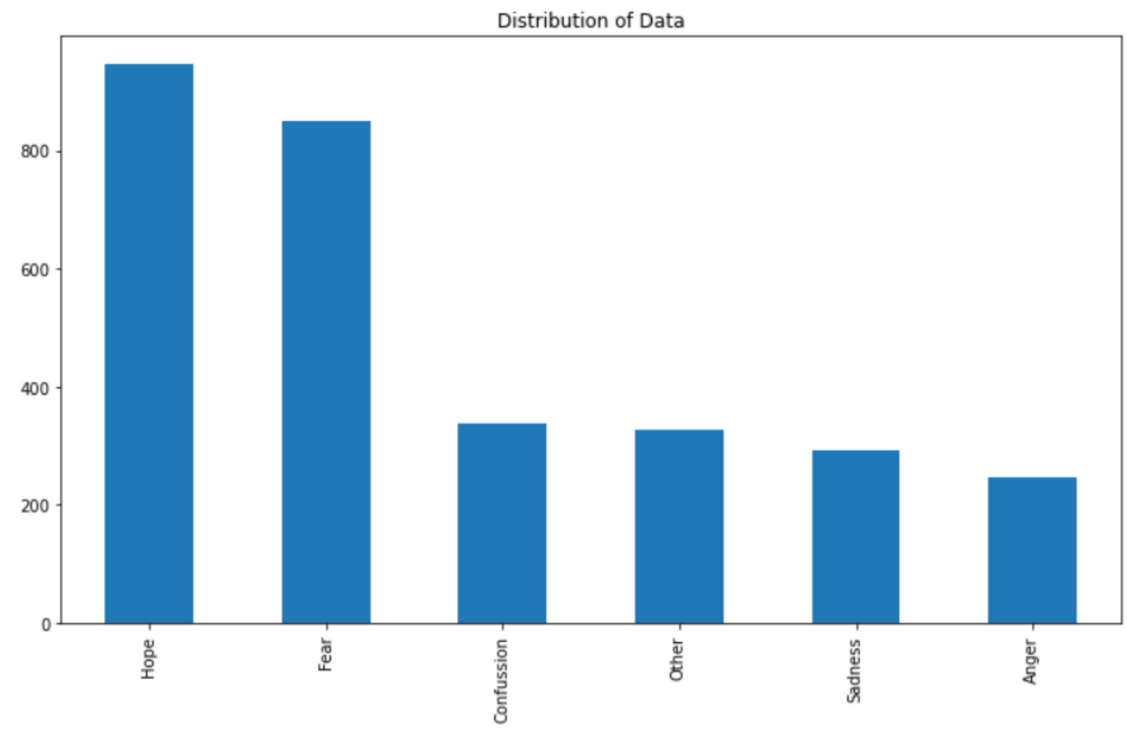


Figure 6 Distribution of data for multi-class sentiment analysis

4.3. Developmental tool for this study

In order to achieve the objective of multi class subjectivity detection and sentiment analysis we used python programming language (3.7 version). We used this programming language starting from preprocessing to evaluate our model including the building of our model. Python is a dynamically semantic, interpreted, object-oriented high-level programming language. Its high-level data structures, combined with dynamic typing and dynamic binding, make it ideal for Rapid Application Development as well as a scripting or glue language for connecting existing components.

In this study, Anaconda Navigator 3.7.0 is used, which is a desktop graphical user interface (GUI) included in the Anaconda® distribution that allows us to manage conda packages, environments, and channels without using command-line commands. It includes different applications like JupyterLab, Jupyter Notebook, Spyder, PyCharm, VSCode, Glueviz, Orange 3 App, RStudio, and others. And also, we use Jupyter Notbook version 3.7.4 application to create live code, visualization, and narrative text.

Table 8 describe list of python libraries used for our work

Table 8 List of libraries used

Python libraries used	Description
Pandas	is an open-source library that allows to you perform data analysis in Python. Used to read, write and manipulate data frames
Numpy	NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.
Nltk	Nltk is a platform used for building Python programs that work with human language data.
String	Byte of array representing Unicode characters in python.
Matplotib	Matplotlib is a cross-platform, data visualization and graphical plotting library for Python.
Sklearn	The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

4.4. Experimental process

After we prepare the environment for the experiment, we loaded the annotated dataset to create and evaluate our model. As its described in the python library part, we use pandas to read the dataset. We prepare the dataset in a CSV file format and so, read_csv function is used to load our CSV file.

```
subjectivitydata = pd.read_csv(r'C:/Users/USER/Desktop/COVIDdataset/Subjectivity.csv')
print (subjectivitydata)
dataset = subjectivitydata.values.tolist()
```

Figure 7 Implementation of loading dataset

We conduct three experiments in this study, Scenario1: Feature extraction, Scenario 2: Subjectivity detection, Scenario 3: Multi-class sentiment analysis.

4.4.1. Scenario 1: Feature extraction

For textual dataset to be understood by the machine feature extraction is needed. Which is a technique used in the transformation of list of word in our dataset into a feature vector which is understandable by machine learning model.

Due to the reason that we are not able to use the prepared data set as it is to feed in to the model feature extraction is required. For our study, Bag of words and term frequency inverse document frequency are used for extraction feature. we select this two feature extraction techniques due to the reason to compare and recommend the best one which make high performance in related to the model. The Python sklearn.feature_extraction module can be used in our study to extract features in a format supported by machine learning algorithms from datasets consisting of formats such as text. Except the vectorizer we implemented, similar dataset, parameters and value are used for both bag of word and term frequency inverse document frequency feature extraction techniques.

The bag-of-words model, abbreviated as BOW, is a technique for extracting features from text for use in modeling, such as machine learning algorithms. A bag-of-words is a text representation that describes how words appear in a document. It consists of two parts: a vocabulary of well-known words and a measure of the presence of well-known words. To convert a collection of text documents to a matrix of token counts in our bag of words, we used Countvectorizer with max_features as a parameter. The preprocessed data is going to be split in to training and test sets by passing a training_data['Message'], training_data['Class'], test_size=0.1, and random_state=100 as a parameter to train test split function. By now our data is splits in to 90:10 in which 90 percent of the data for training and the rest for test purpose.

```
sentences = training_data['Message'].values
lable = training_data['Class'].values
sentences_train, sentences_test, y_train, y_test = train_test_split(training_data['Message'],training_data['Class'],test_size=0.1
                                                    random_state=100)
BOW_vectorizer = CountVectorizer(max_features=500)
BOW_vectorizer.fit(sentences_train)
X_train = BOW_vectorizer.transform(sentences_train)
X_test  = BOW_vectorizer.transform(sentences_test)
print("Train data size", X_train.shape)
print("Test data size", X_test.shape)
```

Figure 8 Feature extraction using countvectorizer

The term frequency-inverse document frequency (TF-IDF) model considers the importance of a word based on how frequently it appears in a document and a corpus. TF represent the frequency of a given word in a document, for a given specific word it calculates the ratio of the number of times a word appears in a document to the total number of words in the document. In our TF-IDF model, we used TfidfVectorizer with max_features as a parameter to convert a collection of raw documents to a matrix of TF-IDF features. The preprocessed data is going to be split in to training and test sets by passing a training_data['Message'], training_data['Class'], test_size=0.1, and random_state=100 as a parameter. By now our data is splits in to 90:10 in which 90 percent of the data for training and the rest for test purpose.

```
Train_X, Test_X, Train_Y, Test_Y = model_selection.train_test_split(training_data['Message'],training_data['Class'],test_size=0.1
                                                                    random_state=100)
print("Train data size",Train_X.shape)
print()
print("Test data size",Test_X.shape)
print()
# encode non asci character
Encoder = LabelEncoder()
Train_Y = Encoder.fit_transform(Train_Y)
Test_Y = Encoder.fit_transform(Test_Y)
# feature extruction
Tfidf_vect = TfidfVectorizer(max_features=500)
Tfidf_vect.fit(training_data['Message'])
Train_X_Tfidf = Tfidf_vect.transform(Train_X)
Test_X_Tfidf = Tfidf_vect.transform(Test_X)
```

Figure 9 TF-IDF with Tfidf vectorizer

4.4.2. Scenario 2: Subjectivity Detection Experiment Result

For any kind of sentiment analysis task identification of subjective sentences from objective ones is a crucial task. In our research work, it’s a major goal to identify these two separate sentences. Especially when it comes in to sentence level sentiment analysis performing, subjectivity detection is a crucial task. To proceed to sentiment analysis, we required subjective statements as input data which shows peoples feeling, emotion, attitude, and the like about some issue. But information available in the world is not only limited to peoples felling rather they also contain objective or factual statements about some issue. Subjectivity detection is required to filter the required type of statement for further analysis.

For the reason that we use labeled data to build and evaluate the model supervise machine learning algorithms are selected. In order to examine the applicability of machine learning algorithm to classify the subjectivity detection model, SVM and Random Forest algorithms are compared with the same dataset and feature categories.

We used 4988 Amharic datasets for our mode as input for all independent models in the training and testing phase. The total dataset was divided into two parts for testing and training the model the main goal of dividing the dataset into a validation set is to avoid overfitting, which occurs when a model becomes extremely good at classifying samples in the training set but is unable to generalize and make accurate classifications on data it has never seen before. We use 90:10 split to classify the dataset in to training set and test set.

The whole dataset was annotated and labeled into three categories, Subjective (peoples feeling about COVID-19), objective (facts related to COVID-19), and non-related (any kind of data which is not related to COVID-19).

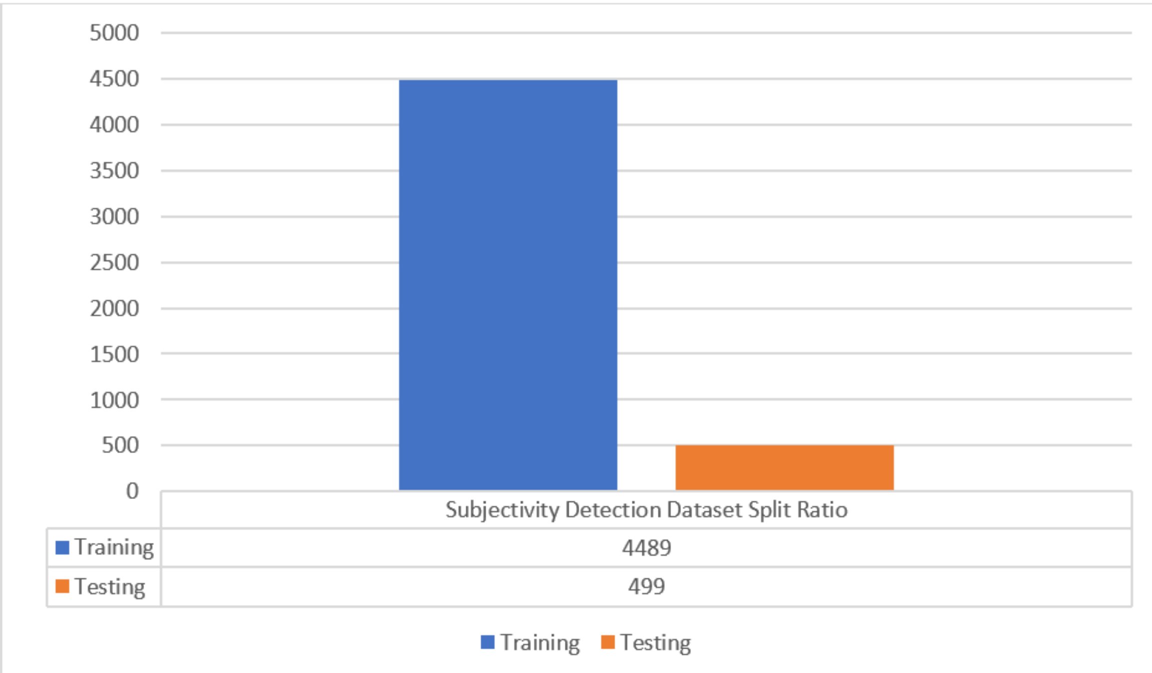


Figure 10 Subjectivity detection Dataset split ratio

4.4.2.1. Experimental result using SVM with BOW

In this work, we designed SVM model with BOW feature extraction model for subjectivity detection with the following libraries and parameters: We chose the SVC library because it uses the one-versus-one approach for multi-class classification. In total, n classes * (n classes-1) / 2 classifiers are built, with each classifier training data from two classes. We used Linear kernel since is the most fundamental type of kernel. When there are numerous features, it proves to be the best function. The linear kernel is commonly used for text classification problems because most of these problems can be linearly separated.

Precision, recall, and F-score are the three-measurement metrics that we use to assess the model's accuracy. We have got 88%, 80.3%, and 83.3% respectively and the total average accuracy from the SVM classification Model is 87.37%. The detailed experiment result from our SVM with BOW subjectivity detection model is discussed in Table 9.

Table 9 SVM with BOW evaluation result

```
              precision    recall  f1-score   support

 Non related       0.88      0.70      0.78       106
   Objective       0.89      0.74      0.81        78
  Subjective       0.87      0.97      0.91       315

    accuracy                           0.87       499
   macro avg       0.88      0.80      0.83       499
weighted avg       0.87      0.87      0.87       499

SVM Accuracy Score:  87.374749498998
```

Confusion matrix for SVM with BOW

A confusion matrix is a technique for summarizing the performance of a classification algorithm.

Table 10 Confusion matrix for SVM with BOW

Confusion matrix for SVM with BOW		
Non related	Objective	Subjective
74	0	32
6	**58**	14
4	7	**304**

The results of the Support vector machine showed that from the 449 test data out of 106 data 74 were correctly classified as *Non related*, 0, and 32 were incorrectly classified as Objective and Subjective respectively; out of the 78 data 58 were correctly classified as Objective 6 and 14 incorrectly classified as Non related and Subjective respectively and out of 315 data 304 were correctly classified as Subjective, 4 and 7 incorrectly classified as Non related and Objective.

4.4.2.2. Experimental result using Random Forest with BOW

The second subjectivity detection experiment was implemented using random forest model with BOW feature extractor. We used *Random Forest Classifier* library from scikit-learn python library.

Precision, recall, and F-score are the three-measurement metrics that we use to assess the model's accuracy. We have got 91%, 85%, and 87.6% respectively and the total average accuracy from the Random Forest classification Model is 90.7%. The detailed experiment result from our Random Forest with BOW subjectivity detection model is discussed in Table 11.

Table 11 Random Forest with BOW evaluation result

```
              precision    recall  f1-score   support

 Non related       0.90      0.78      0.84       106
   Objective       0.91      0.77      0.83        78
  Subjective       0.91      0.98      0.95       315

    accuracy                           0.91       499
   macro avg       0.91      0.85      0.87       499
weighted avg       0.91      0.91      0.90       499

RFC Accuracy Score:  90.78156312625251
```

Confusion matrix for Random Forest with BOW

Table 12 Confusion matrix for Random Forest with BOW

Confusion matrix for RFC with BOW		
Non related	Objective	Subjective
83	3	20
7	**60**	11
2	3	**310**

The results of the random forest classifier showed that from the 449 test data out of 106 data 83 were correctly classified as *Non related*, 3, and 20 were incorrectly classified as Objective and Subjective respectively; out of the 7 data 60 were correctly classified as Objective 7 and 11 incorrectly classified as Non related and Subjective respectively and out of 315 data 310 were correctly classified as Subjective, 2 and 3 incorrectly classified as Non related and Objective.

4.4.2.3. Experimental result using SVM with TF-IDF

The third experiment focused on the implementation of SVM with TF-IDF feature extraction model. In this experiment, we use the same parameters and dataset split ratio so that the comparison become truthful with BOW model.

We have got 92.6%, 91.3%, and 92% for Precision, recall, and F-score respectively and the total average accuracy from the SVM classification Model is 93.78%. The detailed experiment result from our SVM with TF-IDF subjectivity detection model is discussed in Table 13.

Table 13 SVM with TF-IDF evaluation result

	precision	recall	f1-score	support
0	0.96	0.84	0.89	106
1	0.90	0.90	0.90	78
2	0.94	0.97	0.95	315
accuracy			0.93	499
macro avg	0.93	0.90	0.92	499
weighted avg	0.93	0.93	0.93	499

SVM Accuracy Score: 93.3867735470942

Confusion matrix for SVM with TF-IDF

Table 14 Confusion matrix for SVM with TF-IDF

Confusion matrix for SVM with TF-IDF		
Non related	Objective	Subjective
93	0	13
4	**67**	7
2	7	**306**

The results of the Support vector machine showed that from the 449 test data out of 106 data 93 were correctly classified as *Non related*, 0, and 13 were incorrectly classified as Objective and

Subjective respectively; out of the 78 data 67 were correctly classified as Objective 4 and 7 incorrectly classified as Non related and Subjective respectively and out of 315 data 306 were correctly classified as Subjective, 2 and 7 incorrectly classified as Non related and Objective.

4.4.2.4. Experimental result using Random Forest with TF-IDF

The fourth experiment deals with building of random forest model with TF-IDF model We used the same *Random Forest Classifier* library from scikit-learn as we used in BOW model.

We have got 94%, 87%, and 90.3% for Precision, recall, and F-score respectively and the total average accuracy from the Random Forest classification Model is 92.5%. The detailed experiment result from our Random Forest with TF-IDF subjectivity detection model is discussed in Table 15.

Table 15 Random Forest with TF-IDF evaluation result

```
              precision    recall  f1-score   support

           0       0.94      0.80      0.87       106
           1       0.97      0.82      0.89        78
           2       0.91      0.99      0.95       315

    accuracy                           0.93       499
   macro avg       0.94      0.87      0.90       499
weighted avg       0.93      0.93      0.92       499

RFC Accuracy Score:  92.58517034068137
```

Confusion matrix for Random Forest with TF-IDF

Table 16 Confusion matrix for Random Forest with TF-IDF

Confusion matrix for RFC with TF-IDF		
Non related	Objective	Subjective
85	0	21
5	**64**	9
0	2	**313**

The results of the RFC showed that from the 449 test data out of 106 data 85 were correctly classified as *Non related*, 0, and 21 were incorrectly classified as Objective and Subjective respectively; out of the 78 data 64 were correctly classified as Objective 5 and 9 incorrectly classified as Non related and Subjective respectively and out of 315 data 313 were correctly classified as Subjective, 0 and 2 incorrectly classified as Non related and Objective.

4.4.3. Scenario 3: Multi-class Sentiment Analysis Experiment Result

As we explained earlier, in the case of COVID-19 pandemic, simple classification of people's sentiment as binary class such as positive or negative is not enough rather, we need to analyze emotion of people about the pandemic in a Multi class approach. After Subjectivity detection, the output (subjective statement) from subjectivity analysis model used as input for the multi class sentiment analysis model to classify the sentiments as (Hope, Fear, Confusion, sadness, anger and other).

We used 2998 subjective data to build and evaluate our multi-class sentiment analysis model. We label the data based on the above class. Then the entire data was going to be split in to training and testing data for this purpose we use 90:10 split to classify the dataset in to training set and test set.

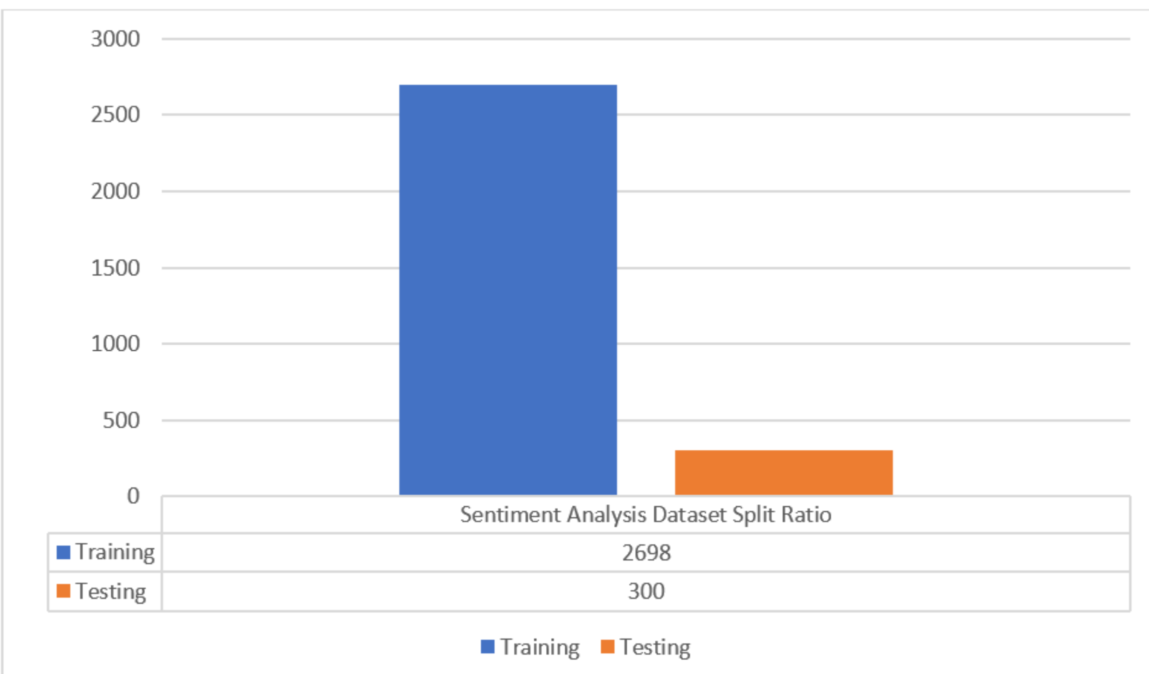


Figure 11 Multi-class Sentiment analysis dataset split ratio

As we seen in the above subjectivity detection model BOW model does not work well when we compare it with TF-IDF model so for our multi-class sentiment analysis model we used TF-IDF feature extraction model.

4.4.3.1. Experimental result using SVM with TF-IDF

This experiment was focused on the implementation of support vector machine with TF-IDF feature extraction model for multi class sentiment analysis. We used SVC library and linear kernel due to the reason we explained earlier in subjective detection model.

Precision, recall, and F-score are the three-measurement metrics that we use to assess the model's accuracy. We have got 93.8%, 87.8%, and 90.5% respectively and the total average accuracy from the SVM classification Model is 90.6%. The detailed experiment result from our SVM with TF-IDF multi-class sentiment analysis model is discussed in Table 17.

Table 17 SVM with TF-IDF evaluation result

```
              precision    recall  f1-score   support

           0       1.00      0.82      0.90        33
           1       0.92      0.80      0.86        30
           2       0.86      0.91      0.88        80
           3       0.89      0.97      0.93       106
           4       1.00      0.92      0.96        25
           5       0.96      0.85      0.90        26

    accuracy                           0.91       300
   macro avg       0.94      0.88      0.90       300
weighted avg       0.91      0.91      0.91       300

SVM Accuracy Score:  90.66666666666666
```

Confusion matrix for SVM with TF-IDF

Table 18 Confusion matrix for SVM with TF-IDF

Confusion matrix for SVM with TF-IDF					
Anger	Confusion	Fear	Hope	Other	Sadness
27	2	4	0	0	0
0	**24**	1	5	0	0
0	0	**73**	7	0	0

0	0	3	**103**	0	0
0	0	0	1	**23**	1
0	0	4	0	0	**22**

The results of the Support vector machine showed that from the 300 test data out of 33 data 27 were correctly classified as *Anger*, 2, 4, 0, 0, and 0 were incorrectly classified as Confusion, Fear, Hope, Other, and Sadness respectively, out of the 30 data 24 were correctly classified as *Confusion*, 0, 1, 5, 0, and 0 incorrectly classified as Anger, Fear, Hope, Other, and Sadness respectively; out of the 80 data 73 were correctly classified as *Fear*, 0, 0, 7, 0, and 0 incorrectly classified as Anger, Confusion, Hope, Other, and Sadness respectively; out of the 106 data 103 were correctly classified as *Hope*, 0, 0, 3, 0, and 0 incorrectly classified as Anger, Confusion, Fear, Other, and Sadness respectively; out of the 25 data 23 were correctly classified as *Other*, 0, 0, 0, 1, and 1 incorrectly classified as Anger, Confusion, Fear, Hope, and Sadness respectively; out of the 26 data 22 were correctly classified as *Sadness*, 0, 0, 4, 0, and 0 incorrectly classified as Anger, Confusion, Hope, and Other, respectively.

4.4.3.2. Experimental result using Random Forest with TF-IDF

This test deals with building of random forest model with TF-IDF feature extractor. We used the *Random Forest Classifier* library from scikit-learn.

We have got 96.3%, 94%, and 95.1% for Precision, recall, and F-score respectively and the total average accuracy from the Random Forest classification Model is 94.3%. The detailed experiment result from our RFC with TF-IDF Multi-class sentiment analysis model is discussed in Table 19.

Table 19 Random Forest with TF-IDF evaluation result

```
              precision    recall  f1-score   support

           0       1.00      0.88      0.94        33
           1       0.93      0.90      0.92        30
           2       0.93      0.93      0.93        80
           3       0.92      0.97      0.94       106
           4       1.00      1.00      1.00        25
           5       1.00      0.96      0.98        26

    accuracy                           0.94       300
   macro avg       0.96      0.94      0.95       300
weighted avg       0.94      0.94      0.94       300

RFC Accuracy Score:  94.33333333333334
```

Confusion matrix for Random Forest with TF-IDF

Table 20 Confusion matrix for Random Forest with TF-IDF

Confusion matrix for RFC with TF-IDF					
Anger	Confusion	Fear	Hope	Other	Sadness
29	2	2	0	0	0
0	**27**	0	3	0	0
0	0	**74**	6	0	0
0	0	3	**103**	0	0
0	0	0	0	**25**	0
0	0	1	0	0	**25**

The results of the RFC showed that from the 300 test data out of 33 data 29 were correctly classified as *Anger*, 2, 2, 0, 0, and 0 were incorrectly classified as Confusion, Fear, Hope, Other, and Sadness respectively; out of the 30 data 27 were correctly classified as *Confusion*, 0, 0, 3, 0, and 0 incorrectly classified as Anger, Fear, Hope, Other, and Sadness respectively; out of the 80 data 74 were correctly classified as *Fear*, 0, 0, 6, 0, and 0 incorrectly classified as Anger, Confusion, Hope, Other, and Sadness respectively; out of the 106 data 103 were correctly classified as *Hope*, 0, 0, 3, 0, and 0 incorrectly classified as Anger, Confusion, Fear, Other, and Sadness respectively; out of the 25 data 25 were correctly classified as *Other*, 0, 0, 0, 0, and 0 incorrectly classified as Anger, Confusion, Fear, Hope, and Sadness respectively; out of the 26 data 25 were correctly classified as *Sadness*, 0, 0, 1, 0, and 0 incorrectly classified as Anger, Confusion, Hope, and Other, respectively.

4.5. Discussion and Summary of the Result

Developing Multi-class Amharic subjectivity detection and sentiment analysis model for COVID-19 was our goal when we start this study. We perform two separate tasks in our study the first one was subjectivity detection which classify a given sentence in to Subjective or Objective and Multi-class sentiment analysis which classify a given subjective statement in to Fear, Hope, Sadness, Anger, Confusion and Other.

We have taken experiments on Support Vector Machine and Random Forest machine learning classification algorithm with TF-IDF and BOW feature extraction techniques by preparing our own dataset from Facebook. Experimental results from proposed models in this study are evaluated using a variety of performance evaluation metrics. Precision, recall, f1-score, and accuracy are all metrics used to evaluate and compare the performance of proposed models to determine which Machine learning algorithm is best for Subjectivity Detection and Multi-class Sentiment Analysis.

The effectiveness of each model with feature extraction techniques was measured by using the above listed evaluation metrics. For Subjectivity Detection SVM with BOW achieved 87% accuracy while SVM with TF-IDF achieved 93.3%. On other hand, RFC with BOW achieved 90.7% accuracy while RFC with TF-IDF achieved 92.5% accuracy. Below Table 21 and Figure 12 summarizes our model performance result

Table 21 Experimental Result Summary for Subjectivity detection

	SVM with BOW	**RFC with BOW**	**SVM with TF-IDF**	**RFC with TF-IDF**
Precision	88%	91%	92.6%	94%
Recall	80.3%	85%	91.3%	87%
F-score	83.3%	87.6%	92%	90.3%
Accuracy	**87.37%**	**90.7%**	**93.78%**	**92.5%**

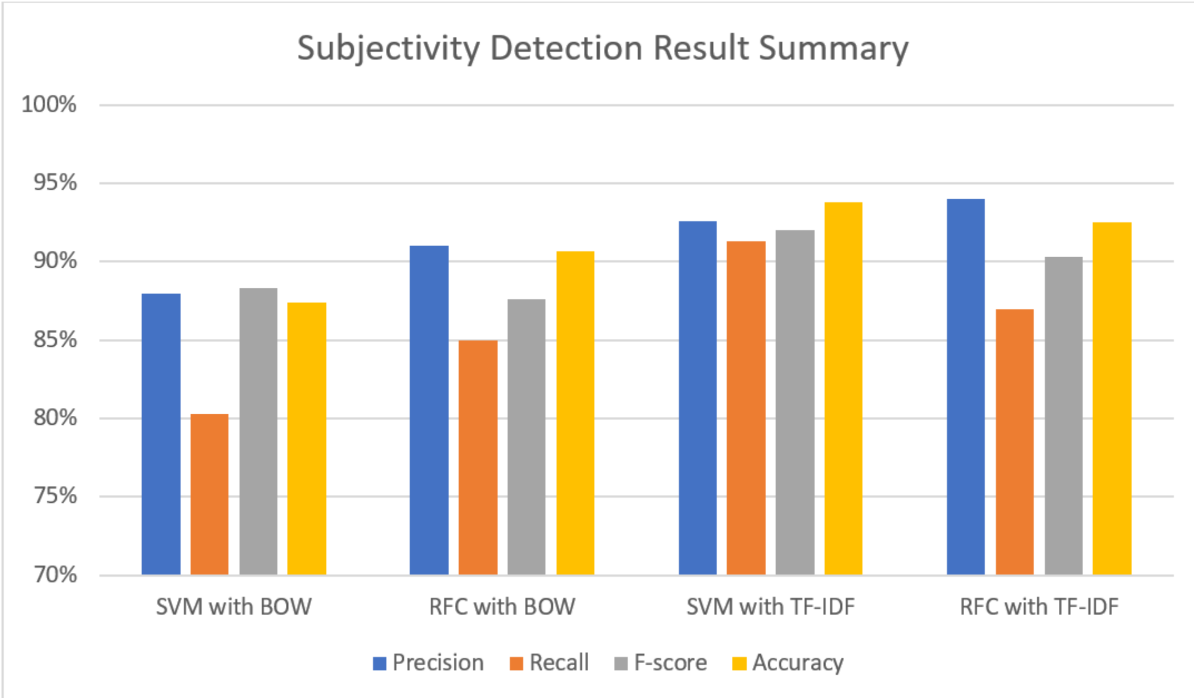


Figure 12 Summary of experimental result for Subjectivity Detection

For Multi-class Sentiment analysis SVM with TF-IDF achieved 90.6% and RFC with TF-IDF achieved 94.3% accuracy. Below Table 22 and Figure 13 summarizes our model performance result.

Table 22 Experimental result summary for Multi-class Sentiment analysis

	SVM with TF-IDF	**RFC with TF-IDF**
Precision	93.8%	96.3%
Recall	87.8%	94%
F-score	90.5%	95.1%
Accuracy	**90.6%**	**94.3%**

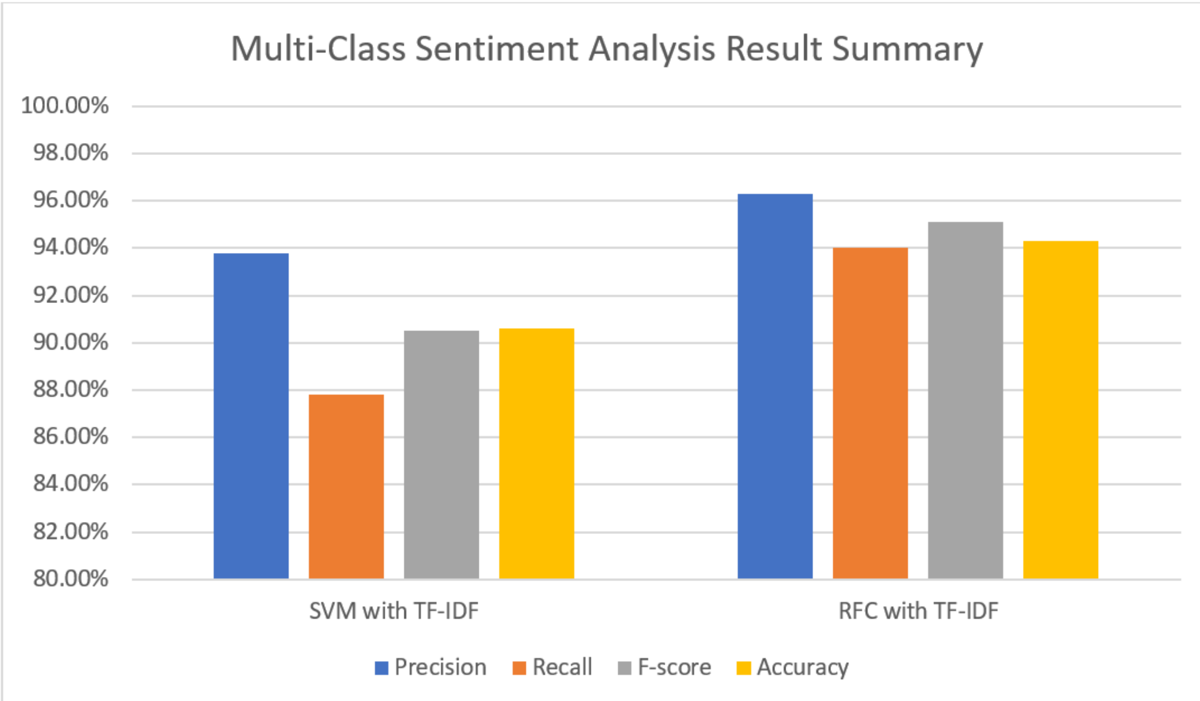


Figure 13 Summary of experimental result for Multi-class Sentiment Analysis

After these extensive experiments made for our proposed Multi-Class Amharic Subjectivity Detection and Sentiment Analysis, we observed that the SVM model with TF-IDF feature extraction technique performed better than the rest for Subjectivity Detection. For sentiment classification RFC model with TF-IDF feature extraction technique performed better than the rest. Table 23 shows comparative analysis with other papers done in Sentiment Analysis.

Table 23 Comparative analysis our work with others paper

Paper Title	**Methods**	**Class**	**Dataset**	**Accuracy**
Sentence level opinion mining for amharic language[14]	Lexicon and Dictionary	3(positive, negative and neutral)	Facebook (800 corpus size)	87% for Pos 94% for Neg 90% for Neu
A machine learning approach to multi-scale sentiment analysis of amharic online posts[15]	Naïve Bayes	5(Very Positive, Very Negative, Neutral, Positive, Negative)	Facebook, Twitter, DireTube (608 corpus size)	43.6% for Unigram 44.3 for Bigram 39.5% for Hybrid

Multi-class sentiment Analysis from Afaan Oromo text based on supervised machine learning approaches[58]	SVM RFC	5(Very Positive, Very Negative, Neutral, Positive, Negative)	OBN Tweeter (1810 corpus size)	90% for SVM 89% for RFC
Sentiment analysis for classifying Amharic opinionated text[59]	NB MNB SVM	3(Positive, Negative and Neutral)	SERTA (576 corpus)	77.6% for NB 74.7% for MNB 78.8% for SVM
Our paper	SVM RFC	**SD** 3(Subjective, Objective and non-related) **SA**6(Fear, Anger, Sadness, Hope, confusion and other)	From Ministry of Health, Ethiopia official Facebook page	**SD** 87.37% for SVM BOW 93.78% for SVM TF-IDF 90.7% for RFC BOW 92.5% for RFC TF_IDF **SA** 90.6% for SVM TF-IDF 94.3% for RFC TF_IDF

As Table 23 presented that our study shows better performance than the other study made in the area of sentiment analysis. Our current work improved by the number of datasets for the classification purpose which help the improvement of the accuracy. Additionally, we performed subjectivity detection in our work which is completely ignored by other researches.

Distinguishing between facts and opinions is possibly one of the most important sentiment analysis subtasks, as neutral statements can very negatively affect the information combination process that enables a polarity classifier to mine and categorize opinions into classes. Most of the time researchers used the total equal number of positive and negative words appear in the sentence to classify as neutral unfortunately those neutral statements may contain fact or some sentiment. This may cause biased in the sentiment classification task and decrease the accuracy of the models. Due to this we applied subjectivity analysis to minimize the biased and to only depend on subjective statement.

Chapter Five: Conclusions and Future work

5.1. Conclusion

In this paper, we have proposed multi-class subjectivity detection and sentiment analysis using machine learning approach: a case study on Amharic social media posts on covid-19. Specifically, we have used the TF-IDF feature extraction technique with a Support Vector Machine and Random Forest models and a proper text preprocessing technique. We have prepared our own dataset from the Ministry of Health, Ethiopia's official Facebook page by using a face pager. We have used a total of 4988 annotated sentences for training and testing purposes. We developed a Subjectivity detection model which classifies a given statement into Objective, Subjective, Non-related, and Multi-Class Sentiment classification based on pre-defined classes such as Hope, Fear, Anger, Confusion, Sadness, and others. The subjectivity detection model uses TF-IDF as a feature extraction technique and SVM to determine whether the input sentence is subjective, objective, or not related. This subjectivity detection model produces subjective statements as an output, and the subjective statements that are identified are used as input for the sentiment analysis model.
In this study, we reviewed the selective state-of-the-art machine learning approaches and feature extraction techniques for sentiment analysis and implement them for selecting the best performing approach and techniques. TF-IDF and BOW feature extraction techniques with Support vector machine and random forest approaches were implemented and evaluated in this study.

The proposed model's classification result is 87.37%, 93.78%, 90.7%, 92.5% for SVM with BOW, SVM with TF-IDF, RFC with BOW, and RFC with TF-IDF respectively. The result shows that SVM with TF-IDF model has the best Subjectivity detection accuracy.

After Subjectivity detection, we use the machine learning models with feature extraction techniques for Amharic multi-class sentiment analysis. Here we have six classes i.e., Hope, Sadness, Anger, Fear, Confusion, and Others. In the simulation Result, we conduct the comparison with different Machine Learning approaches (SVM and RFC). We implemented and evaluated those models, and the results show that RFC outperforms the other model in terms of accuracy. The proposed models can classify Sentiments in the Amharic testing data set 90.6% and 94.3% for SVM with TF-IDF and RFC with TF-IDF respectively. The result shows that RFC with TF-IDF model has the best multi-class sentiment analysis accuracy.

Finally, the accuracy predicted result is 93.78% for the proposed subjectivity detection model and 94.3% for multi-class sentiment analysis model.

5.2. Contribution of the work

Using a machine learning approach, we proposed a multi-class Amharic subjectivity detection and sentiment analysis for COVID-19. The contributions of this study are presented as follows: -

- ✓ We collect and annotate Amharic Text documents from the Ministry of Health, Ethiopia's official Facebook page related to COVID-19.
- ✓ We developed a subjectivity detection model which detects subjective statements for further sentiment classification.
- ✓ We developed a multi-class sentiment analysis model which classifies a given sentence into six sentiments.
- ✓ We have done different Experimental results on selective machine learning algorithms and selected one approach which has the best performance based on our problem domain.
- ✓ The previous researchers only classify the sentiment into positive and negative but in this work, we consider different emotions and we classify those sentiments into Hope, Fear, Anger, Confusion, Sadness, and others which express people feeling in a better way than positive or negative speech.

5.3. Future work

In this research, an attempt is made to design and develop Multi-class subjectivity detection and sentiment analysis using machine learning approach: a case study on Amharic social media posts on covid-19. Arriving at a full-fledged subjectivity detection and sentiment analysis for the specified language is time consuming and involves coordination of team efforts that comprise linguistic Professional, computing science professional and other people is required to make sentiment analysis system with full functionality and a better performance. The following are some of the recommendations for further research and improvement:

- ✓ In this research work we only focuses on sentence level sentiment analysis of Amharic sentence but feature researcher can be scale up the scope in to sentiment analysis for Amharic Document like books. From sentence to Document level.

- ✓ In our work stemming is not a part of our pre-processing task. So, we recommend to other to examine the effect of Amharic stemmer in sentence level sentiment analysis.
- ✓ In this work we only emphases on Amharic text but we also recommend to include image and audio data to perform multimodal sentiment analysis.
- ✓ Finally, we recommend to study the accuracy of different machine learning approaches regarding to the vary in the number of classes.

Reference

[1] W. You, S. Know, and A. The, “Disease what you should know 2019 novel,” 2019.

[2] H. Policy, “COVID-19 in Ethiopia : Assessment of How the Ethiopian Government has Executed Administrative Actions and Managed Risk Communications and Community Engagement,” 2020.

[3] G. Mutangadura, N. Pheto, H. Dagmar, R. Aishwarya, and X. Liu, “Impact of Covid-19 and Responses in Landlocked,” 2021.

[4] J. Abbas, D. Wang, Z. Su, and A. Ziapour, “The role of social media in the advent of covid-19 pandemic: Crisis management, mental health challenges and implications,” *Risk Manag. Healthc. Policy*, vol. 14, no. March, pp. 1917–1932, 2021, doi: 10.2147/RMHP.S284313.

[5] A. D. Andrea, F. Ferri, and P. Grifoni, “Approaches , Tools and Applications for Sentiment Analysis Implementation,” no. January 2016, 2015, doi: 10.5120/ijca2015905866.

[6] H. Zou, X. Tang, B. Xie, and B. Liu, “Sentiment classification using machine learning techniques with syntax features,” *Proc. - 2015 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2015*, pp. 175–176, 2016, doi: 10.1109/CSCI.2015.44.

[7] M. D. Devika, C. Sunitha, and A. Ganesh, “Sentiment Analysis: A Comparative Study on Different Approaches,” *Procedia Comput. Sci.*, vol. 87, pp. 44–49, 2016, doi: 10.1016/j.procs.2016.05.124.

[8] S. Behdenna, F. Barigou, and G. Belalem, “EAI Endorsed Transactions Document Level Sentiment Analysis : A survey,” vol. 4, no. 1, pp. 1–8, 2017.

[9] V. K. Bongirwar, “A Survey on Sentence Level Sentiment Analysis,” vol. 3, no. 3, pp. 110–113, 2015.

[10] D. Chilot, “Public sentiment analysis for Amharic news,” 2019.

[11] K. Schouten and F. Frasincar, “Survey on Aspect-Level Sentiment Analysis,” *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, 2016, doi: 10.1109/TKDE.2015.2485209.

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

[13] B. Liu, “Sentiment analysis and subjectivity,” *Handb. Nat. Lang. Process. Second Ed.*, pp. 627–666, 2010.

[14] Y. Teshome, “Sentence level opinion mining for amharic language,” no. June, 2019.

[15] Wondwossen Mulugeta, “A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts,” *HiLCoE J. Comput. Sci. Technol.*, vol. 2, no. 2, pp. 80–87, 2014.

[16] A. GETACHEW, “Opinion Mining From Amharic Entertainment Texts College of Natural Science Opinion Mining From Amharic Entertainment Texts,” 2014.

[17] Tulu Tilahun, “Opinion Mining From Amharic Blog,” *Unpubl. Masters Thesis,Addis Ababa Univ.*, 2013.

[18] J. Berkhout, D. Ruedin, W. Van der Brug, and G. D’Amato, “Research design,” *Polit. Migr.*, pp. 19–30, 2015, doi: 10.5771/9783845243634-64.

[19] Ž. Đ. Vujović, “Classification Model Evaluation Metrics,” no. July, pp. 2–10, 2021, doi: 10.14569/IJACSA.2021.0120670.

[20] B. Liu, “Sentence Subjectivity and Sentiment Classification,” *Sentim. Anal.*, vol. 1, pp. 70–89, 2015, doi: 10.1017/cbo9781139084789.005.

[21] A. Ramdhani, M. A. Ramdhani, and A. S. Amin, “Writing a Literature Review Research Paper : A step-by-step approach Writing a Literature Review Research Paper : A step - by - step approach,” no. July, 2014.

[22] S. Dhuria, “Sentiment Analysis : An approach in Natural Language Processing for Data

Extraction," vol. 2, no. 4, pp. 27–31, 2015.

[23] O. Bharti, "Sentiment analysis," vol. 5, no. 6, pp. 625–633, 2016.

[24] G. Ritchie, "Survey of natural language processing," *ACM SIGART Bull.*, vol. 5, no. 80, p. 61, 1982, doi: 10.1145/1056176.1056184.

[25] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Am. Med. Informatics Assoc.*, vol. 18, no. 5, pp. 544–551, 2011, doi: 10.1136/amiajnl-2011-000464.

[26] Desalegn, *Desalegn Asfawwesen*. 2015.

[27] F. Gereme, W. Zhu, T. Ayall, and D. Alemu, "Combating fake news in 'low-resource' languages: Amharic fake news detection accompanied by resource crafting," *Inf.*, vol. 12, no. 1, pp. 1–9, 2021, doi: 10.3390/info12010020.

[28] B. A. Hordofa, "Event Extraction and Representation Model from News Articles," vol. 16, no. 3, pp. 1–8, 2020.

[29] B. Ymam, "የአማርኛ ሰዋስው የተሻሻለው ሁለተኛው እትም," *የኢትዮጵያ ቋንቋዎችና ስነ ጽሁፍ*. pp. 1–492, 2000.

[30] M. T. Khan, M. Durrani, A. Ali, I. Inayat, S. Khalid, and K. H. Khan, "Sentiment analysis and the complex natural language," *Complex Adapt. Syst. Model.*, vol. 4, no. 1, 2016, doi: 10.1186/s40294-016-0016-9.

[31] S. Alhojely, *Sentiment Analysis and Opinion Mining: A Survey*, vol. 150, no. 6. 2016.

[32] M. Farhadloo and E. Rolland, "Fundamentals of sentiment analysis and its applications," *Stud. Comput. Intell.*, vol. 639, no. August 2018, pp. 1–24, 2016, doi: 10.1007/978-3-319-30319-2_1.

[33] S. Rani, "Sentiment Analysis: A Survey," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. V, no. VIII, pp. 1957–1963, 2017, doi: 10.22214/ijraset.2017.8276.

[34] Y. Lin, X. Wang, and A. Zhou, "Sentiment analysis and opinon mining," *Opin. Anal. Online Rev.*, no. May, pp. 79–94, 2016, doi: 10.1142/9789813100459_0007.

[35] S. K. Jain and P. Singh, "Systematic Survey on Sentiment Analysis," *ICSCCC 2018 - 1st Int. Conf. Secur. Cyber Comput. Commun.*, pp. 561–565, 2018, doi: 10.1109/ICSCCC.2018.8703370.

[36] B. Jadav and V. Vaghel, "Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis," *Int. J. Comput. Appl.*, vol. 146, no. 13, pp. 26–30, 2016.

[37] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Comput. Sci.*, vol. 17, pp. 26–32, 2013, doi: 10.1016/j.procs.2013.05.005.

[38] A. Tabassum and R. R. Patil, "A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing," *Int. Res. J. Eng. Technol.*, no. June, pp. 4864–4867, 2020, [Online]. Available: www.irjet.net.

[39] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organ. Res. Methods*, vol. 25, no. 1, pp. 114–146, 2020, doi: 10.1177/1094428120971683.

[40] S. Kannan and V. Gurusamy, "Preprocessing Techniques for Text Mining (PDF Download Available)," no. October 2014, 2014, [Online]. Available: https://www.researchgate.net/publication/273127322_Preprocessing_Techniques_for_Text_Mining.

[41] B. Liu, "Sentiment analysis and subjectivity," *Handb. Nat. Lang. Process. Second Ed.*, no. January 2010, pp. 627–666, 2010.

[42] M. Karamibekr and A. A. Ghorbani, "Sentence subjectivity analysis in social domains," *Proc. - 2013 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2013*, vol. 1, pp. 268–275, 2013, doi: 10.1109/WI-IAT.2013.39.

[43] B. Seerat, "Opinion Mining : Issues and Challenges (A survey)," vol. 49, no. 9, pp. 42–

51, 2012.

[44] S. S. Kamble and P. A. R. Itkikar, “Study of supervised machine learning approaches for sentiment analysis,” *Int. Res. J. Eng. Technol.*, vol. 05, no. 04, pp. 3045–3047, 2018, [Online]. Available: https://www.academia.edu/download/56967026/IRJET-V5I4673.pdf.

[45] J. E. T Akinsola, A. Jet, and H. J. O, “Supervised Machine Learning Algorithms,” *Int. J. Comput. Trends Technol.*, vol. 48, no. 8, pp. 2760–2769, 2017, [Online]. Available: http://www.ijcttjournal.org.

[46] P. Chauhan, “Sentiment Analysis: A Comparative Study of Supervised Machine Learning Algorithms Using Rapid miner,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. V, no. XI, pp. 80–89, 2017, doi: 10.22214/ijraset.2017.11011.

[47] S. Kasthuri, L. Jayasimman, and A. N. Jebaseeli, “Procedure of Opinion Mining and Sentiment Analysis Techniques: A Survey,” *Mach. Learn.*, pp. 573–575, 2016, [Online]. Available: http://ijarcce.com/upload/2015/september-15/IJARCCE 19.pdf.

[48] N. Zainuddin and A. Selamat, “Sentiment analysis using Support Vector Machine,” *I4CT 2014 - 1st Int. Conf. Comput. Commun. Control Technol. Proc.*, no. May 2016, pp. 333–337, 2014, doi: 10.1109/I4CT.2014.6914200.

[49] C. C. Chen and Y. De Tseng, “Quality evaluation of product reviews using an information quality framework,” *Decis. Support Syst.*, vol. 50, no. 4, pp. 755–768, 2011, doi: 10.1016/j.dss.2010.08.023.

[50] Y. M. Li and T. Y. Li, “Deriving market intelligence from microblogs,” *Decis. Support Syst.*, vol. 55, no. 1, pp. 206–217, 2013, doi: 10.1016/j.dss.2013.01.023.

[51] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*, vol. 9781107057. 2013.

[52] Z. Chen, A. Mukherjee, and B. Liu, “Aspect Extraction with Automated Prior Knowledge Learning,” pp. 347–358, 2014.

[53] A. Sadia, F. Khan, and F. Bashir, "An overview of lexicon-based approach for sentiment analysis," *Int. Electr. Eng. Conf.*, no. IEEC, pp. 1–6, 2018.

[54] S. Umar, M. M, F. Azhar, S. Malik, and G. Samdani, "Sentiment Analysis Approaches and Applications: A Survey," *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 1–9, 2018, doi: 10.5120/ijca2018916630.

[55] F. Keshtkar and D. Inkpen, "A bootstrapping method for extracting paraphrases of emotion expressions from texts," *Comput. Intell.*, vol. 29, no. 3, pp. 417–435, 2013, doi: 10.1111/j.1467-8640.2012.00458.x.

[56] A. Mudinas, D. Zhang, and M. Levene, "[doi 10.1145_2346676.2346681] Mudinas, Andrius; Zhang, Dell; Levene, Mark -- [ACM Press the First International Workshop - Beijing, China (2012.08.12-2012.08.12)] Proceedings of the First Internatio (1).pdf," 2012.

[57] S. Gebremeskel, "Sentiment mining model for opinionated amharic texts," p. Angeles, L., Advocacy, S., Location, O. (2002)., 2010.

[58] N. Wayessa and S. Abas, "Multi-Class Sentiment Analysis from Afaan Oromo Text Based On Supervised Machine Learning Approaches," *Int. J. Res. Stud. Sci. Eng. Technol.*, vol. 7, no. 7, pp. 10–18, 2020.

[59] M. Kassa, "Sentiment analysis for classifying Amharic opinionated text," *DEGREE MASTERS Sci. Comput. Sci. ADDIS ABABA Univ.*, 2013.

[60] B. T. Aragaw, "Sentence level Amharic Sentiment Analysis Model: A Combined Approach," vol. 151, pp. 10–17, 2015, doi: 10.1145/3132847.3132886.

[61] A. S. Poriyath and A. Joseph, "Sentiment Analysis on Facebook Comments," pp. 1–6, 2021.

[62] Y. Roh, G. Heo, and S. E. Whang, "A Survey on Data Collection for Machine Learning: A Big Data-AI Integration Perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1328–1347, 2021, doi: 10.1109/TKDE.2019.2946162.

[63] and L. B. Gebreegziabher, Mulu, “English-Amharic statistical machine translation,” *SLTU-Workshop Spok. Lang. Technol. Under-Resourced Lang.*, 2012.

[64] C. W. Hsu and C. J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Trans. Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002, doi: 10.1109/72.991427.

[65] R. Katuwal and P. N. Suganthan, “Enhancing Multi-Class Classification of Random Forest using Random Vector Functional Neural Network and Oblique Decision Surfaces,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 2018-July, 2018, doi: 10.1109/IJCNN.2018.8489738.

[66] M. Hossin, “A review on evaluation metrics for Data classification evaluation,” no. August, 2016, doi: 10.5121/ijdkp.2015.5201.

Appendix I: Dataset Annotation Guideline

በጅማ ቴክኖሎጂ ኢንስቲቱት በኢንፎርሜሽን ቴክኖሎጂ ትምህርት ክፍል ለሁለተኛ ዲግሪ ማሟያ ለሁለተኛ ዲግሪ የቋንቋ እና የሳይኮሎጂ ተማሪዎች የቀረበ የስሜት ትንተና ማስተካከል እና ምደባ ቅፅ፡፡

- ከኢትዮጵያ የጤና ጥበቃ ሚኒስቴር ኦፊሴላዊ የፌስቡክ ገጽ የተሰበሰቡ የአማረኛ ፅሁፎች አብረው ተያይዘዋል እነዚህ ፅሁፎች በውስጣቸው የተለያዩ እውነታዎችን እና የሰዎች ስሜቶችን ይዘዋል፡፡
- እነዚህን የአማረኛ ፅሑፎች በማንበብ በተመራማሪዎቹ በተሰጠው ህግ መሰረት አግባብነቱን የጠበቀ የአማረኛ ቋንቋ አገባብ እና ሌሎች የቋንቋ መሰረታዊ ህጎችን በጠበቀ መልኩ እንዲያዘጋጁልን ሲሉ የምርምሩ ባለቤቶች የሚከተሉትን ህጎች አቅርበዋል፡፡

<u>ለፅሑፍ ምደባ የሚረዱ ደንብ እና መመርያዎች</u>

1. እባክዎን በፅሁፍ ውስጥ የተለያዩ የአማረኛ ሰዋሰው እና ሌሎች ተጓዳኝ ችግሮች ካሉ በማስተካከል ይጀምሩ፡፡
2. በተሰጠው ፅሁፍ ውስጥ ስለ ኮሮና ቫይረስ እውነታዊ የሆነ መረጃ ከያዘ Objective ብለው ይወክሉ፡፡ የግለሰቡን ስሜት የሚገልፅ ከሆነ ደግሞ Subjective በማለት ያስተካክሉ ከኮሮና ጋ ያልተያያዘ ፅሁፍ ከሆን Non related ብለው ይወክሉ፡፡
3. የግለሰቡን ስሜት በሚገልፀው ፅሁፍ ውስጥ የሚከተሉት ስሜቶች መኖራቸዉን ያረጋግጡ፡፡

ሀ) ስለ ተስፋ የሚያወራ ፅሁፍ ከሆን.............................Hope የሚል ፅሁፍ ይወከሉበት፡፡

ለ) በፅሁፍ ውስጥ ንዴት የሚያወሳ ከሆነ .... Anger የሚል ፅሁፍ ይወከሉበት፡፡

ሐ) በፅሁፍ ውስጥ ፍርሀት የሚያወሳ ከሆነ.................... Fear የሚል ፅሁፍ ይወከሉበት፡፡

መ) በፅሁፍ ውስጥ ሀዘን የሚያወሳ ከሆነ....................... Sadness የሚል ፅሁፍ ይወከሉበት፡፡

ሠ) በፅሁፍ ውስጥ ግራመጋባት የሚያወሳ ከሆነ.............. Confuses የሚል ፅሁፍ ይወከሉበት፡፡

ረ) በፅሁፍ ውስጥ ከዚህ የተለየ ስሜት የሚታይ ከሆነ .... Others የሚል ፅሁፍ ይወከሉበት፡፡

ይህንን መመሪያ በማንበብ እና በመረዳት ከላይ በተሰጡት ክፍል ወይም መደብ እንዲመድቡልን በአክብሮት እንጠይቃለን።

Appendix II: Labeled Dataset Sample

2	**Messsage**	**Subjectivity**	**Class**
3	በሽታውን ለማግኘት እየተከናወነ ያለው ሌላ የምርመራ ዘዴ ደግሞ	Objective	none
4	መንግስት ሀገር አቋራጭ ትራንስፖርትን ለተወሰነ ጊዜ ቢያግደው ጥ	Subjective	Other
5	በሽታውን ለመቆጣጠር ወደ ክልሎች ሚደረግ ትራንስፖርት መቋረ	Subjective	Fear
6	በብዛት የምያዙ በንክክ ሳይሆን በማይታዎቅ ስለሆነ ታድያ ጥንቃቄ	Subjective	Confussion
7	በተለየ በኢትዮጵያና በሶማሊያ እንዲሁም በኬንያና በሱዳን ድንበሮ	Non-related	none
8	በተለይም ይህን ገንዘብ ያላቸው ተቀናቃኝዎች በድንበሩ አካባቢ ከ	Non-related	none
9	በደቡብ ክልል ያሉ የምርመራ ማሽኖች ግን በርግጥ በትክክል የሚሰ	Subjective	Confussion
10	አምላካችን ሆይ ጩኸታችንን ስማ ልመናችንንም ተቀበል እንደ በደ	Subjective	Hope
11	አቤቱ አምላክ ሆይ ከቀን ወደቀን ነገሮች እየከበዱ የሚች ቁጥር እንደ	Subjective	Sadness
12	አብዛኛዎቹ ምልክቶች በ 10 ቀናት ጊዜ ውስጥ የሚቀንሱ ሲሆን ታ	Objective	none
13	እረ ጉድ ነው ዘንድሮ እንዴት ነው ነገሩ እረ አምላክ ሆይ እኛ በምን	Subjective	Sadness
14	ከህገ ወጥ ሰዎች መውጣትና መግባት ጋርም የአገሪቱ የውጪ ምንዛ	Non-related	none
15	ከኮቪድ-19 በሽታ ከተያዘ ሰው ጋር ንክኪ ያላቸውን ሰዎች ማፈላለ	Objective	none
16	ክፍሉ ሕጋዊ ፈቃድ ይኑርው አይኑረው የማይታወቅ የፖለቲካ ድርጅ	Non-related	none
17	ኸረ ወይኔ ምን ይሻላል ሕዝቡም ጥንቃቄውን በጣም ችላ እያለ ነው	Subjective	Fear
18	የሥራዉ ባህሪ ተጋላጭነት ያለው የሚትሉትን የሥራ አይነቱን ለምን	Subjective	Confussion