SENTIMENT ANALYSIS ON TIGRAY TELEVISION SERVICES: A RULE-BASED APPROACH

A Thesis Presented

by

Kiros Aynalem Gebregiorgis

to

The Faculty of Informatics

of

St. Mary's University

In Partial Fulfillment of the Requirements
for the Degree of Master of Science

in

Computer Science

June 2022

ACCEPTANCE

Sentiment Analysis on Tigray TV Service Rule-Based Approach

By

Kiros Aynalem Gebregiorgis

Accepted by the Faculty of Informatics, St. Mary's University, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

Thesis Examination Committee:

____

Internal Examiner
{Michael Melese Woldeyohannis (Ph.D.), Signature, and Date}

____

External Examiner
{Minale Ashagrie (Ph.D.), Signature, and Date}

____

Dean, Faculty of Informatics
{Alemebante Mulu kumlign (Ph.D.), Signature, and Date}

{Date of Defense}
June 2022

DECLARATION

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for the thesis work have been duly acknowledged.

Kiros Aynalem Gebregiorgis

Full Name of Student

____

Signature

Addis Ababa

Ethiopia

This thesis has been submitted for examination with my approval as advisor.

Million Meshesh (Ph.D.)

Full Name of Advisor

million

Signature

Addis Ababa

Ethiopia

June 2022

Acknowledgments

First and foremost, I would like to express my heartfelt thanks to the Almighty God, for giving me the strength, determination, endurance, and wisdom to bring this thesis to completion. And also, I would like to express my heartfelt and deepest sense of gratitude to my advisor Dr. Million Meshesha for his constructive comments, suggestions, guidance, and inspiring and enlightening ideas. Starting from shaping and reshaping the title of this thesis, his support and encouragement were high and without all these efforts, this work wouldn't have been a success at all. Next, I would like to thank Mr. Asemelsh for their help on linguistic parts and valuable comments and suggestions on this study.

I would also like to thank my family, especially my mother Gezu Fitsum, and my father Aynalem Gebregiorgis for their love and endless support and encouragement during my academic career. I take this opportunity to express my deepest gratitude to my wife Meseret. Thank her for her, love, care, patience, and moral encouragement throughout my study.

Last but not least, I would like to thank my friends for sharing their knowledge and experience, and for providing constructive comments and suggestions for the conclusion of this article.

Table of Contents

List of Figures

List of Tables

List of Algorithms

List of Acronyms and Abbreviations

BOW	Bag of Words
CNN	Convolutional Neural Network
DTM	Document Term Matrix
FNR	False Negative Rate
FPR	False Negative Rate
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
POS	Part Of Speech
PyMySQL	pure-Python MySQL
RNN	Recurrent Neural Network
SA	Sentiment Analysis
SVM	Support Vector Machines (SVM)
TNR	True Negative Rate
TPR	True Positive Rate
TV	Televison

ABSTRACT

Sentiment analysis (SA) is an ongoing research field in the field of text mining. SA is the calculation and processing of the opinions, emotions, and subjectivity of the text. The comments given by viewers of the program reflect whether the program is positive (positive increment) or negative (negative decrement) or neutral. SA can analyze a given text into predefined categories based on emotional terms that appear in self-righteous documents, such as positive, incremental positive, negative, reduced negative, or neutral. These opinions need to be explored, analyzed, and organized in order to make better decisions.

Early related researchers did not fully consider sentiment analysis in Tigrigna which is very important for identifying the polarity of emotions. They also did not consider the irony and ladder of expressions. And they only considered positive and negative polarity, but it is important to consider inverter words that change polarity. In this study, these gaps are attempted using NLP technology.

The sentiment analysis system uses rule-based and dictionary-based methods to resolve polarity. The questionnaire we used to do this study was to prepared and collect comments from Facebook and the website. Audience/non-audience comments were collected from website/Facebook pages, focus group discussions, and distribution of open-ended questioners.

The experiment uses 1633 (one thousand six hundred thirty-three) sentiment comments and four target research fields. The average accuracy, precession, recall, and f-score are 0.84, 0.94, 0.84, and 0.87, respectively. The experimental results using the comment viewer show the effectiveness of the system and the main limitation of this study was our inability to collect sufficient data. Hence, further research needs to be done to prepare a standardized data set that canusedable for experimentation and following the progress of the study.

Keywords: Polarity, Opinionated Documents, Sentiment Analysis, Focus group discussion, NLP Technology; Rule-based Approach.

CHAPTER ONE

INTRODUCTION

1.1 Background

The media landscape of the world and Ethiopia is constantly evolving, and new TV channels and products are constantly being launched to attract available audiences [1]. This is gradually shifting the media market of Ethiopia, where there are many opportunities for many people, to the entire Ethiopian region, prompting media companies to become innovative enough to make ends meet. This has led to many media companies' layoffs to make ends meet [2]. However, the reasons for the decline in viewing revenue have not been fully considered, because there are few data insights in this area. In Ethiopia, the situation has become worse due to the purchase of televisions and cable services that rarely use pay-TV to purchase information. Therefore, it is difficult to detect any changes in audience perception and interest.

Although information technology is not new in the technical field, it has not been fully utilized not only in Ethiopia but also in the world's media field, especially the television field [3]. In the new era of the convergence of media and technology, understanding the audience's views on TV shows also helps to understand which marketing strategy to adopt [4]. This would secure and perhaps enhance the marketing income generated by media houses. However, few companies such as Netflix have made good use of artificial intelligence to determine which shows are more popular than others, resulting in the success of huge shows such as the political drama House of Cards [4].

Currently, Tigray TV has not deployed specific data-driven methods to make the above decisions. Although cable television is not common, sentiment analysis using social media data sets can still be used to obtain information about the viewer behavior of a particular TV show or character. There are currently 311 thousand in Tigray as of 2021 (Tigray TV) [Tigrai TV - YouTube]. There is a need to address the missing link in the audience behavior that is yet to be captured. Thus, there are a few proofs of data being used to check audience perspectives regarding a given TV show.

As pointed out in [3], **Sentiment Analysis** has many challenges to deal with. These challenges include the following.

- **Sarcasm Detection**

In the sarcastic text, people express their negative sentiments using positive words. This fact allows sarcasm to easily cheat sentiment analysis models unless they're specifically designed to take its possibility into account [6].

- **Negation Detection**

In linguistics, negation is a way of reversing the polarity of words, phrases, and even sentences. Researchers use different linguistic rules to identify whether negation is occurring, but it's also important to determine the range of the words that are affected by negation words [6].

- **Word Ambiguity**

Word ambiguity is another pitfall you will face working on a sentiment analysis problem. The problem of word ambiguity is the impossibility to define polarity in advance because the polarity for some words is strongly dependent on the sentence context [6].

- **Multipolarity**

Sometimes, a given sentence or document or whatever unit of text we would like to analyze will exhibit multipolarity. In these cases, having only the total result of the analysis can be misleading, very much like how an average can sometimes hide valuable information about all the numbers that went into it.

Some sentiment analysis models will assign a negative or a neutral polarity to this sentence. To deal with such situations, a sentiment analysis model must assign a polarity to each aspect in the sentence; here, "audio" is an aspect assigned a positive polarity, and "display" is a separate aspect with a negative polarity [6].

1.2. Motivation

Sentiment analysis involves textbook analysis, and the process of natural languages to determine and estimate the private information from source data. In real-world business companies, governments, political movements, politicians and other associations post their products, service, strategy, and other issues on social media platforms to get opinions and feedback from consumers, the public, and citizens. Social media druggies can also express their feeling, and station on the posted motifs through colorful natural languages. But, there are challenges to analyse the sentiments which are written in the Tigrigna language on the posted issues. To the stylish of our knowledge, there's no exploration done in Tigrigna sentiment analysis on social media for the Tigrigna language. Due to the absence of Tigrigna sentiment analysis on social media thousands of commentary, feedback, and opinions written in Tigrigna textbooks cannot be anatomized. In addition, business companies, governments, and other associations cannot know consumers, the public, and citizens' opinions and feedback.

On their product, service, program, and strategies when the sentiment textbooks are written in the Tigrigna language. This motivated us to develop a Tigrigna sentiment analysis system that helps to dissect sentiment textbooks written in the Tigrigna language on social media platforms.

1.3 Statement of the Problem

The media industry is currently facing unprecedented challenges. The company is trying to strike a balance between reducing costs and maximizing profits while keeping up with technological changes [7]. To determine the performance and ratings of TV shows, media companies found this to be a daunting task. The program manager of the Ethiopian Broadcasting Group explained in an interview that the process of determining the popularity of a particular TV program requires the use of focus groups selected by the selected research institution [7]. First, the program group watches the selected TV program and then continues to give the focus group a chance to do the same things.

These focus groups are selected based on the target audience. For example, if a TV show is aimed at a young audience, a young focus group will be collected. Since the national media does not have much influence on the quality of the focus groups provided, the results of the group on the popularity of a particular program may be subject to subjectivity. Then, the same focus group would receive a call after the episode to provide feedback on how the show was performed. Another method used is the free SMS service, which can send feedback to the radio station.

There are various media reports describing the turbulent times that Ethiopian media companies are facing to continue to operate. In 2021, the Ethiopian Broadcasting Corporation announced plans to lay off about 50 employees, which they called 'a restructuring process [8]. This notwithstanding, there is little use of big data insights about TV viewership within the media house. Tapping into the distinct perceptivity given by artificial intelligence would guarantee media houses make the right. These result in challenges in making the right business decisions.

In the case of TV, sentiment analysis has not been fully utilized to provide insight into audience behavior. This leads to decisions made based on perception. However, this should not be the case. With the competence to mine data from prismatic sources including social media at like low costs, machine-learned algorithmic models can be designed to prognosticate spectators' demeanor [8].

It is, therefore, the aim of this study to explore and apply sentiment analysis and the opportunity for research to be done on sentiment analysis to determine the popularity of Tigray TV using a rule-based approach.

1.4 Research Questions

This study attempts to address the problem by answering the following research questions:

I. What are the problems encountered with the current methods used by Tigray TV in gauging the popularity of a show?
II. Which algorithms are used in data mining and predicting the popularity of Tigray TV shows?
III. How a predictive model is developed based on sentiment analysis on Social media?

1.5 Objective of the study

1.5.1 General Objective

The general objective of this research is to design a rule-based sentiment analysis on opinions posted via social media that gauges the popularity of Tigray TV shows for effective determination of data consequent through SA.

1.5.2 Specific Objectives

To achieve the general objective of this research, the following specific objectives are formulated

i. To investigate and identify methods and algorithms associated with the current methods used by media houses in gauging the popularity of a show.
ii. To collect users' opinions on Tigray TV shows and prepare the dataset for experimentation
iii. To design a rule-based sentiment analysis model.
iv. To test the performance of the proposed sentiment analysis.

1.6. Significance of the study

Sentiment analysis is an automated process of analyzing feelings (i.e. attitudes, emotions, thoughts, opinions, etc.) by making use of Natural Language Processing NLP tools. Natural Language Processing purposes to comprehend and create a characteristic language by utilizing essential techniques and tools. Usually, other than distinguishing the feeling, a Sentiment analyzer extracts the behaviors of articulation or expression. Accordingly, some benefits of sentiment analysis of comments given on Tigray TV shows are the following.

Improve Customer Experience.

When your customers buy something, you need to keep them loyal to your brand for as long as possible and become a communicator of your brand. Your customers can become your micro-influencers.

Many factors can incredible customer service, such as on-time delivery, responsiveness to online networks, and adequate compensation for any errors. Measuring sentiment analysis online encourages you to avoid making your customers ignore and angry. For instance, customers drop a brand or product only after one bad customer service experience [8].

Improve lead generation

One of the primary uses of sentiment analysis is the generation of leads. It's possible to produce leads by modifying your marketing movements, perfecting your product quality, and having great patron service. Happy & loyal customers, acting as your brand ambassadors, will eventually bring you new customers [9].

Boosts sales revenue

The biggest advantage of doing sentiment analysis is to increase sales revenue. The result of effective marketing is to increase sales revenue and improve customer service and quality. This study is a reasonable and easy-to-use sentiment analysis software solution. It helps you to listen to what customers say and use the information to direct your business choices. Your customers are your bread and butter. Tuning in to everything they might do, and moving with them, can be a distinct advantage for your bottom line [9].

Other people’s thoughts and feelings have always been important information for most of us in the decision-making process [10]. A rudimental task in sentiment analysis is classifying the opposition of a given textbook at the document, judgment, or quality/aspect standing whether the expressed opinion in a document, a judgment, or something quality/aspect is positive or negative. The meaning of this article is to use focus group discussions and open-ended questioners to collect data about the show from the Tigray TV website to determine the positive or negative polarity of a given text. And can automatically analyze the sentiment of a large number of collected comments before making a decision.

Therefore, this research can help the Tigray TV broadcast corporation to improve its services in the future. The results of the research can be used as an input to the development of a full-fledged opinion mining system for Tigrigna languages. Another significance of the study is that, and the output can be used as input data for recommender and opinion retrieval/search systems. And the system can be used to answer people's opinions or feeling questions. Generally, the significance of the study can be seen from the point of the society, the researcher, and the owners of Tigray TV program Managers.

1.7. Methodology of the study

The purpose of the methodology is to state the choice of way of doing the disquisition. Research methodology refers to the procedural steps used to solve a problem, and figure out issues such as the methods of data collection, processing, and presentation [11].

1.7.1 Research Design

This study follows experimental research. This involves having a set of variables and manipulating them while checking the results. The experimental results show that the accuracy is satisfactory and prove that it is reasonable to calculate the polarity score by the proposed method when the main factor is found to be the body structure. Although this work is based on ontology, it relies on a dictionary of opinion terms to assign weights to sentiment terms. Feature extraction requires ontology.

1.7.2 Data Collection and preparation

Data was collected from Tigray TV broadcast viewers who have given comments using the data online from questionnaires related to TV shows. Some of the sources of the data include Tigray TV Facebook, website, and blog. In addition distribution of open questionnaires was done to have enough data for the experiment. After the data is collected, pre-processing tasks were applied to clean and construct the final data set used for experimentation.

Data preparation tasks are usually performed multiple times depending on the quality and size of the initial dataset. Mainly tasks such as normalization, tokenization, and stemming of the data are performed to come up with the final appropriate dataset for the selected algorithms. Nowadays there are only around a million viewers of Tigray television, due to getting information from different social media and other methodologies, such as the website of Tigray, Facebook, online forms, and questionnaire. The target groups from which comments were collected are university students of Addis Ababa 6 kilo campus, Addis Ababa Gofa Sefer, and Gofa Mebrat hail condominium areas and website/Facebook page. They are good for our study or research because youth spent much time in the Tigray program. Especially, an open-ended questionnaire was distributed among these people (viewers/non-viewers of the program) and focusing group discussion has been done here in these areas. We are used to collecting comments from the Tigray website page (http://www.tmma.gov.et) and Facebook by using video screenshots.

1.7.3 Implementation tools

Different open-source tools and programming languages are used to implement the system prototype. Python programming language is used primarily to implement the system because of the following reasons:

- It is suitable for natural language processing.
- Simple and powerful programming language with excellent features and extensive libraries
- It contains a lot of packages that help us to do code re-usability

The sentiment analysis system is developed and tested on Lenovo laptop computer:

- Laptop computer with windows 10 ultimate operating system, Intel Core i7 with 2.4 GHz processor speed, 8.0GB RAM, and 1 TB hard disk capacity
- Added software components used to develop and test our system are NLTK, Notepad++, and PyMySQL.

Python is a simple yet powerful programming language with exceptional functionality for processing linguistic data [12]. It is highly readable, allows data and methods to encapsulate, and contains an extensive library including components for graphical programming, numerical programming, and web connectivity. For this work, version 3.6 of the python 64-bits is used. It is used to develop the prototype of the sentiment analysis system. NLTK is a natural language processing toolkit that can be used to build NLP programs in python. It is an open-source toolkit that contains an open-source python module, linguistic data, and documentation for research and development in the natural language processing field [12]. It provides a basic class for representing data relevant to NLP, and a standard interface for performing tasks such as text classification, part-of-speech tagging, and syntactic parsing. In this work, NLTK is used for sentence and word tokenization to split the input sentiment sentences into lists of words.

Notepad++ is a free source code editor and the replacement for Notepad which supports several languages. In this work, we used the Notepad++ version of 7.6.1 with 64-bits which is used to build and edit the lexicons.

PyMySQL is the package that contains pure Python on the MySQL client library and is used to connect and access MySQL databases.

1.7.4. Evaluation methods

The experiment is done to measure the overall performance of the developed sentiment analysis system. In this research work, a total of 1633 (One thousand six hundred thirty-three) sentiment sentences were used to test the accuracy of the system. The following exists as of evaluation of sentiment analysis using a sentiment dictionary.

- A Survey of Sentiment Analysis describes sentiment analysis research work from basic to the most advanced methods [12]. It interprets research based on recent research, that is, the technology of finding, extracting, organizing, and integrating evaluation information from text. That paper discussed methods of using vocabulary networks, methods of using co-occurrence information, and methods of using surrounding context information.
- There is a paper on the development and evaluation of an emotional search system, which uses susceptible words as search input for search [14]. This is to improve the recall rate and accuracy rate by correcting the vicinity of the correct feature quantity distribution.
- There is also a report on extracting reputation and evaluation expressions from the Web [15]. It proposes a method of statistically extracting evaluation expressions from the corpus, using pre-collected positive and negative reputations as the corpus.

1.8. Scope and Limitation of the study

This work deals with Tigrigna sentiment analysis on Tigray TV show. This research work only considers that the input sentiment text is grammatically correct, analyzes and classifies sentiment text written in the Tigrigna language, and only focuses on sentiment analysis at the sentence level. Therefore, grammatically incorrect text, slang, and emotions expressed through images/pictures, audio, video, and other emotional symbols are not the focus of this research. In addition, words or texts that have indirect or hidden meanings such as an idiom/ብቅኔዝዝረብ / and ambiguous words are not incorporated in this research work and many other factors affect the popularity of a show such as the characters or any other external factors that could have an impact on the views. The analysis will also limit to comments given only Tigrigna language.

Many different methods have been used to try to solve the sentiment classification problem. One of the most widely used methods involves classifying a single word or phrase with sentiment and then calculating an overall sentiment rating for a target document using some weighting [16]. Though there are different techniques applied for sentiment analysis (see section 2.), in this study an attempt is made to use the natural language processing technology for Tigrigna sentiment analysis to distribute the polarity of comments given by viewers of Tigrigna TV shows.

1.8. Thesis Organization

The remaining part of this thesis is organized as follows. Chapter Two presents the review of related literature which includes: sentiment analysis, approaches to sentiment analysis, and the linguistic behaviors of English, Amharic, and Tigrigna languages. Chapter three presents related works done on monolingual and multilingual sentiment analysis systems and related areas. The Fourth Chapter deals with the data preparation design of the rule-based approaches which are rule-based the dictionary-based approaches to identify the given comment as either positive or negative or neutral on social media. The Fifth Chapter presents the experimental results of the sentiment analysis system. Finally, the conclusion, recommendation, and future works are presented in Chapter Six.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

The purpose of this chapter is to review relevant literature and studies carried out in studying TV audience behavior using opinion mining and sentiment analysis. The various approaches taken in different scenarios to determine popularity using sentiment analysis are reviewed in this chapter, as well as the different approaches and algorithms used in classifying text.

2.2. Introducing sentiment analysis

Sentiment analysis (SA), also known as opinion mining, has attracted increasing interest. It is a hard and challenging task for language technologies, and achieving good results is much more difficult than some people think. The task of automatically classifying a text written in a natural language into a positive or negative feeling, opinion, or subjectivity [9] is sometimes so complicated those even different human annotators disagree on the classification to be assigned to a given text. Personal interpretation by an individual is different from others, and this is also affected by cultural factors and each person's experience. And the shorter the text and the worse written, the more difficult the task becomes, as in the case of messages on social networks like Twitter or Facebook [17].

Sentiment analysis is becoming a popular area of research and social media analysis, especially around user reviews and Tweets. It is a special case of text mining generally focused on identifying opinion polarity, and while it's often not very accurate, it can still be useful. Sentiment Analysis is one of the interesting applications of text analytics. Although the term is often associated with sentiment classification of documents, generally speaking, it refers to the use of text analytics approaches applied to the set of problems related to identifying and extracting subjective material in text sources [18]. It is concerned with the identification of opinions in a text and their classification as positive, negative, and neutral. Sentiment analysis refers to a broad area of natural language processing, computational linguistics and text mining that aims to determine the attitude of a speaker or writer concerning some topic Wondwossen Mulugeta [19] defined sentiment mining as a recent disciple at the crossroads of information retrieval, text mining and computational linguistics which tries to detect the opinions expressed in the natural language texts. Sentiment mining is a complex field as it involves the processing and interpretation of natural language. Hence, it must deal with natural language’s inherently ambiguous natures, the importance of context, and other complications that do not lend themselves to automation [9].

As noted by X Cheng [17], the main activities needed for building a sentiment mining system are the following:

- Development of linguistic resources e.g. build a lexicon of subjective terms.
- classification of text (entire documents, sentences) based on their content (e.g. classifying a news article either as positive or negative about the subject),
- extraction of opinion expression from text, including relations with the rest of content (e.g. recognizing an opinion, which is expressing it, who/what is the target of the opinion)
- Mining tools and visualization tools to extract meaningful information from the mined articles based on the sentiment tags.

2.3 Steps in sentiment analysis

Sentiment Analysis is the process of determining whether a piece of writing (product/movie review, social media, etc) is positive, negative, or neutral [20]. It can be used to identify the customer, or follower's attitude towards a brand through the use of variables such as situation, tone, emotion, etc. Marketers can use sentiment analysis to research the public view of their company and products or to analyze customer satisfaction. Organizations can also use this analysis to gather serious responses about problems in newly released products.

Sentiment analysis, not only benefits companies, understand how they're doing with their customers, but it also gives them a better picture of how they stack up against their competitors. For example, one may ask, if the company has 20% negative sentiment, is that bad? It depends. If the participants have a roughly 50% positive and 10% negative sentiment, while the company is 20% negative, that merits more discovery to understand the drivers of these opinions. Knowing the sentiments associated with competitors helps companies evaluate their performance and search for ways to improve [21].

How to Perform Sentiment Analysis?

Several tools provide automated sentiment analysis results. Regardless of what tool you use for sentiment analysis, these steps are the following [22].

Step 1: Data collection:

The first step of sentiment analysis involves collecting data from user-generated content contained in blogs, forums, and social networks.

These data are disorganized and expressed in different ways through the use of different vocabulary, slang, writing context, etc. Manual analysis is almost impossible. Therefore, text analysis and natural language processing are used for extraction and classification.

Step 2: Text preparation:

The main concern of text preparation is in cleaning the extracted data before analysis. No writ continents and contents that are irrelevant for the analysis are identified and eliminated; The "Subject" and "Body" are the columns that I will apply text pre-processing procedures. I pre-processed the news articles following the standard text mining procedures to extract useful features from the news contents, including tokenization, removing stop-words and lemmatization.

Tokenization

The first step of pre-processing text data is to break each sentence into individual words is called tokenization. Using a single word instead of a sentence will break the connection between the words. However, it is a common method for analyzing large amounts of text data. It is efficient and convenient for a computer to analyze text data by checking which words appear in the article and the number of times these words appear, and it is sufficient to give insightful results.

After tokenization, each news article will be converted into a list of words, symbols, numbers, and punctuation. You can also specify whether to convert every word to lowercase. The next step is to delete useless information. For example, symbols, digits, and punctuations. We will use spacy combined with remex to remove them.

Stop words

After some transformation, the news article is much cleaner, but we still see some words we do not wish, for example, "ወይ", "ንሕና", etc. The next step is to remove the useless words, namely, the stop words. Stop words are words that frequently appear in many articles, but without significant meanings. Examples of stop words are "እነ", "ናይ" and "ናቱ". These are the words that the deleted words will not interfere with the understanding of the article.

Lemmatization

Remove stop words and symbols, numbers, and punctuation, and each news article will be converted into a series of meaningful words. However, to count the number of occurrences of each word, it is necessary to remove the grammatical tense and convert each word to its original form.

Lemmatization is taking a word into its original lemma, and stemming is taking the linguistic root of a word.

Step 3: Sentiment detection:

Extracted sentences of the reviews and opinions are examined. Sentences with subjective expressions (opinions, beliefs, and views) are retained, and sentences with objective communication (facts, factual information) are discarded;

Step 4: Sentiment classification:

In this step, subjective sentences are classified into positive, negative, good, bad; like, and dislike, but classification can be made by using multiple points;

Step 5: Presentation of output:

The main goal of sentiment analysis is to convert unstructured text into meaningful information. After the analysis is complete, the test results are displayed on graphics such as pie charts, bar charts, and line charts. You can also analyze time and display it graphically, constructing an emotional timeline with selected values (frequency, percentage, and average) over time.

2.4 Sentiment Analysis and Classification Techniques

Classification is used to assign a sentiment sentence into a particular category (also referred as a class) or categories based on their sentiment polarities. A sentiment analysis system can classify sentiments in the sentence using different approaches. This Section discusses the sentiment analysis and classification techniques.

2.4.1 Lexicon-Based Approach

The lexicon-based approach involves calculating the sentiment polarity of a sentence from the semantic orientation of words or phrases in the sentence and uses the sentiment of lexicons as a collection of known and precompiled sentiment terms with their polarity values [23, 24]. The system classifies the sentiments into specific categories of sentiment classes which are positive, negative, or neutral based on the positive or negative sentiment terms that occurred in the sentence. This can be done using rule-based classifier methods, and if there are positive words only or more positive words than negative words in the sentence, the semantic orientation of the sentence is classified as positive. If there are negative words only or more negative words than positive words in the sentence, the semantic orientation is classified as negative. If there are equal numbers of positive and negative sentiment terms or there are no sentiment terms in the sentence, the semantic orientation of the sentence will be neutral and categorized in the neutral class. In this approach, overstatement and understatement words are considered (i.e., overstatement words increase the semantic orientations and understatement decreases the semantic orientations of the sentiment words in the sentence).

2.4.2 Machine Learning Approach

Machine learning is the sub-field of computer science that gives a computer to learn without being explicitly programmed [25]. The machine learning approach performs sentiment classifications based on training algorithms, the classification is on a set of selected features for a specific mission and tests on other sets whether it can detect the right feature and give the right classifications. In sentiment classification, the machine learning algorithm uses supervised and unsupervised sentiment classification models. The supervised sentiment classification model uses a large number of labeled training datasets, while the unsupervised sentiment classification model uses it when it is difficult to find the labeled training datasets. A supervised sentiment classification model can classify sentiment sentences based on the algorithms like Naïve Bayes, Maximum Entropy, and SVM classifiers [26].

Naïve Bayes classifier is a simple probabilistic model-based the on Bayes theorem [27] along with a strong independence assumption. It computes the posterior probability of class based on the distribution of words in the document. It uses Bayes‟ theorem to predict the probability that a given feature set belongs to a particular label.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Where, p(label) is the prior probability of a label or the likelihood that a random feature set the label, p(features | label) is the prior probability that a given feature set is being classified as a label p(features) is the prior probability that a given feature set occurs.

Maximum Entropy classifier is a probabilistic classifier that belongs to the class of exponential models. The models that fit the training examples are computed, where each feature corresponds to a constraint on the model. The model with the maximum entropy over all models that satisfies these constraints is selected for classification. Maximizing the entropy ensures that no biases in the classification system [28].

SVM classifier performs by constructing a hyperplane with maximal Euclidean distance to the closest training examples. This can be seen as the distance between the separating hyperplane and two parallel hyperplanes at each side, representing the boundary of the examples of one class in the feature space. It is assumed that the best generation of the classifier is obtained when the distance is maximal. If the data is linearly not separable, a hyperplane will be chosen that splits the data with the least error possible. SVM yields the best accuracy in sentiment classification [27].

2.4.3 Hybrid Approach

Both lexicon-based and machine learning-based approaches are combined to benefit from their synergy effect that rises to a hybrid approach. Researchers have proved that this combination gives improved performance in sentiment classification [24]. In the hybrid approach, sentiment lexicons are used as seed resources to detect sentiment polarities and the results from the lexicon-based method are used to train machine learning algorithms. Then, sentiment sentences are analyzed using machine learning classifiers based on the knowledge acquired from the training and the lexicon resources.

2.5 Sentiment Lexicon Generation

Words or phrases that convey positive or negative sentiments are vital for sentiment analysis and classification systems. Positive sentiment words are used to express some desired states or qualities while negative words are used to express some undesired states or qualities. Table 2-1 shows an example of sentiment wordlists.

Table 1 Sentiment Wordlists

Language	Positive sentiment words	Negative sentiment words
Tigrigna	ጽቡቅ/Sbuq ልባም/lebam ወሓለ/weHale	ሕማቕ/HmaQ ሃሳስ/hasas በሰሮ/bsero

Overall, positive and negative sentiment words are called sentiment lexicons. The sentiment lexicons can be generated using three different approaches [9]: manual approach, dictionary-based approach, and corpus-based approach.

i. ***Manual approach*** involves generating sentiment-bearing word lists manually from various resources and labeling their sentiment polarities (i.e., positive or negative) and strength of polarity values. This approach is effective for resource-poor languages and provides the best performance for sentiment analysis systems under specific domains.

ii. ***Dictionary-based approach*** involves using a dictionary that contains synonyms and antonyms of words. In this method, it uses a few seeds of sentiment words to bootstrap based on the synonym and antonym structure of the dictionary. Specifically, this method works as first a small set of sentiment words (seeds) with known positive and negative semantic orientation is collected manually, then the algorithm grows this set of words by searching in WordNet [25], thesaurus [26,29], and other online available dictionaries for their synonyms and antonyms. The seed lists will be added with the newly found words. The process iteratively keeps on adding the words until no more new words are found. At last manual inspection is used to remove errors in the lists.

iii. ***Corpus-based approach*** helps to solve the problem of finding sentiment words with context-specific orientations. This method depends on syntactic patterns or patterns that occur together along with a seed list of sentiment words to find other sentiment words in a large corpus. The corpus-based approach can be performed using the statistical method and sematic method [28].

Statistical method: this method helps to find the co-occurrence pattern or seed sentiment words using statistical techniques. The polarity of the word is determined based on the occurrence frequency of the word in the large annotated corpus of texts. If the word occurs more frequently among positive texts, then its polarity is positive. If the word occurs frequently among negative texts, then its polarity is negative. But, if it has an equal occurrence frequency, then it is a neutral word. If two words appear together frequently within the same context, then there is a high probability that they have the same polarity. In this case, the polarity of an unknown word can be determined by calculating the relative frequencies of co-occurrence with other words. This can be done using PMI [30], the semantic orientation of a given phrase is calculated by comparing its similarity to a positive word ("excellent") and its similarity negative word ("poor").

Semantic method: this technique gives sentiment values directly and uses different principles to compute the similarities between words. In this case, the principle gives similar sentiment values to semantically close words. These semantically close words can be obtained by getting the list of sentiment words and iteratively expanding the initial set with synonyms and antonyms and then determining the sentiment polarity for unknown words by the relative count of positive and negative synonyms of the word [31].

2.6. The State of the Media in Ethiopia

The media industry in Ethiopia is a thriving market, In the TV station field, Citizen TV leads with the 3% share of viewers in Tigray. Other stations are KANA TV (34 percent), EBC1 News (18 percent) and EBS TV (16 percent) closing the top four watched TV stations in Ethiopia (Kantar-Geopoll Media Measurement for Ethiopia March 2020). A 2020 report by GeoPoll.

The State of the Media in Kenya

To get an aggregate image of the state of the media in Kenya, thorough research employing both qualitative and quantitative research methods was conducted. Over three months, we talked to audiences across the country, and media personnel from different entities, reviewed documents, and utilized online sources. On the whole, the state of the media in Kenya is worrying. Legacy media have lost their sting and social media, which is on the ascendency, is a jungle requiring to be tamed. The COVID-19 pandemic has complicated hitherto existent economic challenges exacerbated by a continuous brain drain. This has weakened the media undermining trust in it and making it all but impossible to fully meet the information needs of Kenyans. But hope still abounds as the Kenyan media is resilient. The presence of international media supports organizations in the country could be exploited to address the critical skills, resources, and policy gaps that need to be plugged in for the media to thrive. This is of paramount necessity as the country gears itself for both possibly a referendum and a general election in short order. Price Water house Coopers (PwC PwC) places the Box entertainment and media demand at USD 865 million. It had also been estimated that the expenditure would cross USD 3 billion by 2021 (PwC, 2013) [32].

2.7. Tigrigna language

Tigrigna is a Semitic language family related to Amharic and Arabic that is primarily spoken in Ethiopia’s Tigray region and central and southern Eritrea. In Ethiopia’s Tigray region, it is the official language and medium of instruction for primary schools. This language was derived from the Geez language, which first appeared in writing in the 13th century [33]. It is written with a variation of Geez (Ethiopic) scripts. Tigrigna alphabets, like Amharic alphabets, are written in a tabular fashion with seven columns from left to right, with the first column representing the base letter and the remaining columns representing their derived vocal sounds.

Tigrigna, like Amharic, has a morphologically rich language, as both are Semitic languages that descended from Geez [33]. In Tigrigna, the root of verbs and most nouns are characterized by a series of consonants known as root/radicals, similar to Amharic. By concatenating vowels and non-root consonants that go with a specific morphological category around the root consonants, roots are employed to generate additional genuine words. Tigrigna creates inflectional and derivational morphemes using various affixes for non-concatenating morphological properties.

Tigrigna, like Amharic, is a highly inflectional language, with the given root of the Tigrigna word appearing in various forms. Nouns, adjectives, and verbs are the highly inflected Tigrigna word classes, and they are explored as follows.

Nouns: Number, gender, person, definiteness, and cases are all inflected into Tigrigna nouns, resulting in an inflected word containing affixes.

Adjectives: The inflected word with affixes to the specified adjectives is the result of marking Tigrigna adjectives for number, gender, person, and degree. Adjectives can be inflected by adding affixes like -ኦ/, -ት/t, -ቲ/ti, -አቶም/’atom, -አዊ/’awi, -ኣት/’at, -አዊት/’awit, ክ/k-, ዜ/z- etc. Table 4 shows the inflected Tigrigna adjective words.

Table 2 Tigrigna Inflection of Adjectives

Male	Female	Plural form	Comparative degree	Superlative degree
ዕጉስ	ዕግስቲ	ዕጉሳት	ይዕገስ	ዝተዓገሰ
ቀይሕ	ቀያሕ	ቀየሕቲ	ይቀይሕ	ዝቀይሐ
ቀጢን	ቀጣን	ቀጠንቲ	ይቀጥን	ዝቀጠነ
ፀሊም	ፀላም	ፀለምቲ	ይፅልም	ዝፀለመ
ረጉድ	ረጓድ	ረጎቲ	ይሮግድ	ዝረጎደ

Verbs: Tigrigna verbs are inflected for any combination of person, gender, number, case, tense, aspect, and mood [33]. Matching as perfective, imperfective, gerundive, jussive and imperative by employing affixes. Number, gender, person, tense, mood, and aspects are inflected into Tigrigna verbs, resulting in an inflected word with affixes to the verb stem. Perfective verbs have morphological variety due to suffixes such as ኣ/'a, ኣት/'at, ኩ/ku,ና/na, ኣቶም/'atom, ኣተን/' aten and ኣ/'a that indicates for the person, gender and numbers to the perfect verb stem. For instance, ቆረፀ, ቆረፀት, ቆረፅኩ, ቆረጽና, ቆሪፃቶም, ቆሪፀን and ቆረፃ (koretsa, koretset, koretsiku, koritsatom, koritsen and koretsa). This perfect verb is formed from the stem verb ቆረፀ/koretse which means to break". Imperfective verbs are also formed by adding affixes on the verb stem and markers for gender, person and number.

For example,ትቆርፅ, ክቆርፅ, ክቆርፃ, ክቆርፁ, ዜይቆረፁ, ዜቆረፀት, ዜቆረፅና, ይቆርፁ, ይቆርፃ (tikorts, kikorts, kikortsa ,kikortsu , zeykortsu, zeykoretset, zeykoretsina, ykortsu ykorts), in this example, the stem verb ቆርፅ/koretsewhich means 'let's break' uses both prefix and suffixes, the morphemes that attach in the stem verb such as ት/t, ክ/k, ዜ/z, and ይ/y are prefixes and ኣ/a, አ /e,, ኣት/et, ኣና/ena and ኡ/u are suffixes.

The gerundive verb from inflected by adding suffixes at the end of the gerundive verb to indicate person, gender, and number. For example, ሰሪሗ, ሰሪሕኻ, ሰሪሕኺ, ሰሪሑ, ሰሪሕና, ሰሪሕኹም, ሰሪሕኽን, ሰሪሖም and ሰሪሗን (seriHa, seriHXa, seriHXi, seriHu, seriHna, seriHXum, seriHXn, seriHom, seriHan), from this example, the stem verb ሰሪሕ/seriH varies its morphology and generate various verbs that markers of the person, gender and number. The suffixes ኣ/'a, ኻ/Xa, ኺ/Xi, ኡ/'u, ና/na, ኹም/Xum, ኽን/Xn, ም/m and ን/n attached with the inflected stem verb.

Jussive verbs, also known as mood verbs, convey a command for the first and third persons, whereas imperative verbs express the second person in singular and plural forms.

Tigrigna words are derived from other Tigrigna word classes in a significant way. Nouns, adjectives, and verbs are some of the Tigrigna word classes with a lot of derivation.

Nouns: Tigrigna nouns are derived from other word classes by adding affixes and using compound words [32]. In the case of compound words, the new noun is constructed from two separate words. Example ቤት and ትምህርቲ, ቤት and ብሌዕ provided ቤትትምህርቲ which means school and ቤትብሌዕ which means restaurant or nouns constructed by adding affixes like -ነት and -ኛ. Example, ሰብ + -ነት gives ሰብነት, ዒፋር + -ኛ generates ዒፋርኛ. In addition, nouns can be derived from verbs and adjectives.

Adjectives: similar to nouns, Tigrigna adjectives can be derived from nouns, verbs, and adjectives themselves [32]. Adding morphemes like -ኣዊ/'awi, -ዊ/wi, -ኣም/'am, and -ታይ/tay to nouns such as: ኢትዮጵያ/'ityoPya, ሃገር/hager, ነገር/neger and መቐለ/meQele that generates the following new adjectives: ኢትዮጵያዊ/'ityoPyawi, ሃገራዊ/hagerawi, ነግራም/negram and መቐለታይ/meQeletay respectively. Similarly, it can be constructed from the root verbs. For instance, the root verb ኩቡር by infixing the vowel -ኡ-, generates the adjective ኩቡር.

Verbs: unlike nouns and adjectives, Tigrigna verbs can be derived only from verbal roots and stems [33].

For example, ቅ-ብ-ር, ቅኣብኣር which provides ቀበር, and it can also be constructed from verbal stems by adding affixes like ተ and ኣ to the stem verb.

2.8. Challenges of the language in sentiment analysis

Sentiment analysis has gone a hot motif in the scientific sector. The adulthood of sentiment analysis exploration and methodologies are concentrated in the English manual. As a result, there are apparent restrictions on the researchers that are interested in sentiment analysis for the Tigrigna language [33]. Furthermore, the majority of researchers concentrate on the formal Tigrigna language [33]. Because the majority of users on social media use informal Tigrigna, sentiment research becomes more difficult. This encourages us to investigate the difficulties in English and the sentiment for informal Tigrigna language, such as the many Tigrigna dialects. Examples Tigraway, Agame, Tselimat zemen, harif, miwiziwaz, Hiray, eshi, chele, etc.

2.8.1 Challenges in sentiment analysis

When it comes to sentiment analysis challenges, there are quite a few things that companies struggle with to obtain sentiment analysis accuracy. Sentiment or emotion analysis can be difficult in natural language processing simply because machines have to be trained to analyze and understand emotions as a human brain does. This is in addition to understanding the nuances of different languages. As data science continues to evolve, sentiment analysis software can tackle these issues better. Here are the main roadblocks in analyzing sentiment [34].

1. Tone

Tone can be difficult to interpret verbally, and even more difficult to figure out in the written word. Things get even more complicated when one tries to analyze a massive volume of data that can contain both subjective and objective responses. Brands can face difficulties in finding subjective sentiments and properly analyzing them for their intended tone [34].

2. Polarity

Words such as “love” and “hate” are high on positive (+1) and negative (-1) scores in polarity. These are easy to understand. But there are in-between conjugations of words such as “not so bad” that can mean “average” and hence lie in mid-polarity (-75). Sometimes phrases like these get left out, which dilutes the sentiment score [34].

3. Sarcasm

People use irony and sarcasm in casual conversations and memes on social media. The act of expressing negative sentiment using backhanded compliments can make it difficult for sentiment analysis tools to detect the true context of what the response is implying often resulting in a higher volume of “positive” feedback than is actually [34].

4. Emojis

The problem with social media content that is text-based, like Twitter, is that they are inundated with emojis. NLP tasks are trained to be language-specific. While they can extract text from even images, emojis are a language in themselves. Most emotion analysis solutions treat emojis like special characters that are removed from the data during the process of text mining. But doing so means that companies will not receive holistic insights from the data [34].

5. Idioms

Machine learning programs don't necessarily understand a figure of speech. For example, an idiom like "not my cup of tea" will boggle the algorithm because it understands things in the literal sense. Hence, when an idiom is used in a comment or a review, the sentence can be misconstrued by the algorithm or even ignored. To overcome this problem a sentiment analysis platform needs to be trained in understanding idioms. When it comes to multiple languages, this problem becomes manifold [34].

6. Negations

Negations, given by words such as not, never, cannot, were not, etc. can confuse the ML model. For example, a machine algorithm needs to understand that a phrase that says, "I cannot go to my class reunion", means that the person intends to go to the class reunion [34].

7. Comparative sentences

Comparative sentences can be tricky because they may not always give an opinion. Much of it has to be deduced. For example, when somebody writes, "the Galaxy S20 is larger than the Apple iphone12," the sentence does not mention any negative or positive emotion but rather states a relative ordering in terms of the size of the two phones[34].

8. Employee bias

Employee feedback is valuable when it comes to shaping company culture, improving sales tactics, and reducing employee turnover. Many companies, though, find themselves struggling to parse information because of biases. These can be either from the employee, or from the perspective of the surveyor who may not take the responses of an ex-employee seriously [34].

9. Multilingual data

All the problems listed above get compounded when a mix of languages is thrown in. Each language needs a unique part-of-speech tagger, lemmatize, and grammatical constructs to understand negations. Because each language is unique, it cannot be translated into a base language like say, English, to extract insights. A simple example is if an idiom “like a fish takes to water” is translated into say, German, the idiom would have lost its meaning [34].

10. Audio-visual data

Videos are not the same as text data. The challenge is not only that videos need to be transcribed but that they may have captions that need to be analyzed for brand logos. Social media videos also come with comments in addition to the video data [34].

2.8.2 Challenges of the language for sentiment analysis

Lexical Sparsity

Sparsity is the lexical magnitude of words in a language, the higher the number of forms affecting the words of a language, the higher the sparsity that language has:

In English, the inflectional forms enjoyed, enjoying, and the derivational form enjoyable can easily be mapped with the positive sentiment word enjoy. As for morphologically rich languages, inflection is not limited to tenses and numbers but spans subjects, objects, pronouns, clitics, and gender as well. A simple example is a word love ፍቅሪ in Tigrigna; it can be inflected as I-love-you (masculine) የፍቅረካኧየ I-love-you (feminine) የፍቅረኪእየ, I-love-you (plural) የፍቅረኩም we-loved-them ነፍቅረኩምእየwe-will-love-you (plural) ነፍቅረኩምኢና reaching over 100 inflections on this front. On another front, the vocabulary derived from a word consists of prefixes, infixes, suffixes, and diacritics.

The lexical sparsity of a morphologically rich language becomes very large if there is no consistent orthography, where each form of a word may be spelled differently. Inconsistent orthography is common for languages that are solely spoken, such that if transcribed there is no standard spelling to follow, as a result, each individual expresses their tongue in-text differently.

2.9. Related works

Xiaoying Xu et al [35] proposed the principles and guidelines for manually creating a large Chinese sentiment dictionary for opinion mining. Two experiments were conducted in their work: The first experiment was to study the reliability of manual subjective labeling of terms. The second experiment is to observe the effectiveness of dictionaries in judging subjective sentence polarity. This paper shows that for the establishment of the first large-scale artificially annotated Chinese sentiment dictionary, the consistency of different annotators and the reliability of the sentence polarity judgment system are key issues. Therefore, annotation principles and guidelines need to be established. The author establishes the subjectivity and polarity English principle of terms: terms are guided by opinion mining, the lexicon constructed is only used for opinion mining subjective sentences, and words express polarity in subjective sentences. Need to have a clear emotional word annotation guide and qualified annotators to achieve the principles of building a dictionary. During the trailing push and wordbook structure, the main resource used was HowNet [20]. HowNet is an online common-sense knowledge base that reveals the connotation of the relationship between concepts and the relationship between attributes in Chinese and English equivalent word dictionaries. According to their analysis, they answered the question “What kind of words can be selected in the dictionary?” If a term has a subjective meaning in conceptual meaning or emotional meaning overtones, it can indicate the polarity in a subjective sentence, it must be selected in our dictionary.

The experimental results of the first experiment show that although the polarity of words is common in Chinese, the polarity of word meanings can be reliably marked, and due to its large scale, it can be a very useful basic resource for opinion mining and other aspects. Related areas. To evaluate dTo in practical applications (the second experiment), they built a simple sentence emotion recognition system, which includes text pre-processing, polar word detection and weight assignment, link construction, and polarity propagation. Text pre-processing is used to segment words and tag POS. In the polarity word detection, every polarity word is checked whether it is a polarity word defined in the sentiment lexicon and get the corresponding sentiment polarity if found. Every polarity word and modifier word gets the initial weight defined in the sentiment lexicon. If the polarity word is linked to a modifier word, the polarity value should be multiplied by a coefficient in the polarity propagation step. The results of the second experiment show that using the sentiment lexicon can achieve an accuracy of more than 70%.

Sigrid Maurel et al [21] use a combination method (combination of symbols and statistics) to classify opinion texts in French. The symbolic method includes a system for extracting information suitable for the corpus based on the rules of syntactic and semantic English. This method English the text sentence by sentence and extracts relationships that convey emotions. Statistical methods are based on machine learning techniques. It processes the text in one step and assigns a global opinion to the entire text at the end. The hybrid approach is used in their work to increase the quality of the results. As they indicated the experimental results show that a combination of statistical and symbolic (hybrid) approaches gives more accurate results than either method used separately. In the ground of NLP Sentiment analysis, as we give clarity about this field is the one and the most widely held research areas. Many researchers were ducted in the area of SA. Most research on sentiment analysis is focused on the English language as we indicated in the portion of related works above. Few Related research is done on Amharic languages.

The earlier related researcher did not enough consider the problem of sentiment analysis in Tigrigna which is very important to identify the polarity of the sentiment. And did not consider the irony and stairs of expressions and also they only considered positive and negative polarities but important to consider the inverter words that will change the polarity. In our paper, the gaps are held by using NLP Tec the sentiment analysis system was applied to solve the polarity by using the Rule-based and Dictionary approach. The comments were collected from the viewers/non-viewers from the website/Facebook page, focusing on group discussion, and distributing an open-ended questionnaire. The experiments are conducted using 1633(One thousand six hundred thirty-three) sentiment comments with four target research areas. The average Accuracy, Precession, Recall, and f-score respectably are 0.85, 0.93, 0.86, and 0.87. Experimental results using viewers of comments shows the effectiveness of the system.

2.9.1: Foreign works

In Japan, TV ratings are provided by Video Research Ltd, an agency that specializes in TV ratings, which started audience measurement in 1962[36]. Proposed the use of various concepts of multimedia to identify how viewers relate to the video and speech content of a given TV show. The TV ratings were acquired from Video Research, an agency that specializes in TV ratings. By using systematic random sampling, a total of 6,600 households were surveyed in Kanto, Kansai, and Nagoya, Japan. The situations were hung on samples from the homes handpicked. Using this data, they can detect an audience's behavior with its TV sets through what they refer to as the 'Boundary of a TV show.' This refers to a user tuning into a station at the start and off at the end. Another factor considered is what the researchers termed 'Transition.' This refers to the act of switching from one channel to another perhaps due to a lack of interest. The model used is shown below.

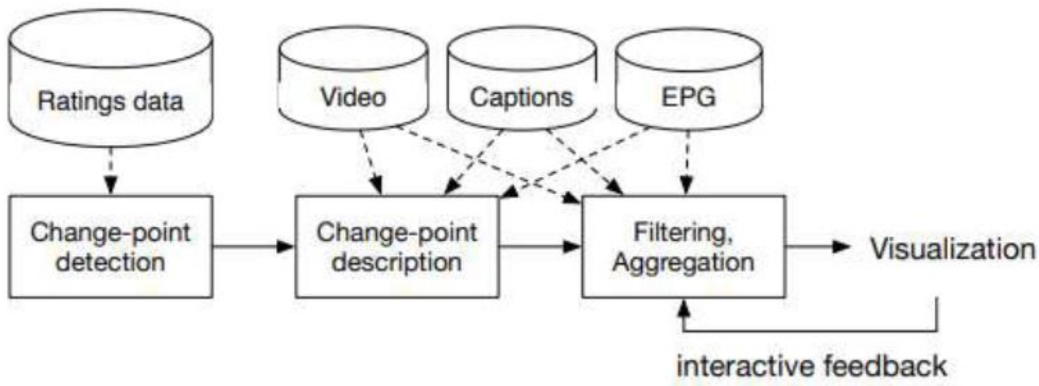

Figure 1 Using Multimedia Contents [37]

As illustrated above, the rating data would provide data on the point in a show in which viewers change their tuning to another show. Then, other insightful data would be collected from multimedia elements such as video and captions. The final product would be aggregated to provide a visualization which can be analysed.

The Nielsen Company provides TV companies with insights into how the audience perceives the TV programmes. A set of households are provided with a diary and urged to record their viewing habits for a given period. Once the insights are picked, a programme is allocated a share of the audience. Thus, for instance, if 10 percent of the households said they watch Programme X, then a conclusion is made that the programme's household rating is 10 percent. The analysis only makes use of households, not head counts. Thus, the analysis provides an aggregate figure, thus implying that the share would be higher than the rating [36].

2.9.2. Local related works

Selama [38] has evolved a sentiment mining version for opinionated Amharic texts. In this study, to decide the sentiment of reviews, the writer used period counting techniques. Based on this technique, it counts the range of sentiment phrases inside the sentence via the way of means of checking every phrase inside the organized lexicon, if the phrase seems with inside the lexicon, it took into consideration because the phrase is a sentiment phrase and assigns polarity weight values automatically, while it took into consideration as a non-sentiment phrase. The writer evolved popular and precise Amharic sentiment lexicons which encompass sentiment phrases and contextual valence shifter phrases. Sentiment terms were tagged as "+" and "-" in the lexicon, which is "+" for positive and default values assigned +2 and "-" for negative and default values assigned -2. But, the very last polarity weight cost changed into calculated via the way of means of thinking about the outcomes of context valance shifters like diminished, intensifier, and negation phrases across the sentiment phrases inside the sentence, and sooner or later English the sentiment sentence into high-quality, bad and impartial instructions primarily based totally on polarity weight values. For the reason of the experiment, the researcher gathered 303 evaluated sentences from distinct domain names and 955 opinion phrases: 411 high-quality and 544 bad phrases. However, the organized opinion phrases are not enough, and inflected Amharic sentiment phrases are not treated in this study work.

Tulu [39] advanced characteristic degree opinion mining version for Amharic texts. The goal of the takes a look at changes to decide an opinion on the capabilities of the domains. In this take a look, the writer first extracted capabilities of the area after which decided the evaluations inside the extracted capabilities through using a few rules. An opinion phrase inside the sentence changed into detected from Amharic popular reason opinion phrase lexicons that include a complete of 1001 sentiment phrases that are 578 bad and 423 wonderful phrases. The creator accumulated 484 Amharic opinions manually from hotels, universities, and hospitals for experimental activities. The effectiveness of the device changed to evaluating the use of precision, do not forget, and F-degree metrics via special experiments. From test one, the writer was given the result of a median precision of 95.2%, a median does not forget of 26.1% for characteristic extraction and a median precision of 78.1%, common do not forget of 66.8% for opinion phrase determination. From the test, a median precision of 79.8%, a median do not forget of 34% for characteristic extraction, and common precision of 80%, and a median do not forget of 92.7% for opinion phrase determination. The energy of this takes a look at changes to decide an opinion on the capabilities of the area inside the assessment sentence. However, the writer used the handiest adjectives as sentiment phrases to decide the evaluations of the assessment sentences, however, sentiment phrases aren't the handiest adjectives, however additionally consist of adverbs, verbs, and nouns. In addition to this, the sentiment phrases aren't sufficient.

Salama [38] explored sentiment mining for opinionated Amharic texts in the film area wherein opinions are manually gathered from lovers via the way of means of dispensing questionnaires. The researcher Used Lexicon primarily-based method totally, and the proposed version has the subsequent components: reprocessing, sentiment phrases detection, weight venture and propagation, polarity classification, polarity electricity illustration, and sentiment lexica. After the opinions are pre-processed, every period is checked for life inside the sentiment lexica on the sentiment phrases detection component. The detected sentiment phrases are assigned weight and the values of sentiment phrases that might be connected to contextual valence shifters are propagated inside the weight assign and polarity propagation component. Based on the weights of the sentiment values, the opinions are categorized into predefined categories: high quality, poor or neutral. Finally, the polarity electricity of the opinions is rated. The sentiment lexica are constructed manually from exceptional reasserts primarily based totally on the concepts and guidelines. This lexicon is a set of predefined sentiment phrases. They're phrases that specify an opinion closer to an item which includes precise that specific high-quality opinion and awful which specific a poor opinion, that are manually gathered and saved inside the dictionary. After having the lexicon, the overview record is pre-processed and each legitimate period inside the overview is checked whether or not it's far a sentiment phrase or not.

This is achieved via a detection mechanism wherein the entire lexicon is scanned for each period. If the period exists inside the dictionary, then the period is a polarity phrase (high quality or poor). Terms inside the dictionary are tagged inside the lexicon with a pc interpretable cost for high-quality opinion phrases and poor opinion phrases. Based on this procedure, given an overview record, if there are greater high-quality phrases than poor then it's far taken into consideration to be high quality. If there are greater poor than high-quality phrases then it's far taken into consideration to be poor. If there are identical numbers of high quality and poor phrases then it's far neutral.

Final experimental results found in the movie review show on average 0.92 % accuracy, 0.95 % recall, and 0.94% precision.

The creator confirmed the result received in the use of the unmarried widespread cause dictionary and the use of sentiment lexica: the overall cause lexicon and the area precise lexicon. And finally, the result of the test carried out the use of the 2 lexica and thinking about the contextual valence shifter terms.

Mengistu [39] additionally has executed a study on Sentiment evaluation for classifying Amharic opinionated textual content into positive, terrible, or impartial through the use of ML technique in ERTV, Fana broadcasting, and diretube.com domains.

The writer hired three gadgets studying class strategies (Bayes, Multinomial Naïve Bayes, and Support vector machines) the use of n-grams presence, n-grams frequency, and n-grams-TF-IDF capabilities choice techniques. The experiments have performed the use of 576 Amharic opinionated texts gathered from ERTA, Fana Broadcasting, and diretube.com manually. The Experiment suggests that uni-grams –period frequency characteristic election techniques carry out the exceptional for all algorithms (support Vector gadget, Naïve Bayes, and multinomial Naïve Bayes). Based on their relative overall performance of class, aid vector gadget registers with 78.8curacy outperform, Naïve Bayes with 77.6%, and multino Naive Bayes with 74.7%. as proven from the result received SVM carried out higher than NB and MNB algorithm. Tilahun has carried out the take a look at the identity of opinion mining from Amharic weblog the use of rule primarily based totally approach. He used characteristic degree opinion mining and summarization strategies evaluated on 484 critiques manually gathered from the hotel, college, and medical institution domain. Getachew works at identifying opinion mining from Amharic enjoyment textual content the use of gadget studying approaches (Naïve Bayes, Decision Tree, and Maximum Entropy). The test performed the use of 616 Amharic optioned texts. The take a look at received 90.9 %, 83.1%, and 89.6% the use of Naïve, Bayes, Decision Tree, and Maximum entropy algorithms respectively. However, the take a look at did now no longer manage negation, due to the fact the take, a look at makes use of unigram as a characteristic for class. The result best indicates high quality and terrible polarity however it did now no longer encompasses incremental high quality and decremented terribly.

In a summary, one of the study's paintings associated with this take a look at is Selama‘s paintings [38]. He has finished Sentiment evaluation on Amharic opinionated textual content with the use of rules primarily based totally technique on film review. To classify the opinionated textual content, the researcher has constructed a lexicon that includes statistics on approximately Amharic phrase senses. All the phrases which are not inside the listing of the dictionary are disregarded then the elegance could be unclassified consequently there's 29neglect extra some other approaches to resolve this problem.

Summary of a number of the associated works in phrases in their objective/goal, Methods used, Data source, and result found in remark are summarized below in Table 3.

Table 3 Summary of previous works

Author	Objectives/goals	Methods/ Techniques	Problem domain	Result
(Selama,2010)	Design and develop a sentiment mining model for opinion at Amharic documents.	Lexicon/dictionary-based	Movie view	Better results were found by using both basic Lexica and domain lexica with contextual valance shifter terms
(Mengistu, 2013)	Explore the possibility of applying Sentiment analysis and building a model.	Supervised ML (NB,MNB,SVM) and n-grams & n-grams-TF-IDF techniques for feature selection	Entertainment reviews in ERTA& Fana Broadcasting Corporate program	Support Vector machine achieves the best result using unigrams term frequency
(TuluTilahun,2013)	Develop feature level OM & summarization model for the Amharic Language.	Rule-based approach	Amharic blog	Two experiments are conducted.
(Abreham Getachew, 2014)	Appling OM to create a classification model for the Amharic entertainment reviews.	ML Approach (NaïveBayes, Decision Tree & Maximum Entropy)	Entertainment review	By combining the two methods they can improve the results.

2.9 Research gap

The previous study did not take into account the problem of sentiment analysis in tigrigna, which is critical for determining the polarity of a feeling and they didn't think about the irony or the stairwells of expressions. Furthermore, they only evaluated positive and negative polarities, overlooking the importance of inverter words that modify the polarity. In our study, we used rule-base techniques with a dictionary approach to fill the existing information gap in sentiment analysis towards determining TV show popularity.

CHAPTER THREE

METHODS

3.1 Overview

In this research, the focus is to detect the overall polarity of the audience comments on Tigray TV shows and calculate the polarity (positive, negative, or neutral) by adding and decreasing words. Now in this research, we will use Dictionary-based methods and rule-based methods. A wordbook is no other than a list of words that participate in a tier. For example, you can have a dictionary for positive and negative expressions. In our approach we need to process the input text, splitting, stemming, or lemmatizing, and extracting information from it. Tokenization and stemming for the source material written in Tigrigna Language should be done before we compare each lexicon with positive and negative dictionaries. Stemming should be done to disambiguate the root word that should be compared with the positive and the negative dictionaries. The design of the dictionary largely depends on the specific topics you want to perform sentiment analysis on.

3.2 The proposed system architecture

The general architecture of the proposed system of sentiment analysis for Tigray TV broadcast is shown in figure 2. The system contains different components based on the processes required. These components are: pre-processing (Tokenization, Normalization, and Stemming), sentiment sentence analysis, and sentiment scores. If the cumulative result is greater than zero the sentence or document is positive and if the aggregate result is less than zero the text is negative and if it is zero called neutral. The sentiment lexicon is also a portion of the general systems architecture. The input to the system is a corpus of documents in text format. The documents in this corpus are converted to text and pre-processed using a variety of linguistic tools such as tokenization and stemming. The arrangement may also develop a set of lexicons and linguistic resources. The main component of the system is a sentence or document analysis module, which uses language resources to make emotional annotations on pre-processed documents. The explanations may be friendly to whole documents (for document-based sentiment) to individual sentences (for sentence-based sentiment).

The user, who is in this case labeled as the analyst, enters the social media associated with a TV show or the name of the show. We focus on Facebook, the website of Tigray TV, and group discussion conversation data collected manually.

After this, the data collected is then put in a structured format, and stored in a database. The cleaned Facebook would then undergo text pre-processing removing any abbreviations, punctuation marks, and stop words. The tools proceed to the ***Naïve Bayes classifier*** to be marked as positive, negative, or neutral, and an aggregate score is provided.

These annotations are the output of the system and can be presented to the user using various visualization tools.

Rule-based approach

If a comment contains only positive emotions and no negative emotions, it is classified as Positive emotions. If a comment contains only negative emotions and no positive emoticons, it is classified as Negative emotions. If a comment emotions no emoticons Neutral emotions, we apply the sentiment lexicon-related rules. In the following annex, one emotion is considered to be positive.

Samples Tigrigna Positive Sentiment Words listed in annex one

Positive = 2

ስሉጥ =2 ጥዑም =2 ግርም =2 ሓጎስ =2 ፍሽኽ ==2 ፍቅሪ =2

While, the following annex emotions are considered to be negative:

Samples Tigrigna Negative Sentiment Words listed in annex two

ኪሕዶም =1 ኮራዪ=1 መሕዘኒ=1 መከራ =1 ሰረዘ=1 ደንቆሮ=1 ጽሉልቲ =1

Negative = 1

If none of the above rules apply, the comment is classified as Neutral

Neutral = 0

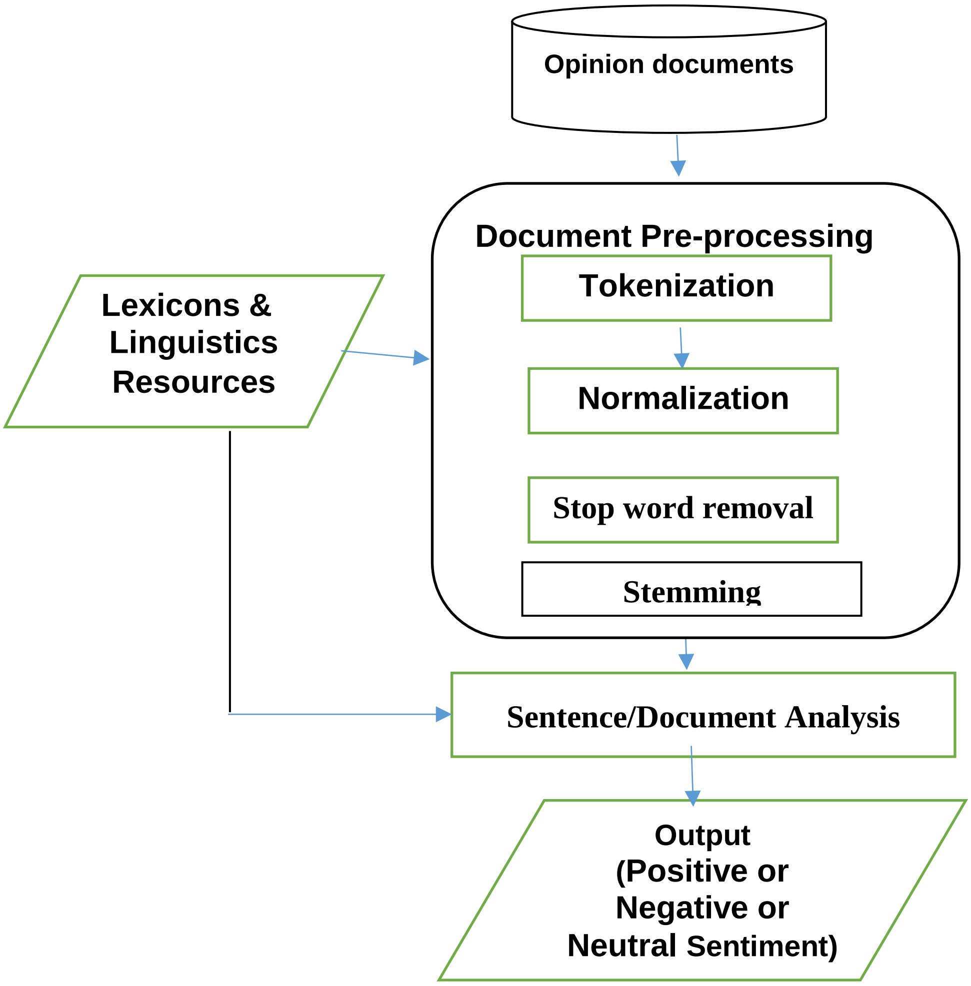

Figure 2 the proposed lexical Architecture

3.3. Defining a structure for the text

Simply define the primer as a list of words. Or define a more detailed structural bearing for every possible attribute of the processed text (word lemma, word form, multiple token Changes...) [40].

Structures to implement the polarity of sentiment will be as follows [40]:

- Each input text to a list of sentences
- Each input sentence to a list of tokens
- Each input token to a tuple of three elements:
    - **A word form** (the exact word that appeared in the manual)
    - **A word lemma** (a general kind of the word) and
    - A list of associated tags

To conduct all experiments we have considered the Tigray TV reviews domain. Those comment and feedback reviews manually collected from Tigray TV viewer is relatively easier and more manageable than any other domains (politics and religion). After collecting comment reviews from the opinion holders coded into a computer and categorize them into labeled classes positive (ኣወንታ), negative (ኣሉታ), and neutral (ዘይሻራዊ).

For this study initiall have collected data to complete the objectives of this study, we have collected data from manually gathered and other resources for analysis to evaluate the performance of the Sentiment Analysis. Before implementing the actual task of the SA, pre-processing of the data has to be done to get more insights about the data. Often input data have been in raw data format is taken and then processed for feature extraction [40].

For detailed analysis, we used pre-processing techniques like removing punctuations and stop words, applying stemming algorithms (reduces the tokens or words to their root form), and tokenization as described below:-

3.4. Pre-processing the text

Once we have decided on the Structures to implement the polarity of the given text, we can pre-process this structured text. By pre-process, we mean in NLP akin to Tokenize, Split into sentences.

A simple definition could be that data pre-processing is a data mining technique to turn the raw data gathered from diverse sources into cleaner information that’s more suitable for work. In other words, it’s a preliminary step that takes all of the available information to organize it, sort it, and merge it [41].

Let’s explain that a little further. Data science techniques try to extract information from chunks of data. These databases can get incredibly massive and usually contain data of all sorts, from comments left on social media to numbers coming from analytic dashboards. That vast amount of information is heterogeneous by nature, which means that they don’t share the same structure that’s if they have a structure, to begin with.

Raw data can have missing or inconsistent values as well as present a lot of redundant information. The most common problems you can find with raw data can be divided into 3 groups:

Missing data: you can also see this as inaccurate data since the information that isn't there creates gaps that might be relevant to the final analysis. Missing data often appears when there's a problem in the collection phase, such as a bug that caused a system's downtime, mistakes in data entry, or issues with biometrics use, among others.

Noisy data: this group encompasses erroneous data and outliers that you can find in the data set but that is just meaningless information. Here you can see noise made of human mistakes, rare exceptions, mislabels, and other issues during data gathering.

Inconsistent data: inconsistencies happen when you keep files with similar data in different formats and files. Duplicates in different formats, mistakes in codes of names, or the absence of data constraints often lead to inconsistent data that introduces deviations that you have to deal with before analysis.

If those issues are not well-handled, the final output would be plagued with faulty insights. That's especially true for more sensitive analysis that can be more affected by small mistakes, like when it's used in new fields where minimal variations in raw data can lead to wrong assumptions [41].

The Need for Data Pre-processing

By now, it is surely clear that data pre-processing is so important. Since mistakes, redundancies, missing values, and inconsistencies all compromise the integrity of the set, we need to fix all those issues for a more accurate outcome [41].

Thus, before using that data for the intended purpose, there is a need to organize and "clean" it as much as possible. There are several ways to do so, depending on what kind of problem the researcher wants to tackle**.**

3.4.1. Tokenization

Tokenization is performed which attempts to identify words in the document corpus. The common method of representing the document text is using the bag of words approach where the word from the document corresponds to a feature and the documents are considered as a feature vector [4]. This demonstrates that the words can as it were separate the given archives within the categorization handle. So, the punctuation marks and numbers are impertinent elements of the manual documents. In the present study, the punctuation marks and digits are removed and replaced with space.

The words are considered as important for the given reports, and they are isolated by space. Algorithm 1 illustrates the tasks of removing punctuation marks towards simplifying the tokenization process of the present study. First, the content of the file is read line by line. Second, check whether the word within the list contains punctuation marks of the Tigrigna language; if punctuation marks exist within the word, replace them with space. This step continues until the end of the page reaches. After removing punctuation marks, the content of the file is passed for the tokenization process.

Algorithm 1 Removing punctuation marks

```
Open the file for processing
Do
        Read the content of the queue line by line
        Assign the content to string
        For word in string split by space
                If word contains punctuation marks
                        Replace punctuation marks with space
        end for
While end file
```

The over calculation tokenizes the content reports as takes after: to begin with, the substance of the record is read line by line. Third, check whether the word within the list contains punctuation marks of the Tigrigna language; if punctuation marks exist within the word, replace them with space. This step continues until the end of the line reaches. Similarly, digits are also removed using the python built-in function called “sub“ which takes a digits symbol “d+” as an argument and removes digits from the whole list of words.

Tigrigna language has compounds composed completely of different word formations. For the most part, space and hyphens are utilized to partition words. When the hyphen is used, the two words are treated as one word. However, when words are separated by space, their meaning is different (Leslau, 1998). For example, “ቤትትምህርቲ”, “ስነስርዓት”, “መራሕቲስድራ”, and “ኣብያተፅሕፈት” are compound words separated by space. However, the words “ቤት”, “ስነ”, “መራሕቲ”, and “ኣብያተ” do not have meaning when they are used separately. This creates a problem in the text categorization process. Hence, this study prepares the compound words list file and combines them using algorithm 2 as follows during the tokenization process.

Algorithm 2 Tokenization Algorithm of Combining Tigrigna compound words

```
Open the file for processing
Do
        Assign the content to string
        for word in string
        split by space
        if word in compound word list
                Combine first word with next word
        end for
While end file
```

Tokenization and combining multiple words result in a group of words of the Tigrigna language. They need further processing to remove stop words and apply stemming for grouping similar words together.

3.4.2 Stemming

Tigrigna language has words with the same root and is written in different forms. So stemming is used to stem words of different forms to their root words. In this study, the researcher stems the suffix and prefix of the term by identifying them in the given documents. The suffix is written at the end of the root word while the prefix is placed before the root word. Together the prefix and suffix modify the meaning of the word.

The steaming process used in this study is based on the affix removal algorithm developed for Tigrigna language sentences. Generally, the process takes the word from the file as an input. The algorithm has both suffix dumping and prefix dumping. The prefix removal algorithm is shown in algorithm 3 below, it checks whether the word starts with prefix lists or not. If the word starts with the prefix-list, it replaces the prefix of the word with space. If the word does not contain a prefix, it then continues checking the end of the word for the suffix.

Algorithm 3 Prefix striping

```
Read prefix list file
Open the file for processing
Do
        Read  the file line by line
        Assign the content to string
        for word in word-list do
                If word starts with prefix then
                        If length of word is greater than twothen
                                Remove prefix from the word
                        end if
                end if
        end for
While end of file
```

Similar to the prefix striping, the suffix stripping algorithm is shown in algorithm 4 also strips the suffix if the word ends with one of the suffixes in the suffix list.

Algorithm 4 Suffix striping

```
Read suffix list file
Open the file for processing
Do
        Read the file line by line
        Assign the content to string
        For word in string split by space
                If word ends with suffix then
                        If length of word is greater than twothen
                                Remove suffix from the word
                        end if
                end if
        end for
while end of file
```

As shown in algorithms 3 and 4, the affix removal algorithm checks the existence of the prefix and suffix and then removes them from the word. The prefix striping algorithm removes the prefix which is placed before the root word. For example “ስለዝወሰደ” (sleziwesede) contains “ስለዝ” (slez-) prefix. As a result, it is stemmed into “ወሰደ” (wesede). Similarly, the suffix stripping algorithm 4 removes the suffix which is written at the end of the root word. For instance, “መጓዓዝቲ” (megaez-ti) has “ቲ” (-ti) suffix at the end. So it is stemmed into “መጓዓዝ” (megaez). In the Tigrigna language, prefix and suffix also exist in a single word. For example, the word “መመንደርና” (memenderna) has a prefix “መ” me - and suffix “ ና” (-na). First, the– prefix “ መ” me - is removed using algorithm 3. As a result, the word “መንደርና” (menderna) is stemmed into “መንደርና” (menderna) but it has a suffix “ና” (- na). Then the suffix “ና” (-na) is removed from “መመደርናና” (memenderna) and it is stemmed into “መንደርና” (menderna) using algorithm 4. As a result, the stemming helps to define words in the same context with the same term and accordingly reduce their dimensionality.

3.4.3. Stop words detection

Tigrigna language has very frequently appearing terms which are called stop words These terms do not distinguish one document from other documents. The stop words are identified after stemming because some stop words exist in the Tigrigna text documents in their affix form. So stemming conflates these terms into their base words. As shown in Table 3 below, the Tigrigna stop words do not discriminate document meaning but they provide an arrangement in the language. These are propositions like "ንስኺ" (You), pronouns such as "እቲ" (the), and verbs to be like "እዮም" (are), etc.

Table 4 Example stop words of Tigrigna documents

Stop Word	Meaning	Stop Word	Meaning
ኣብ	At	እዮም	Are
ናብ	To	እቲ	The
ካብ	From	ንስኺ	You
ኣነ	I		

In the present study, stop words are recognized manually by consulting books and dictionaries of the Tigrigna language. The consulted books and dictionaries help to identify prepositions, conjunction, articles, pronoun, and auxiliary verbs of the Tigrigna language. After identifying the stop words in the Tigrigna documents, algorithm 5 detects and removes the stop words from the document corpus.

Algorithm 5 Stop word removal

```
Read stop word list file
Open the file for processing
Do
Read the file line by line
        Assign the content to string
        For word in string split by space
                If word in stop word list then
                        Remove word from the index term
                end if
        end for
While end file
```

As shown in algorithm 5 above, the framework peruses the list of halt words from the halt word list record and iteratively expels them from the collection of file terms

3.5. Normalization

Normalization indicates the consistency of characters. Since the Tigrigna writing system has homophone characters which means characters with the same sound have different symbols. For example, the character ‘ሰ’(Se) and ‘ሠ’(Se) are used interchangeably as “ሰላም” (selam) and “ሠ” (selam) to mean “Peace” These different symbols must be considered as similar because they do not affect meaning. Inconsistency of words may be because of an unnecessary increase in the number of words representation that causes large data size processing and considers the word like a different term. In this research, such a kind of inconsistency in writing words will be handled by replacing characters of the same sound to their common standard with form.

Algorithm 6 Normalization Algorithm

Open the file for normalization

While not end of corpus file do

For each character in the corpus

If the character is ኀ(HA) or any order then

Changed to ሀ(HA)

Else if it is ሠ (SA) or any order of it then

Changed to ሰ(SA)

Else if ……

End if

End for

3. End while

3.6. Sentence analysis for polarity detection

Due to the subjective nature of Sentiment Analysis, it is of no surprise that a key indicator of sentiment polarity is sentiment words. Good (ፅቡቅ), great (ዝበለፀ), and brilliant (ንቁሕ) can be considered positive words while bad (ሕማቅ), worse (መፅልኢይ) and disastrous (መሕዘኒ) can express negative attitudes. Therefore, sentiment lexicons, which gather sentiment words, are used extensively by the research community. Sentiment lexicons can be organized in three types, attending to which information is contained in them [13]: (i) those that contain only sentiment words (a list of words), (ii) the ones that are formed by both sentiment words and polarity orientations (a list of words with only positive and negative annotations), and (iii) the lexicons that offer sentiment words with orientation and intensity [13] (a list of words with scalar numerical values).

The most popular approach that makes use of sentiment lexicons is keyword matching, also called keyword spotting. This technique consists in detecting the presence of certain sentiment bearer words, thus obtaining the sentiment estimation as an aggregate of the associated sentiment values. Although this method is certainly simple and computationally cheap, it is also limited, as it happens in the case of domain adaptation.

In general, the steps to decide the polarity of the given text are the following:

- Stemming is done on every word of the input word.
- Identify each disambiguated lexicon for each input word.
- Compare positive, negative, and neutral there is a positive, negative, and incremental or decrement in the dictionary.
- Calculate the scores, output the sentiment polarity as positive, negative, or neutral

The sentiment is a state-of-the-art, lexicon-based classifier that exploits a sentiment lexicon built by combining entries from different linguistic resources. In the lexicon, each negative word receives a sentiment score ranging from 1 to 2, which represents its prior polarity (i.e., the polarity of the term out of its contextual use). Similarly, positive words are associated with a score between 4 and 5, while neutral words receive scores equal to 3. Positive and negative emotions are also included in the dictionary. Based on the assumption that a sentence can convey mixed sentiment, Sentiment outputs both positive and negative sentiment scores for any input text written in Tigrigna. It determines the overall positive and negative scores of a text by considering the maximum among all the sentence scores, based on the prior polarity of their terms. Intensifiers, i.e., exclamation marks or verbs such as ‘really’, are treated as booster words and increase the word sentiment scores. Negations are also treated and determine the inversion of the polarity score for a given word. Therefore, the overall positive p and negative n sentiment scores issued by the tool range from 1 (absence of positive/negative sentiment) to 5 (extremely positive/negative). Based on their algebraic sum, Sentiment can also report the overall trinary score, i.e. the overall positive (score = 2), negative (score = 1) and neutral (score = 0). Examples are provided in TABLE 1. The rationale for classification reported in the second column of the table is obtained by enabling the ‘explain’ option in Sentiment.

Table 5 examples of sentiment detection

Input Text	Final Sentiment Score			Overall Score
	Negative Score	Positive Score	Neutral Score	
ካብ ዝሓለኹዎን ላዕሊ እዩ	0	5	0	Positive (overall result = 2)
ትግራይ ተለቪዥን ጸማማት ጥራሕ እዩ	1	0	0	Negative (overall result = 1)
መብዛሕቱ አብ ፖሎቲካ ጉዳይ ዘድሀበ ስለዝነበረ ናይ ኢኮኖሚይ ማሕበረሰብ ጉዳያት ብዝበለፀ ትኹረት ይንእስ ከምዝነበረ::	0	0	3	Neutral (overall result = 0)

3.7. Performance measurement

We used the following performance metrics to evaluate experimental results registered by the proposed sentiment analysis in this study. A confusion matrix is used as a form of visioning the performance of a classifier. It is displayed in a table format in which the columns represent the actual values (true and false) and the rows represent the predicted values (positive, negative, and neutral). It can freely be generalized to formulate class classifiers. The table reports the results of a classifier in terms of the number of true positive (TP), false positive (FP), false negative (FN), and true negative (TN).

A confusion matrix is a matrix of size 2×2 for the binary organization with actual values on one axis and predicted on another as shown in table 4 below.

Table 6 Confusion matrix

Labelled	Truth	
	TP	FP
	FN	TN

Based on the confusion matrix, it is possible to calculate the accuracy, precision, recall, and f-score of sentiment analysis results.

Accuracy is well-defined as the overall percentage of positive and negative predictions that are correctly done [16].

$$Accuracey = \frac{\text{TP+TN}}{TP+TN+FP+FN} \qquad (3.1)$$

Accuracy performance metrics can be decisive when dealing with imbalanced data. In this blog, we will learn about the Confusion matrix and its associated terms, which look confusing but are trivial. The confusion matrix, precision, recall, and F1 score give better intuition of prediction results as compared to accuracy.

True Positive (TP) is the model that correctly predicts the positive class (prediction and actual both are positive). True Negative (TN) is on the other hand model that correctly predicts the negative class (prediction and actual both are negative). False Positive (FP) is the model that gives the wrong prediction of the negative class (predicted-positive, actual-negative). FP is also called a TYPE I error. Finally, False Negative (FN) — the model wrongly predicts the positive class (predicted-negative, actual-positive). FN is also called a TYPE II error.

With the help of TP, TN, FN, and FP, other performance metrics that can be calculated include precision, recall, and F-score

Precision calculates out of all the positive predicted, what percentage is truly positive using equation 3.2.

$$Precision = \frac{TP}{TP + FP} \qquad \text{.......(3.2)}$$

The precision value lies between 0 and 1.

Recall measures out of the total positive, what percentage are predicted positively using equation 3.3. It is the same as TPR (true positive rate).

$$Recall = \frac{TP}{TP + FN} \qquad \text{.......(3.3)}$$

When comparing different models or systems, it will be difficult to decide which is better (high precision and low recall or vice-versa). Therefore, there should be a metric that combines both of these. One such metric is the F1 score [16].

F1 Score is the harmonic mean of precision and recall as shown in equation 3.4. It takes both false positives and false negatives into account. Therefore, it performs well on an imbalanced dataset.

$$F1\ score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

…….(3.4)

F1 score gives the same weightage to recall and precision.

CHAPTER FOUR

DATA PREPARATION

Here in this study, we are using the rule-base approaches which are rule-based, and the dictionary-based approaches to identify the given comment as either positive or negative, or neutral.

4.1. Data collection

The information is gathered from Tigray TV broadcasting watchers which are given remarks from Tigray TV Facebook, web site, and during the focus group discussion conversation and by distributing an open-ended questionnaire. The target groups from which comments are collected include university students of Addis Ababa 6 kilo campus and youth from Addis Ababa Gofa Sefer and Mebrat Hail condominium area. They are good for our study or research because youth spent much time in the Tigrigna program. Especially, for the Tigrigna speakers, an open-ended questionnaire distributed among these people (viewers/non-viewers of the program), and focusing group discussion have been done here in this area. We collected comments from the Tigray website page and Facebook by using video screenshots. From all information assortment strategies, we have collected a sample size of 1633 comments for this study (see table 7).

Table 7 Summary of the data collected

Source	Positive	Negative	Neutral	Total
Addis Ababa 6 kilo campus	116	74	18	208
Addis Ababa Gofa sefer	110	10	0	120
Gofa Mebrat hail condominium area	140	25	9	174
Social Media	564	544	23	1131

After the data is collected, pre-processing tasks were applied to construct the final data set (data that are used as input for experimentation). Data preparation tasks are usually performed multiple times depending on the quality and size of the initial dataset. Tasks such as normalization, tokenization, and stemming of the data were performed to come up with the final appropriate dataset for the selected algorithms.

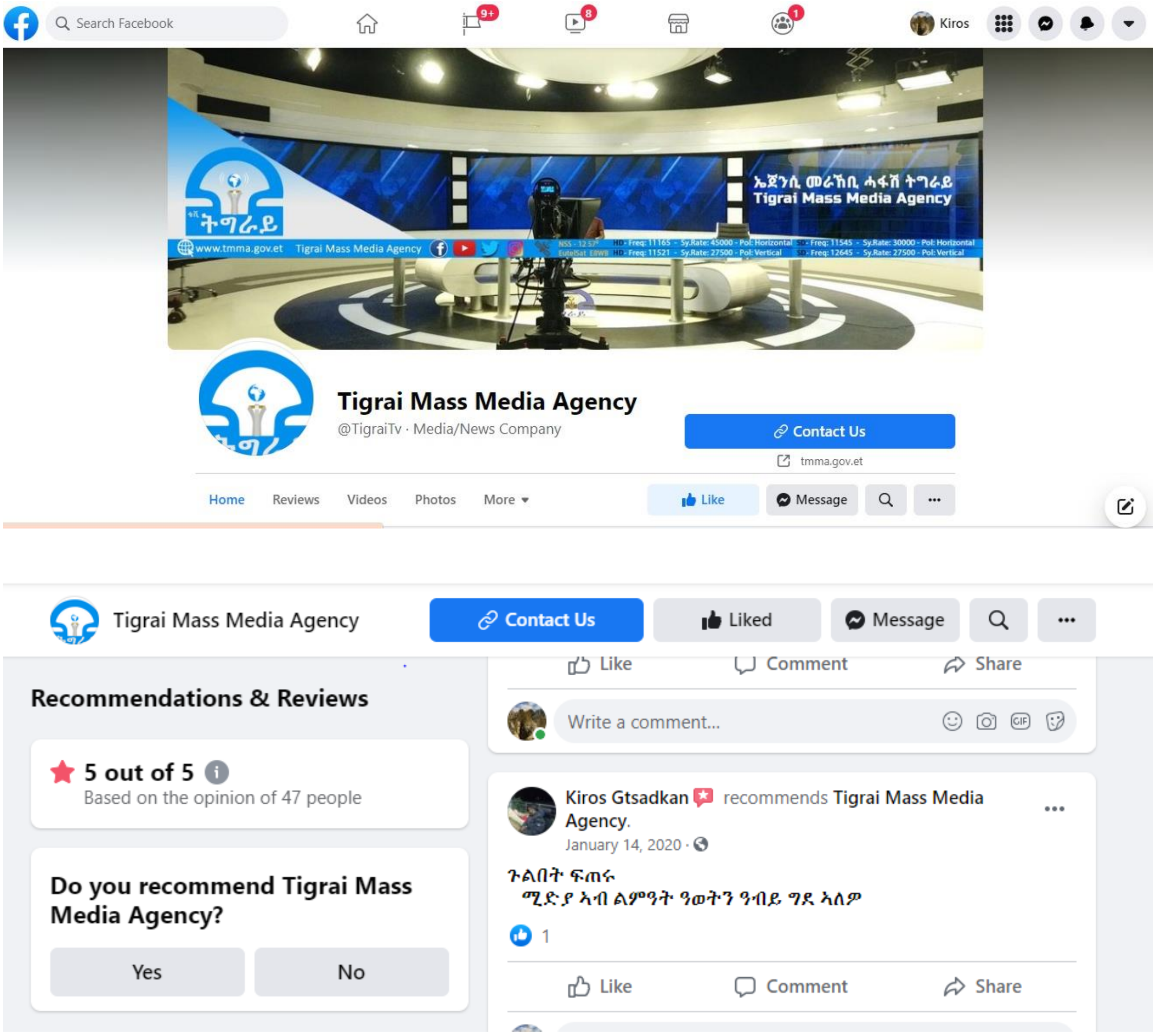

Figure 3 Sample face book comments by screen shot.

4.2. Normalization

There are many abbreviations (short) words in the Tigrigna writing system. These short words can be single or compound short forms of Tigrigna language words abbreviated by the slash (/) or period (.). To make the corpus usable for stemming, short words should be expanded to their expanded form. And Tigrigna language normalizer to extend the short form of Tigrigna language, separated by periods or slashes, such as ላ.ማይጨው or ላ/ማይጨውላ ዕላይማይጨው (short form of La elay Machew from the study area of the central area) represented by the Unicode value taken from the Ethiopian Unicode table.

There are two types of compound Tigrinya short-form words: there is a space or no space between the two words but they need a space after they are expanded. For example, ቤትፅ/ት =

ቤትፅሕፈት, which means "office" is a compound short word with a space between the two words ቤት and ፅ/ት. However, the compound short word ዓ.ም have no space but after expanding it needs a space between the two words and expanded as ”ኣመተምሕረት‟ which means ”Year.‟ Therefore, by considering such conditions, we have implemented a Tigrigna language short-form word expander. Here we present python code for normalization abbreviation of words used in Tigrigna writing.

```
In [11]: #Normalization
         import fileinput
         normalFile = open("C:/Users/Kiros/Desktop/Sentiment Code/Untitled Folder 1/ResultNormalization.txt", "a+", encoding='utf-8');
         fields = {"ፕ/ር ": "ፕሮፌሰር", "ዶ/ር":"ዶክተር","ጠ/ሚኒስተሩ":"ጠቅላይ ሚኒስተር","ጠ/ሚኒስትሩ":"ጠቅላይ ሚኒስትሩ","ወ/ሮ":"ወይዘሮ",
                   "ዓ.ም":"ዓመተ ምህረት","ጽ/ቤት":"ጽህፈት ቤት"}
         f = open('C:/Users/Kiros/Desktop/Sentiment Code/Untitled Folder 1/Reviews.csv', encoding='utf-8');
         for line in f:
             line = line.rstrip()
             if not line:
                 continue
             for f_key, f_value in fields.items():
                 if f_key in line:
                     line = line.replace(f_key, f_value)
             normalFile.write(line + "\n");
         normalFile.close();

In [19]: train = pd.read_csv('Result.txt')
         print(train.head(10))
```

Figure 4 Python code for Tigrigna words Normalization

Social media texts often don't conform to rules of spelling, grammar, or punctuation. Among its challenges are:

- Unstructured data: Most common challenges we face in NLP are around unstructured data. Data generated from Facebook conversations, and comments are highly unstructured. It’s a huge challenge to process that data and get useful information out of it.
- Abbreviations: no consistency is kept in abbreviating ‘ዓመተምሕረት’, meaning ‘AD’, can be abbreviated as ዓም, ዓ.ም, ዓ.ም., ዓ/ም, ዓ-ም. All the aforementioned problems pose challenges since the same word is treated in different forms in the process of feature preparation for the text classifier. Classes learned by the competitive layer are referred to as subclasses and the classes of the linear layer are called target classes.
- Misspelling: wouls (would), rediculous (ridiculous)

- Omitted Punctuation: እዩ ('ዩ)
- Slang: that was well mint (that was well good)
- Wordplay: ቀውውውው that was soooooo great (that was so great)
- Disguised Vulgarities: sh1t, f**k
- Informal Transliteration: This concerns only multilingual text. Variations in transliteration occur due to long vowels, borrowed words, accents/dialects, and double consonants.

There is some progressive work that has been done so far on Tigrigna NLP tasks with promising results including part of speech tagging, morphological analyzer, named entity recognition, base phrase chunking, and text classification. Various techniques have been widely employed for each task to enhance the accuracy and handling of linguistic exceptions. However, there have not been ready-made pre-components and well-organized datasets. Besides these limitations, there has not been any undergoing research on event extraction from unstructured Tigrigna text due to difficulties in the syntactic and semantic status of the class of functional verbs. The other challenge is identifying event arguments. In our case temporal event arguments have been considered. However, it has a challenge in Tigrigna texts. Tigrigna texts have been represented in various forms such as; sequence of words, Arabic, and Geez'e script numerals. As such it needs extra normalization and a syntactic analyzing scheme to tackle the temporal argument. Semitic languages like Arabic and Tigrigna have much more complex morphology than English. The morphological variation limits the research progress on Natural language processing, in general. However, in our case, we mainly concentrate on extracting events and their arguments with the advantage of hand-crafted rules and the rule-based approach of the sentiment analyzer.

4.3 Tokenization

Tokenization is the process of splitting text into words (tags) to obtain contextual words for disambiguation purposes [8]. It is also defined as the process of splitting a string into a list of fragments or tags. This means that the given string is split into a list of tokens to make natural language processing usable [40]. In this study, before processing the data, all punctuation marks (see Figure 5) in the Tigrigna language were removed from the text. Words are treated as tokens. Punctuation marks are converted to spaces. And spaces are used as word separators.

Figure 5 below shows python code for removing punctuations and tokenizing Tigrigna sentences into words.

Let's first save the phrase as a variable named "ResultTokenization.txt":

```
In [31]: import fileinput
normalFile = open("C:/Users/Kiros/Desktop/Sentiment Code/Untitled Folder 1/punctuations.txt", "a+", encoding='utf-8');
punctuations = '\!\@\#\$\%\^\«\»\&\*\(\)\…\[\]\{\}\;\"\"\›\'\'\"\'\:\,\.\‹\/\<\>\?\\\\|\`\´\~\-\=\+\፡\።\፤\;\፦\፥\፧\፨\፠\፣\-'
filename = open("C:/Users/Kiros/Desktop/Sentiment Code/Untitled Folder 1/ResultNormalization.txt", "r", encoding='utf-8');
file = filename
no_punct = ""
lst = []
for line in file:
    line = line.strip()
    for char in punctuations:
        line = line.replace(char,'')
    #lst.append(line)
    normalizedData.write(line + "\n");

filename.close()
```

```
In [37]: import re
import fileinput
normalFile = open("C:/Users/Kiros/Desktop/Sentiment Code/Untitled Folder 1/Foreign_remove.txt", "a+", encoding='utf-8');
#punctuations = '[A-Za-z0-9]'
filename = open("C:/Users/Kiros/Desktop/Sentiment Code/Untitled Folder 1/ResultNormalization.txt", "r", encoding='utf-8');
file = filename
no_punct = ""
for line in file:
    line = line.strip()
    #for char in punctuations:
    clearData = re.sub('[A-Za-z0-9]','',line)
     #line = line.replace(char,'')
    #lst.append(line)
    normalFile.write(clearData + "\n");
filename.close()
```

Figure 5 Punctuation and Foreign languages removing

4.4. Stemming

As described in [33], the Tigrigna language uses prefixes, suffixes, and infixes to create inflections and derivative word forms. Stemming is a technology that reduces words to their root forms by deleting derived affixes and inflectional affixes (prefixes, infixes, and suffixes), and will lead to the WSD system and information retrieval tasks. A word's stem is its most elementary form which may or may not have a semantic interpretation.

Figure – below shows python code designed for stemming and conflating words in Tigrigna writing.

```
def stem(word):
s=codecs.open("suffix.txt",'r', encoding = 'utf-8' )
suffix= ("suffix.txt",'r')
suf = s.read()
for suffix in suf.split():
```

```
if len(word)>=1:
if word.endswith(suffix):
word=word [:-len(suffix)]
if word.endswith('ንክ'):
word=word.replace('ክ','')
word=word.replace('ን','ና')
          if len(word)>=1:
return word
else :
 Word=word.replace ('ንክ','')
return word
```

Figure 6 Sample Python code for suffix removal

This stemmer utilized five-stage rules to eliminate Tigrigna word variants. The initial step takes the word to be stemmed as information and eliminates twofold letter reduplication. The subsequent advance eliminates the prefix-addition pair. This progression takes the yield of the initial step as info and checks if the words contains match with any prefix-addition pair. On the off chance that the word contains, a match and the excess string have a length more prominent than the base length, then, at that point, the prefix and postfix are taken out from a word. The third step eliminates prefixes and takes the yield of prefix-postfix stripping. In eliminating a prefix, checking for the match in the prefix rundown and tallying the length of the excess string is finished. The fourth step eliminates postfixes by tolerating the yield from the past stem and checks if the word contains any match from the rundown of additions. On the off chance that the word has a match and the excess string is more noteworthy than the least length, the addition is eliminated from the word. In the last advance, the calculation stems reduplication of the single letter. This calculation has a recording decide that is applied get-togethers step is applied for checking some spelling special cases and making rearrangement.

4.5. Stop words

Stop words are the most common words found in any natural language which carry very little or no significant semantic context in a sentence [42] and so that the process is not over-influenced by very frequent words [27].

Figure – below present's python code written for removing stop words in Tigrigna writing using a stop word list.

```
In [3]: stopwords=['እዩ','ነቲ','ናበይ','እንትኾውን','እና','ላዕሊ','ወዘተ','ካሊእ','ግን','ነገር','ወይስ','እዮም','እዚ','ካብዚኦም','እዝኦም','ንሳቶም','
```

```
In [4]: stopwords=list(stopwords)
        print(stopwords)
```

```
['እዩ', 'ነቲ', 'ናበይ', 'እንትኾውን', 'እና', 'ላዕሊ', 'ወዘተ', 'ካሊእ', 'ግን', 'ነገር', 'ወይስ', 'እዮም', 'እዚ', 'ካብዚኦም', 'እዝኦም', 'ንሳ
ቶም', 'ግደፎ', 'ብዝኾነ', 'ብታሕቲ', 'ማለት', 'ካልኦት', 'ሓደ', 'ምኽንያቱ', 'ይኾውን', 'በዚይ', 'ይኹን', 'ክልተ', 'ብኻሊዕ', 'ብዚ', 'ሎ
ሚ ዘበን', 'ነይሩ', 'ይኾውን', 'ምኽንያቱ', 'ነዚኦም', 'ተወሳኺ', 'ከምዝነበረ', 'ተሓሒዙ', 'ዘይብሎም', 'እንደገና', 'እንተ', 'ግና', 'ንሳቶም',
'ነይሮም', 'ጻሓር', 'ናይዚ', 'ተወሳኺ', 'ብዚ', 'ይከኣል']
```

Figure 7 Stop words python code

4.6. Data set for experiment

The data sets for experimental purposes are taken from four sources. The data set is taken from the Facebook page and general TV viewer review site. The data set taken contains 1380 positive and 630 negative reviews. The reviews were collected from a variety of different products, including questionnaires, and TVs review sources.

In [11]:
```python
import pandas as pd
df = pd.read_csv('Book8.csv')
```

In [12]:
```python
review_df = df[[ 'ውጽኢት','ሪኢቶ','መጠቃለሊ']]

print(review_df.shape)
review_df.head(5)
```

```
(1637, 3)
```

Out[12]:

	ውጽኢት	ሪኢቶ	መጠቃለሊ
0	2	ስሉጥ	አዎንታዊ
1	2	ስልጥቲ	አዎንታዊ
2	2	ብስሉጥ	አዎንታዊ
3	2	ስሉጣት	አዎንታዊ
4	2	ሰሊጠ	አዎንታዊ

In [13]:
```python
df.columns
```

Out[14]:

	ውጽኢት	ሪኢቶ	መጠቃለሊ
735	1	ኪሕጾም	አሉታዊ
736	1	ኪሕደን	አሉታዊ
737	1	ካሕዳም	አሉታዊ
738	1	ክሕዳቶም	አሉታዊ
739	1	ክሕዳተን	አሉታዊ

In [15]:
```python
review_df["መጠቃለሊ"].value_counts()
```

Out[15]:
```
አዎንታዊ     930
አሉታዊ      653
ዘይሻራዊ      50
Name: መጠቃለሊ, dtype: int64
```

In [20]:
```python
df = pd.read_csv("Book8.csv")
df.tail(10)
sns.countplot(df["ውጽኢት"])
df["መጠቃለሊ"].value_counts()
```

Out[20]:
```
አዎንታዊ     930
አሉታዊ      653
ዘይሻራዊ      50
Name: መጠቃለሊ, dtype: int64
```

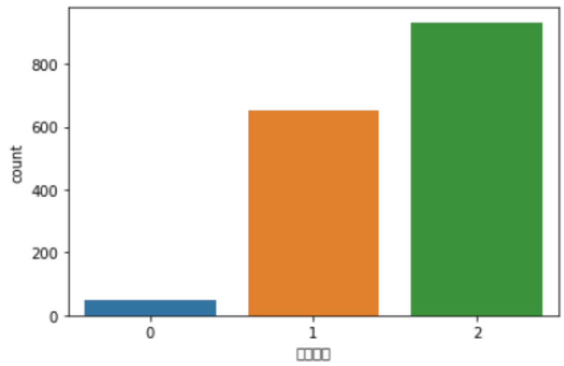

Figure 8 Data set for experiment

CHAPTER FIVE

RESULTS WITH EXPERIMENTS

The experiments in this study are conducted based on four target areas which are Addis Ababa University (Addis Ababa 6 kilo campus), Addis Ababa Gofa Sefer, Mebrat hail condominium area, and the Facebook page of Tigray TV. During this study, the data were collected through questionnaires, interviews, focus groups, and collected comments on the Facebook page by video screenshots. The total number of the dataset (comments) used in this study for the experiment is 1633 (One thousand six hundred thirty-three).

5.1. Experimental setup

The sentiment analysis system is developed and tested on a Toshiba laptop computer with windows 10 ultimate operating system, Intel Core i7 with 2.6 GHz processor speed, 8.0GB RAM, and 1 BT GB hard disk capacity. Other software components used to develop and test our system are Python, NLTK, Notepad++, and PyMy.SQL.

Python is a simple yet powerful programming language with excellent functionality for processing linguistic data [12]. It is highly readable, allows data and methods to encapsulate, and contains an extensive library including components for graphical programming, numerical programming, and web connectivity. For this work, version 3.5 of the python 32-bits is used. It is used to develop the prototype of a sentiment analysis system.

NLTK is a natural language processing toolkit that can be used to build NLP programs in python. It is an open-source toolkit that contains the open-source module, linguistic data, and documentation for research and development in the natural language processing field [12]. It provides a basic class for representing data relevant to NLP, and a standard interface for performing tasks such as text classification, part-of-speech tagging, and syntactic parsing. In this work, NLTK is used for sentence and word tokenization to split the input sentiment sentences into lists of words.

Notepad++ is a free source code editor and the replacement for Notepad which supports several languages. In this work, we used the Notepad++ version of 7.5.1 with 32-bits which is used to build and edit the lexicons.

PyMySQL is the package that contains a pure-Python MySQL client library and is used to connect and access MySQL databases.

5.1.1 Number of experiments conducted

In this research work, different datasets were collected for language identification, to build sentiment lexicons, and to evaluate the performances of the sentiment analysis system.

Language identification: Tigrigna languages were detected in word level using lexicon based approaches. To build the lexicon Tigrigna wordlists were used. Tigrigna wordlists were collected from manual pre-processing activities were made for Tigrigna wordlists in order to avoid unimportant words and phrases. After the pre-processing activity is completed, a total of 4,205 Tigrigna words were used for Tigrigna language.

Sentiment lexicon: the sentiment lexicon is built manually by collecting sentiment words from various resources. Tigrigna sentiment words and the sentiment words are validated by linguistic expertise. Some lists of the sentiment words are provided on the Appendix portion.

Table 9 belowshows the total number of sentiment words prepared for use to build the sentiment lexicons.

Table 8 Sentiment Words in Sentiment Lexicons

Sentiment words			
Language	Positive	Negative	Neutral
Tigrigna	1380	630	36

Finally, for performance evaluation of the sentiment analysis system, 1633 sentiment sentences have been manually collected from Facebook and websites in the domain of movies, music, and politics. These domains were collected because of the absence of any prepared available written sentiment sentences in the Tigrigna language. At the time of collecting the data, the sentiments on the posted topics are selected randomly. After collecting the sentiment data, we made manual sentiment classification with the help of linguistic expertise. The manual sentiment classification helps to compare the results generated by the system. The manual sentiment classification is used to assign the polarity of the sentiment sentences and helps to compare the results of the system after experimenting.

In this research work, sentiment sentences are categorized into three basic classes based on their semantic orientations: positive, negative, and neutral sentiment classes. The positive and negative sentiment classes are classified further into positive, strongly positive, weakly positive, negative, strong negative, and weak negative sentiment classes. The sentiment sentences can contain Tigrigna and mixed language of words since different opinion holders can express their feelings for the same topic from different perspectives; it can use different languages or can combine languages for their expressions. For example, consider this sentence "nice movie film so ቀፀለሉ!!!" or "too much happy አዝዩፀቡቅ music" In this example, the first sentiment sentence contains Tigrigna and English words in the movie topic, while the second sentiment sentence contains English in the music topic. In this study, such kinds of sentences are considered, and the sentiment is determined based on the sentiment words in the sentence. However, there are also pure Tigrigna sentiment sentences on the topic from opinion holders. Finally, the collected sentiment sentences were categorized into predefined categories: positive, strongly positive, weakly positive, negative, strong negative, weak negative, and neutral sentiment classes.

5.2 Results and Discussion

The experiment is done to measure the overall performance of the developed sentiment analysis system. In this research work, a total of 1633 sentiment sentences were used to test the accuracy of the system. Below we present the test results achieved in this study.

5.2.1 Effect of Stemming on the Tigrigna Sentiment Analysis

In this section, we investigate the effect of stemming on the performance of Tigrigna sentiment analysis. To show the effect of stemming, we conducted sentiment analysis using stemmed and unstemmed datasets without changing the polarity of unstemmed datasets. To compare results of Naive Bayes with stemming of this paper.

Table 9 show the effect with and without stemming.

Performance measure	20% test and 80% training data	
	with stemming	without stemming
Precision	77.235	81.26
Recall	71.51	77.89
F score	69.76	77.89
Accuracy	74.93	81.36

Table 9 and 10 show that the accuracy of sentiment analysis with stemming is lower than without stemming by an average of 0.53%, 0.43% and 0.9% for Uni-gram, Bi-gram and Tri-gram Naïve Bayes models respectively.

Stemming gives the same F score for uni-gram and Bi-gram implementation of NB as without stemming. However, the F score of Tri-gram implementation of NB reduces from 77.89% to 69.76% and from 89.0% to 88.0% respectively.

Table 10 effect with and without Stemming

Performance Measure	20% test data and 80% training data					
	Without Stemming			With Stemming		
	Uni-gram	Bi-gram	'Tri-	Uni-gram	Bi-gram	Tri-gram
Precision	87.0	90.0	90.0	86.0	89.0	89.0
Recall	84.0	89.0	89.0	84.0	89.0	88.0
F Score	85.0	89.0	89.0	85.0	89.0	88.0
Accuracy	86.03	89.1	89.1	85.5	89.27	88.8

Tables 9 and 10 show that the accuracy of sentiment analysis with stemming is lower than without stemming by an average of 0.53%, 0.43%, and 0.9% for Uni-gram, Bi-gram, and Tri-gram Naïve Bayes models respectively. Stemming gives the same F score for uni-gram and Bi-gram implementation of NB as without stemming. However, the F score for Tri-gram implementation of NB reduces from 78.89% to 70.76% and from 90.0% to 89.0% respectively.

Stemming reduces inflected words to their stem or root form resulting in a reduced number of features or vocabulary size. Sometimes the stem of inflected words need not be identical to the morphological root of the word. For instance, the word "ዕውት" can be inflected into "ዕውተኛ", "ዕውተኛታት", "ብዓወት", "ዓወት", "ዝተዓወተ", "ዝተዓወቱ" etc. In this example, all those words have the same root that is "ውት". This shows less dependence between inflected words and the root word "ውት". Merging inflected forms of words to a common stem form results in weak dependence between features. However, the weak dependence of features in a sentence due to stemming makes the network not capture the dependency between these features. In conclusion, the experiments conducted show that stemming from the preprocessing step hurts sentiment analysis.

5.2.2 Evaluation Results

The role of this activity is to describe the evaluation metrics of the designed system and followed by its test results. We have used precision, recall, and F-measure evaluation metric to measure the effectiveness of the selected approach. In this study, the experiment is done based on the sentiment polarity class and evaluates each evaluation metric corresponding to each sentiment polarity class. The sentiment polarity classes are classified into seven categories:

positive, weakly positive, strongly positive, negative, weak negative, strong negative, and neutral classes. At last, the precision, recall, and F-measure were calculated in the experiment for each polarity class.

In this experiment, we considered the effects of contextual valence shifter terms and took an account of negation terms in sentiment sentences. The negation and contextual valance shifter terms include the amplifiers, diminishes and conjunction terms in sentiment sentences, due to this, the sentiment polarity classes are classified into seven categories which are positive, strongly positive, weakly positive, negative, strong negative, weak negative, and neutral classes. Table 11 shows the evaluation results of the experiment on each sentiment polarity class.

Table 11 Evaluation Result of the Experiment

No	Target area	Sample size	Accuracy	Precession	Recall	F score
1	Addis Ababa 6 kilo campus	208	0.85	0.89	0.90	0.85
2	Addis Ababa Gofa sefer	120	0.85	0.96	0.86	0.90
3	Gofa Mebrat hail condominium area	174	0.72	0.94	0.77	0.84
4	Facebook page	1131	0.92	0.97	0.82	0.89
Average			0.84	0.94	0.84	0.87

5.2.3 Frequency of result distribution

We used the following performance metrics to evaluate our experiment results and our classifiers

A confusion matrix is used as a form of visualizing the performance of a classifier. It is displayed in a table format in which the columns represent the actual values (true and false) and the rows represent the predicted values (positive, negative, and neutral). It can easily be generalized for multi-class classifiers. The table reports the results of a classifier in terms of the number of true positives (TP), false positives (FP), false negatives (FN), true negatives (TN), False neutral (Fneu), and true neutral (Tneu).

Table 12 Experimental results by frequency

Source	TP	FN	Fneu	TN	FP	Fneu	Tneu	FN	FP
Addis Ababa 6 kilo campus	103	8	5	61	11	2	14	3	1
Addis Ababa Gofa sefer	95	15	0	7	3	0	0	0	0
Mebrat hail condominium area	104	29	7	17	3	5	4	2	3
Social Media	515	110	104	175	14	22	117	33	41
	817	162	116	260	31	29	135	38	45

The values for true positive, true negative, false positive and false negative were as shown below:

Table 13 Confusion matrix for the model

	Actual 0 (negative)	Actual 1 (positive)
Predicted 0(negative)	162	260
Predicted 1(positive)	260	817
Predicted 0(neutral)	145	135

The researchers come up with four target areas which are the Addis Ababa 6-kilo campus, Addis Ababa Gofa Sefer, Gofa Mebrat hail condominium area, and website/Facebook page. For all target areas 1633 (One thousand six hundred thirty-three) comments are collected about the program and from each target area, there would be accuracy, precession, recall, and f-score. Generally, the accuracy of this study is 84%.

5.2.4 Discussion

As shown in Table 12 above, the experimental results are different for each polarity class. In this Sub section, we discuss the results of each polarity class and the reasons for the variation of the results in each polarity class. In Table 12, the precision of positive and strong positive polarity classes are higher than the remaining polarity classes, because much of the test datasets contain positive sentiments.

In negative and strong negative polarity classes, the precision is higher than their recalls. This indicates the classification approach predicts the right sentiment classification in negative and strong negative sentiment polarity classes and the system achieves better correctness in the classification. In the neutral polarity class, the system achieves the lower precision, but the highest recall from the other polarity classes. The reason is the effect of misspelling sentiment words in sentiment sentences, the equal occurrences of sentiment terms in the sentence, the size of the sentiment words in the sentiment lexicons, and the presence of sarcastic sentences in sentiment sentences. For example, the sentence "I appreciate the way how he expresses about narrow-minded politicians", is a sentiment sentence and it has positive polarity, however, due to an equal number of sentiment term occurrence (i.e., one positive term (appreciate) and one negative term (narrow)) in the sentence, the system assigns the sentence wrongly into neutral polarity class. Other examples also, "this is a beautiful cultural wedding song!!!!!!" in this example, the sentiment sentence contains a wrong spelling sentiment word (i.e., beautiful) and the system considers all the terms in the sentence to be neutral and the sentence has neutral polarity values. If the sentiment sentence contains misspelled sentiment terms, an equal number of sentiment terms, or not correctly identified the language of the sentiment terms or the sentiment term does not occur in the sentiment lexicons, then the system automatically assigns the polarity of the sentence into neutral polarity classes. Therefore, an incorrect polarity classification of the sentiment sentence into a neutral polarity class influences the performance results of neutral polarity classes. In general, the absences of a spelling checker, the existence of sarcastic sentences, and ambiguous words in sentiment sentences have their negative impacts/influences on the overall performance of the system in sentiment classifications.

5.4 Comparison with previous works

Table 14 Contribution of the Thesis

Author (year)	Problem solved	Approach	Result	Gaps
Joshua M. Mutisya	Sentiment analysis is yet to be fully exploited to provide insights into viewership behaviour. This results in decisions being made on the basis of perception.	Tweets are classified as positive, negative, and objective (neutral). A Twitter corpus was created by fetching tweets using the Twitter API and	A series of four experiments were carried out to determine which classifier delivered the best results. Indeed, the SVM model with bigram features had the highest accuracy score, of 89 percent.	The use of multimedia content to pass coded messages on Twitter is on the rise, especially through the use of emoticons, short videos, GIF content as well as memes. Included in building the corpus necessary for training the model.

		annotating by using emoticons.	The system was then deployed to carry out similar classifications on Tweets obtained via the Twitter search API.	It would be useful if such type of content is considered.
we	This research, therefore, attempts to address the knowledge gaps or problems in sentiment analysis of sentiment sentences, which are usually available on social media.	A rule-based approach is classified as positive, negative, and objective (neutral).	Attempts to address the knowledge gaps or problems in sentiment analysis of sentiment sentences, which are usually available on social media.	Model ambiguous sentiment words help to perform proper sentiment classification

5.3 Contribution of the thesis

The findings of this study can have considerable contributions as explained here.

- Use Tigrigna language to study sentiment analysis on social media and help serve as a bridge for another TV sentiment analysis-related work.
- Developed an algorithm to recognize the language of the input text from emotional sentences.
- Developed an algorithm to detect and determine the polarity value of sentiment terms from Tigray TV sentiment sentences.
- Constructed the Tigrigna sentiment dictionary for sentiment word detection and polarity weight determination in sentiment sentences. Developed a sentiment analysis system to analyze and classify the sentiment polarity of sentiment sentences.

CHAPTER SIX

CONCLUSUION AND RECCOEMMENDATION

6.1 Conclusion and recommendation

There are many social media sites in different languages all over the world. The content of information available on social media can be in different languages. Due to the high availability of text data on social media, it needs to process social media text content written in the Tigrigna language and an easy-to-use format. In this regard, sentiment analysis attempts to use various methods to analyze and classify sentiment sentences based on sentiment polarity.

This research, therefore, attempts to address the knowledge gaps or problems in sentiment analysis of sentiment sentences, which are usually available on social media. To solve this problem, this research investigates an open-source tool that provides various services. This research work has studied the language behavior of Tigrinya, especially its morphological characteristics. A comprehensive review of sentiment analysis in social media and related fields was conducted. An understanding of technology and language behavior allows for the development of requirements for each component of the sentiment analysis system. The Tigrigna sentiment analysis system consists of a preprocessor, a language identifier, a morphological analyzer, a sentence builder using roots, a sentiment word detector, a sentiment word polarity weight determiner, and a sentiment classifier component.

When developing the sentiment analysis system, this research uses language identification components to perform morphological analysis and detect sentiment words in the corresponding language in the input text. These languages come from different families and have their own written. To maintain data uniformity and consistency, the Ethiopic script representations are less complex and contain small numbers of letters. Analyzing sentiment sentences is based on sentiment terms in the Tigrigna sentiment lexicons and some rules for detecting and determining the polarity weights of sentiment terms were built into the sentiment sentences.

Finally, to verify the goal, the sentiment analysis system was evaluated using social media data. The evaluation comes from the performance test of the system. The success of the demonstration and performance test clearly shows the feasibility of providing the system for social media sentiment text analysis. To evaluate the performance of the system in sentiment analysis, 764 sentiment sentences were collected from different topics published on Facebook and the website. To prove the effectiveness of the system accuracy, the recall, precision, and

F-measure evaluation index were carried out. The result shows that an average precision of 84.00%, an average recall rate of 83.00%, and an average F-measure of 87.00% were obtained from the experiment.

In general, the values of the evaluation metrics are encouraging, the large size of positive and negative sentiment terms in the lexicon, appropriate language identification, concerning the effects of valance shifters, and negation term terms sentiment sentences are assumed to improve the performance of the sentiment polarity classifications. This research shows promising results, but more comprehensive future work will improve more discoveries.

6.2 Future Work

The main purpose of this study was to design and develop a sentiment analysis system on social media. To develop a full-fledged sentiment analysis system on social media, it needs coordinated teamwork from linguistic expertise and computer science knowledge. Even if the system had already demonstrated good performance in the realistic setting, still it needs further progress. Future work may address the following to have a full-fledged sentiment analysis system.

- Build a more comprehensive list of positive, negative, and neutral emotion words to improve system performance.
- Language identification at the phrase level and text level helps to include sentiment classification using phrase patterns.
- Develop sentiment terms for specific domains to consider domain-related sentiment terms and improve system performance.
- Modeling ambiguous sentiment words helps to perform proper sentiment classification.
- Modeling satire or satirical sentences can help reduce misclassification of emotional polarity.
- Analyzing emotions at the feature level helps to determine emotions towards specific features/aspects of the subject.
- In the future, more corpora should be developed for Tigrigna sentiment analysis using Machin Learning.

Reference

Ref number

[1] D. Okoth, "Adapting to survive: Radio and digital media in Eastern Africa", Kas. (2021 Jul. 01). Available: https://kas.de/documents/285576/285625/Media+and+the+pandemic+-+The+fight+for+survival.pdf/3aca44b4-f20e-22c4-a1ad-d704b0bc0236?version=1.0&t=1624356800394. [Accessed 01 June. 2022].

[2] D. H, "THE RELATIONSHIP BETWEEN MEDIA OWNERSHIP AND MEDIA CREDIBILITY", Deemed to be University, vol. 8, no. 15, pp. 2–5, 2019. Available: https://worldwidejournals.com/global-journal-for-research-analysis-GJRA/recent_issues_pdf/2019/March/the-relationship-between-media-ownership-and-media-credibility_March_2019_0867288144_0113961.pdf. [Accessed 05 June. 2022].

[3] "Sentiment Analysis Challenges: Everything You Need to Know", Repustate, 11 (2021 Nov. 11). Available: https://repustate.com/blog/sentiment-analysis-challenges-with-solutions/. [Accessed 05 June. 2022].

[4] J. Nabi, "Machine Learning — Text Processing | by Javaid Nabi | Towards Data Science", Towardsdatascience. (2018 Sep. 19). Available: https://towardsdatascience.com/machine-learning-text-processing-1d5a2d638958. [Accessed 05 May. 2022].

[5] monkeylearn. "Sentiment Analysis Guide", Monkeylearn. (2020 February. 01). Available: https://monkeylearn.com/sentiment-analysis/. [Accessed 05 May. 2022].

[6] R. Eremyan, "Four Sentiment Analysis Accuracy Challenges in NLP | Toptal", Toptal. (2020 Jul. 11). Available: https://toptal.com/deep-learning/4-sentiment-analysis-accuracy-traps. [Accessed 05 May. 2022].

[7] marcobonzanini. "Marco Bonzanini", Marcobonzanini. (2021 Sep. 21). Available: https://marcobonzanini.com/. [Accessed 05 April. 2022].

[8] S. MacDonald, "Customer Complaints: Why Angry Customers Are Good for Business", Superoffice. (2021 May. 04). Available: https://superoffice.com/Blog/Customer-Complaints-Good-for-Business/. [Accessed 05 May. 2022].

[9] getthematic. "Sentiment Analysis | Comprehensive Beginners Guide | Thematic | Thematic", Getthematic, Available: https://getthematic.com/sentiment-analysis/. [Accessed 05 May. 2022].

[10] M. Liu, "[PDF] Mining and summarizing customer reviews | Semantic Scholar", Semanticscholar, (2020 Jul. 22). Available: https://semanticscholar.org/paper/Mining-and-summarizing-customer-reviews-Hu-Liu/cdcf7cb29f37ac0546961ea8a076075b9cc1f992. [Accessed 06 April. 2022].

[11] V. Techo, "ResearchGate | Find and share research", Researchgate. (2016 Jun. 01). Available: https://researchgate.net/. [Accessed 06 April. 2022].

[12] S. Bird, Natural Language Processing with Python, 1st ed. Vol. 100-150. O'Reilly Media, Inc, 2009. Available: https://oreilly.com/library/view/natural-language-processing/9780596803346/.

[13] A. Mccallum, "Please Wait.. | Cloudflare", Researchgate. (2021 Mar. 01). Available: https://researchgate.net/publication/2408220_A_Comparison_of_Event_Models_for_Naive_Bayes_Text_Classification. [Accessed 06 April. 2022].

[14] J. Zhang, "Personalized Web Search Using Emotional Features", Researchgate. (2019 Sep. 01). Available: https://researchgate.net/publication. [Accessed 06 March. 2022].

[15] P., & L. (2015, October 13). An Introduction to Sentiment Analysis (Opinion Mining). Https://Www.Meaningcloud.Com.

[16] O'Neill, A. (2009, November 01). Sentiment Mining for Natural Language Document. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.353.1006&rep=rep1&type=pdf

[17] K. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications", Researchgate. (2015 Jun. 01). Available: https://researchgate.net/publication. [Accessed 06 February. 2022].

[18] Gebremeske, S. (2010, November 01). SENTIMENT MINING MODEL FOR OPINIONATED AMHARIC TEXTS. Http://213.55.95.56/. http://213.55.95.56/bitstream/handle/123456789/3029/Selama%20Gebremeskel.pdf?sequence=1&isAllowed=y

[19] W. Philemon, "A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts", Hilcoe, 01 Sep. 2020. Available: https://hilcoe.net/wp-content/uploads/2020/08/V2N2Paper12.pdf. [Accessed 06 February. 2022].

[20] Dong, Z. (2010, September 01). HowNet and Its Computation of Meaning. Https://Aclanthology.Org/. https://aclanthology.org/C10-3014.pdf

[21] Step-By-Step Twitter Sentiment Analysis: Visualizing Multiple Airlines' PR Crises [Updated for 2020]. (2017, May 26). Https://Ipullrank.Com/Step-Step-Twitter-Sentiment-Analysis-Visualizing-United-Airlines-Pr-Crisis.

[22] Lee, P. (2008, January 12). Opinion Mining and Sentiment Analysis. https://www.researchgate.net/publication/215470760_Opinion_Mining_and_Sentiment_Analysis

[23] Vineet, P. R. A. B. U. (2013). Serendio: Simple and Practical lexicon based approach to Sentiment Analysis. Https://Aclanthology.Org/. https://aclanthology.org/S13-2091.pdf

[24] Rajput, R., & Solanki, A. (2016, February 2). Review of Sentimental Analysis Methods using Lexicon Based Approach. Https://Ijcsmc.Com/. https://ijcsmc.com/docs/papers/February2016/V5I2201636.pdf

[25] Machine learning ", En. (2021 Jul. 01). Available: https://en.wikipedia.org/wiki/Machine_learning. [Accessed 06 March. 2022].

[26] Pk, K. (2017, October). Insights to Problems, Research Trend and Progress in Techniques of Sentiment Analysis. Https://Www.Researchgate.Net. https://www.researchgate.net/publication/320248660_Insights_to_Problems_Research_Trend_and_Progress_in_Techniques_of_Sentiment_Analysis

[27] S. Ray , "6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python and R", Analyticsvidhya. (2017 Sep. 11). Available: https://analyticsvidhya.com/blog/2017/09/naive-bayes-explained/. [Accessed 06 January. 2022].

[28] N. Mehra, "Sentiment Identification Using Maximum Entropy Analysis of Movie Reviews", Researchgate. (2021 Jan. 01). Available: https://researchgate.net/publication. [Accessed 06 January. 2022].

[29] L. Lee, B. P., & S.V. (2002). Sentiment classification using machine learning techniques. Https://Www.Bibsonomy.Org. https://www.bibsonomy.org/bibtex/1c6d2803bbde23d9d9fc1acf00182a84b/oliver_awm

[30] Turney, P. D. (2002 November. 01). Thumbs Up or Thumbs Down? {S}emantic Orientation Applied to Unsupervised Classification of Reviews. Https://Www.Researchgate.Net. https://www.researchgate.net/publication/248832100_Thumbs_Up_or_Thumbs_Down_Semantic_Orientation_Applied_to_Unsupervised_Classification_of_Reviews

[31] Min Kim, S. (2004, August). Determining the sentiment of opinions. Https://Dl.Acm.Org. https://dl.acm.org/doi/10.3115/1220355.1220555

[32] Mwita, C. (2020). The Kenya Media Assessment 2021. Https://Internews.Org/. https://internews.org/wp-content/uploads/2021/04/KMAReport_Final_20210325.pdf

[33] Reda, D. T. (2012). Zebenawi Sewasw Kuankua Tigirigna. Https://Www.Mereb.Shop/. https://www.mereb.shop/rs/?prodet=true&pid=90282761&vid=99

[34] Repustate Team. "Sentiment Analysis Challenges: Everything You Need to Know", Repustate. (2021 Jun. 11). Available: https://repustate.com/blog/sentiment-analysis-challenges-with-solutions/.

[35] Jayaswal, V. (2020, September 14). Performance Metrics: Confusion matrix, Precision, Recall, and F1 Score. Https://Towardsdatascience.Com. https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262

[36] Video Research . "Opportunities and Achievement in Japan", Videor, 2020. Available: https://videor.co.jp/eng/company/index.html. [Accessed 06 January. 2022].

[37] Z. Fang, "Sentiment analysis using product review data", Journalofbigdata. (2015 May. 05). Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0015-2. [Accessed 06 February. 2022].

[38] Fikre, T. (2020, July 4). EFFECT OF PREPROCESSING ON LONG SHORT TERM MEMORY BASED SENTIMENT ANALYSIS FOR AMHARIC LANGUAGE. Http://213.55.95.56/. http://213.55.95.56/bitstream/handle/123456789/22638/Turegn%20Fikre.pdf?sequence=1&isAllowed=y

[39] T. HAILU , "OPINION MINING FROM AMHARIC BLOG", 213.(2021 Apr. 04). Available: https://213.55.95.56/bitstream/handle/123456789/3440/Tulu%20Tilahun.pdf?sequence=1&isAllowed=y. [Accessed 06 February. 2022].

[40] Šandrih, " IMPACT OF TEXT CLASSIFICATION ON NATURAL LANGUAGE PROCESSING APPLICATIONS", Academia. (2020 Jan. 1). Available: https://academia.edu/. [Accessed 06 January. 2022].

[41] Miller, R. (2019, December 13). Data Preprocessing: what is it and why is important. Https://Ceoworld.Biz/2019/12/13/Data-Preprocessing-What-Is-It-and-Why-Is-Important/

[42] Lemaire, V., Salperwyck, C., & Bondu, A. (2015, March 1). A Survey on Supervised Classification on Data Streams. Https://Www.Researchgate.Net/Publication/270787580.

Annexes

Annexone: Sample Tigrigna Positive Sentiment Words

ስሉጥ	ጥዑም	ግርም	ሓጎስ	ፍሽኽ	ፍቅሪ	ዕንበባ
ስልጥቲ	ጥዑምቲ	ግርምቲ	ሕጉስቲ	ፍሽኽቲ	ፍቅርቲ	ዕምብብቲ
ብስሉጥ	ብጥዑም	ብግርም	ብሓጎስ	ብፍሽኽታ	ብፍቅሪ	ብዕንበባ
ስሉጣት	ጥዑማት	ግሩማት	ሕጉሳት	ፍሹኻት	ፍቁራት	ዕምበባታት
ሰሊጠ	ጥዕመ	ገሪመ	ሓጉሰ	ፈሺኽ	አፍቅረ	ዓምቢበ
ሰሊጥና	ጥዒምና	ገሪሙና	ተሓጉስና	ተፈሺኽና	ተፈቂርና	ዓምቢብና
ሰሊጠን	ጥዒመን	ተገሪመን	ተሓጉሰን	ተፈሺኽን	ተፈቂረን	ዓምቢበን
ሰሊጥኩም	ጥዒምኩም	ተገሪምኩም	ተሓጉስኩም	ተፈሺኹም	ተፈቂርኩም	ዓምቢምኩም
ሰሊጦም	ጥዒሞም	ተገሪሞም	ተሓጉሶም	ተፈሺኾም	ተፈቂሮም	ዓምቢቦም
ሰሊጣቶም	ጭዒማቶም	ገሪማቶም	አሐጉሳቶም	አፈሸኻቶም	አፍቂራቶም	ዐዕቢባቶም
ሰሊጣተን	ጥዒማተን	ገሪማተን	አሐጉሳተን	አፈሸኻተን	አፍቂራተን	ዐዕቢባተን
ክሰልጠን	ክጥዕመን	ክገርመን	ክሕጎሳ	ክፍሽኻ	ክፍቀራ	ክዕበባ
ክሰልጥ	ክጥዕም	ክገርም	ክሕጎስ	ክፍሽኽ	ክፍቀር	ክዕበብ
ክትሰልጥ	ክትጥዕም	ክትገርም	ክትሕጎስ	ክፍሽካ	ክፍቀራ	ክዕብባ
ሱሉጥነት	ጥዑምነት	ጉሩምነት	ሑስነት	ፍሽኽታነት	ፉቁራት	ዑምበባታት
ምስላጦም	ምጥዓሞም	ምግራሞም	ምሕጓሶም	ምፍሻኾም	ምፍቃሮም	ምዕምባቦም
ጽቡቅ	ለጋስነት	ምዕቡል	ርህሩህ	ዕምበባ	ወሓለ	ዘደንቕ
ጽቡቅቲ	ለጋስ	ማዕበለ	ራህሪሀ	ዓምበበ	ዉሒልና	ተደነቐ
ብጽቡቅ	ለገሰ	ማዕቢልና	ራህሪህና	ዓምቢብና	ወሓለታት	ተደኒቕና
ጽቡቃት	ለጊስና	ምዕቡላት	ርህሩሃት	ዓምበብቲ	ውሒለን	ተደነቕቲ
ጸቢቀ	ለጋሳት	ማዕቢለን	ራህሪሀን	ዓምቢበን	ውሒልኩም	ተደኒቐን
ጸቢቅና	ለጊሰን	ማዕቢልኩም	ራህሪህኩም	ዓምቢብኩም	ውሒሎም	ተደኒቕኩም
ጸቢቀን	ለጊስኩም	ማዕቢሎም	ራህሪሆም	ዓምቢቦም	ውሒላቶም	ተደኒቖም
ጸቢቅኩም	ለጊስክን	ማዕቢላቶም	ራህሪሃቶም	ዓምቢባቶም	ውሒላተን	ተደኒቓቶም
ጸቢቆም	ለገስቲ	ማዕቢላተን	ራህሪሃተን	ዓምቢባተን	ውሒልክን	ተደኒቓተን
ጸቢቃቶም	ክልግስ	ማዕቢልክን	ራህሪህክን	ዓምቢብክን	ውሕለይቲ	ተደኒቐክን
ጸቢቃተን	ትልግስ	ምዕብልቲ	ርህርህቲ	ዕምበባታት	ወሓለት	ተደኒቆም
ጸቢቅክን	ክልግሱ	ምምዕባሎም	ረህርህነት	ዓምቢቦም	ክውሕል	ተደኒቐ
ክጽብቅ	ክትልግስ	ምምዕባላተን	ክርህርህ	ዓምበበት	ትውሕል	ተደነቐት
ትጽብቅ	ክልግሳ	ምምዕባላቶም	ትርህርህ	ክዕምብብ	ክውሕሉ	ክተደንቕ
ክትጽብቅ	ክትልግሱ	ምዕባለ	ክርህርሁ	ትዕምብብ	ክትውሕል	ክደንቑ
ጽብቅነት	ምልጋስኩም	ምዕባልነት	ክትርህርህ	ክዕምብቡ	ክውሕላ	ክተደንቕ
ክጽብቁ	ምሌጋሳቶም	ክምዕብል	ክርህርሃ	ክትዕምብብ	ምውሓሎም	ከደንቓ
ክትጽብቁ	ምልጋሳተን	ክምዕብሉ	ክትርህርህ	ክዕምብባ	ምውሓላተን	ድንቒ
ክጽብቃ	ምልጋሶም	ክትምዕብል	ምርህራሆም	ምዕምባቦም	ምውሓሎም	ምድናቑም
ምጽባቆም	ምልጋሳተን	ክምዕብላ	ምርህራሃተን	ምዕምባበተን	ውሓለ	ምድናቓተን
ምጽባቃተን	ምልጋሳቶም	ክምዕብሉ	ምርህራሃቶም	ዝዓምበበ	ዝውሓለ	ዘደነቐ
ምጽባቃቶም	ዝለገሰ	ትምዕብል	ዝራህረሀ	ዝዓምበበት	ዝውሓለት	ዘደነቐት
ዝጸበቀ	ዝለገሰት	ዝማዕበለ	ዝራህረሀት	ዝዓምበብካ	ዝውሓልካ	ዘደንቐካ
ዝጸበቀት	ዝለገስካ	ዝማዕበለት	ዝራህረህካ	ዝዓምበብኩም	ዝውሓልኩም	ዘደነቕኩም
ዝጸበቅካ	ዝለገስኩም	ዝማዕበልካ	ዝራህረህኩም	ዝዓምበብክን	ዝውሓልክን	ዘደነቕክን
ዝጸበቅኩም	ዝለገስክን	ዝማዕበልኩም	ዝራህረህክን	ዝዓምበብና	ዝውሓልና	ዘደነቕና
ዝጸበቅክን	ዝለገስና	ዝማዕበልክን	ዝራህረህና	ዕውት	ዋሓለ	ዘገርም

ዝጸበቅና	መልክዐኛ	ዝማዕበልና	ቀሲና	ተዓወተ	ውሕልና	ተገረመ
ትሕጎስ	መልክዐ	ስልጡን	ቅሱን	ተዓወትና	ዋሕለለ	ተገሪምና
ተሓጎሰ	መልኪዕና	ሰልጠነ	ቀሰነ	ተዓወትቲ	ውሕሉል	ተገረምቲ
ሕጉስ	መልክዐኛታት	ሰልጢና	ቊስናት	ትዕወት	ውሕልነት	ትገርም
ሕጉሳት	መልኪዐን	ስልጡናት	ቀሲነን	ተዓዊተን	ዉዕዉዕ	ተገሪመን
ሕጉስቲ	መልኪዕኩም	ሰልጢነን	ቀሲንኩም	ተዓዊትኩም	ዋዕወዐ	ተገሪምኩም
ተሓጊሰ	መልክዕ	ሰልጢንኩም	ቀሲኖም	ተዓዊቶም	ዋዕዊዕና	ተገሪሞም
ተሓጊስና	ክምልክዕ	ሰልጢኖም	ቀሲናቶም	ተዓዊታቶም	ዉዕዉዓት	ተገሪማቶም
ተሓጊሰን	ትምልክዕ	ሰልጢናቶም	ቀሲናተን	ተዓዊታተን	ዋዕዊዐን	ተገሪማተን
ተሓጊስኩም	ክምልክዑ	ሰልጢናተን	ቀሲንክን	ተዓዊትክን	ዋዕዊዕኩም	ተገሪምክነ
ተሓጊሶም	ክትምልክዕ	ሰልጢንክን	ቅስንቲ	ተዓወተ	ዋዕዊዖም	ተገረመት
ተሓጊሳቶም	ክምልክዓ	ስልጥንቲ	ቅሳነት	ተዓወተት	ዋዕዊዓቶም	ክተገርም
ተሓጊሳተን	ክምልክዑ	ስልጥንነት	ክቀስን	ክዕወት	ዋዕዊዓተን	ክግረሙ
ተሓጊስክን	ምምልካዕኩም	ክስልጥን	ክቀስኑ	ክዕወቱ	ዋዕውዕክን	ክትግረም
ክሕጎስ	ምምልካዓቶም	ትስልጥን	ክትቀስን	ክትዕወት	ውዕውዕቲ	ክግረማ
ክሕጎሱ	ምምልካዓተን	ክስልጥኑ	ክቀስና	ክዕወታ	ውዕዋዐ	ምግራሞም
ክትሕጎስ	መልኪዕክን	ክትስልጥን	ክትቀስን	ዕውቲ	ዋዕወዐት	ምግራማተን
ክሕጎሳ	ምልኩዕቲ	ክስልጥና	ምቅሳኖም	ዕውታት	ትውዕውዕ	ምግራሞም
ምሕጋሶም	ምምልካዖም	ክትስልጥኑ	ምቅሳናተን	ምዕዋቶም	ክውዕዉዕ	ዘገረመ
ምሕጋሳተን	ምምልካዓተን	ምስልጣኖም	ምቅሳናቶም	ምዕዋታተን	ክውዕውዑ	ዘገረመት
ምሕጋሳቶም	ምምልካዓቶም	ምስልጣናተን	ዝቀሰነ	ተዓዊቶም	ክትውዕውዕ	ዘገረምካ
ታሕጋስ	ዝመልክዐ	ምስልጣናቶም	ዝቀሰነት	ዝተዓወተ	ክውዕውዓ	ዘገረምኩም
ታሕጋስነት	ዝመልክዐት	ዝሰልጠነ	ዝቀሰንካ	ዝተዐወተት	ምውዕዋዖም	ዘገረምክን
ተሓጎሰ	ዝመልካዕካ	ዝሰልጠነት	ዝቀሰንኩም	ዝተዓወትካ	ምውዕዋዐን	ዘገረመና
ዝተሓጎሰ	ዝመልካዕኩም	ዝሰልጠንካ	ዝቀሰንክን	ዝተዓወትኩም	ምውዕዋዖም	ዝሳነ
ዝተሓጎሰት	ዝመልክዕክን	ዝሰልጠንኩም	ዝቀሰና	ዝተዓወትክን	ዝዋዕወዐ	ተሳነየ
ዝትሓጎስካ	ዝመልክዕና	ዝሰልጠንክን	በሊጽና	ዝተዓወትና	ዝዋዕወዐት	ተሳኒና
ዝትሓጎስኩም	ማራኺ	ዝሰልጠና	ብሉጽ	እዱብ	ዝዋዕወዕካ	ተሳኒኩም
ዝትሓጎስክን	ማረኸ	ስምረት	በለጸ	አደበ	ዝዋዕወዕኩም	ተሳኒየን
ዝትሓጎስና	ማሪኽና	ሰመረ	በሊጽና	አዲብና	ዝዋዕወዕክን	ተሳኒዮም
ልብምነት	ማራኺት	ሰሚርና	ብሉጻት	አዲብኩም	ዝዋዕወዕና	ተሳኒያቶም
ለባም	ማሪኽን	ሰሙራት	በሊጸን	እዲብቲ	ውዕዋዐ	ተሳኒያተን
ለባማት	ማሪኹም	ሰሚረን	በሊጽኩም	እዱባት	ዉቁብ	ተሳኒክን
ትልብም	ማሪኽን	ሰሚርኩም	በሊጾም	ክእደቡ	ውቃበ	ተሳናይት
ክልብም	ማራኽነት	ሰሚሮም	በሊጻቶም	ክእደብ	ውቁባት	ተሳነየ
ክልብሙ	ክማርኽ	ሰሚራቶም	በሊጻተን	ክትእደብ	ዉቅብቲ	ተሳነየት
ክትልብም	ክማርኹ	ሰሚራተን	በሊጽክን	ክእደባ	ወቂበን	ትሳነ
ክልብሙ	ክትማርኽ	ሰሚርክን	ብልጽቲ	ተአዲበን	ወቂብኩም	ክሳነ
ክልብማ	ክማርኻ	ሰምርቲ	ክበልጽ	ተአዲብኩም	ወቂቦም	ክሳነዪ
ለቢመ	ክማርኹ	ስምርነት	ትበልጽ	ተአዲቦም	ወቂባቶም	ክትሳነይ
ለቢሞና	ምምራኹ	ክሰምር	ክበልጹ	ተአዲባቶም	ወቂባተን	ክሳነያ
ለቢመን	ምምራኻቶም	ክሰምሩ	ክትበልጻ	ተአዲባተን	ወቂብክን	ክሳነዩ

ለቢምኩም	ምምራኻተን	ክትሰምር	ክበልጻ	ተአዲብክን	ውቅብቲ	ምስናዮም
ለቢሞም	ማራኺት	ከሰምራ	ክበልጹ	ትእደብ	ክውቅቡ	ምስናያተን
ለቢማቶም	ምምራኹም	ክትሰምሩ	ምብሊጾም	ተአዱቦም	ዉቃበ	ምሰናዮም
ለቢማተን	ምምራኹን	ምስማሮም	ምብላጻተን	ተአደበ	ውቅብና	ዘስማዕማዕ
ለቢምክን	ምምራኻቶም	ምስማራተን	ምብላጻቶም	ተአደበት	ክውቁብ	ተስማዕሚዐ
ለበምቲ	ዝማረኸ	ምስማራቶም	ብልጹነት	ምእዳቦም	ትውቅብ	ተስማዕሚዕና
ምሌባሞም	ዝማረኸት	ዝሰመረ	ዝበሆጸ	ምእደባተን	ክውቁቡ	ተስማዕኩም
ምሌባማተን	ዝማረኽካ	ዝሰመረት	ዝበለጸት	ዝተአደበ	ክትውቁብ	ትስማዕማዕ
ምሌባማቶም	ዝማረኽኩም	ዝሰመርካ	ዝበለጽካ	ዝተአደበት	ክውቅባ	ተስማዕሚዐን
ዝለበመ	ዝማረኽክን	ዝሰመርኩም	ዝበለጽኩም	ዝተአደብካ	ምውቃቦም	ተስማዕሚዕኩም
ዝለበመት	ዝማረኽና	ዝሰመርክን	ዝበለጽክን	ዝተአደብኩም	ምውቃበን	ተስማዕሚያም
ዝለበምካ	ምቹእ	ዝሰመርና	ዝበለጽና	ዝተአደብክን	ምውቃቦም	ተስማዕሚዕኩም
ዝለበምኩም	መቸአ	ርቡሕ	በሪህና	ዝተአደብና	ዝወቀበ	ተስማዕሚዒተን
ዝለበምክን	መቺእና	ትረብሕ	ብሩህ	ኩርዕ	ዝወቀበት	ተስማዕሚዕክን
ዝለበምና	ምቾኣት	ረብሓ	በርሀ	ከርዐ	ዝወቀብካ	ክስማዕማዕ
ለዋህ	መቺአን	ረብሐ	በሪህና	ኮሪዕና	ዝወቀብኩም	ክስማዕመው
ለወሀ	መቺኡኩም	ረብሒና	ብሩሃት	ኩርዓት	ዝወቀብክን	ክትስማዕማዕ
ለዊህና	መቺኦም	ርቡሓት	በሪሀን	ክኮርዕ	ዝወቀብና	ክስማዕመዒ
ለዋሃት	መቺኣቶም	ረቢሐን	በሪህኩም	ክኮርው	ዘዕግብ	ክስማዕመው
ክልውህ	መቺአተን	ረቢሕኩም	በሪሆም	ክትኮርዕ	አዕገበ	ምስምማያም

AnnexTwo: Sample Tigrigna Negative Sentiment Words

ኪሕዶም	ኮራዪ	መሕዘኒ	መከራ	ሰረዘ	ደንቆሮ	ጽሉልቲ
ኪሕደን	ኮራይት	መሕዘኒት	መከራታት	ሰረዘት	ደንቆረ	ፈዛዝ
ካሕዳም	ኮሪይና	ከሕዝኑ	መውት	ሰረዙ	ደንቆረት	ፈዘዘ
ክሕዳቶም	ኮሪያቶም	ክተሕን	ዝሞተ	ሰሪዘን	ደንቂረን	ፈዚዞም
ክሕዳተን	ኮሪያተን	ከሕዝና	ዝሞተት	ሰሪዞም	ዘደንቆረ	ፈዛዛት
ካሕዳማት	ክኩሪይ	ከሕዝኑ	ዝሞቱ	ምስራዝ	ዝደንቆረት	ፈዚዘን
ክፉእ	ኮረየት	አሕዘንቲ	ዝሞታ	ሰቓይ	ዝደንቆሩ	ዝፈዘዘ
ከፍአ	ክትኩሪይ	መሕዘኖ	መድረበዳ	አሳቐየ	ዝደንቆርና	ዝፈዘዘት
ከፊኦም	ክክርያ	አሕዚኖም	መድረባዳታት	አሳቐየት	ደንቆረታት	ዝፈዘዙ
ከፊአን	ኮረይቲ	አሕዚነን	መጺጽ	ስቐያት	ድንቁርና	ዝፈዘዛ
ከፊኣቶም	ምክራየን	አሕዚንና	መጸጽቲ	ዝተሳቐየ	ዲያብሎስ	ፋሕሻዋት
ከፊኣተን	ምክርዮም	አሕዚናተን	ክመጽጽ	ዝተሳቐየት	ዲያብሎሳት	ፋሕሻው
ከፊእና	ምክራያቶም	ሕዙን	መጸጸ	ተሳቐየ	ድንዙዝ	ፋሕሸወ
ዝከፍአ	ምክራያተን	ሕዝንቲ	ዝመጸጸ	ሰነፍ	ደንዘዘ	ፋሕሺዎም
ዝከፍአት	ዝኮረየ	አሕዘነ	ማእሰርቲ	ሰነፈ	ደንዘዘት	ፋሕሺወን
ዝከፍኡ	ዝኮረያ	ዘሕዝን	ተኣሰረ	ሰነፈት	ደንዚዞም	ዝፋሕሸወ
ዝከፍኣ	ዝኮረዩ	ሓደገኛ	ተአሰረት	ሰነፋት	ደንዚዘን	ዝፋሕሸወት
ክፉአት	ዝኮረይና	ሓደጋ	ዝተኣሰረ	ሰኒፍና	ዱንዙዛት	ፎካስ
ትከፍእ	ዝኮረይኩም	ብሓደጋ	ዝትኣሰረት	ሰኒፎም	ደንዚዛተን	ፎኪስ
ክትከፍእ	ዝኮረይክን	ሕልኽልኽ	ዝተኣሰሩ	ሰኒፈን	ደንዚዛቶም	ፎከሰ
ክፉኣት	ሃሰየ	መሕለኽለኺ	ዝተኣሰራ	ዝሰነፈ	ዝደንዘዘ	ፎከስቲ
መክፈኢ	ሃሰየት	ዘሕለኽለኽ	መእሰርቲ	ዝሰነፈት	ዝደንዘዘት	ፎካሳት
ዘክፍእ	ሃሰይና	ሕልኽልኽነት	እሱራት	ዝሰነፉ	ዝደንዘዙ	ዝፎከሰ
ክፉኣት	ሃሰይኩም	ሕማቕ	እሱር	ሰኒፎም	ዝደንዘዛ	ዝፎከሱ
ወንጀለኛ	ሃሰይክን	ሓመቕ	ምምራር	ሰኒፈን	ደነዝ	ፎኪሰን
ወንጀል	ሃስያተን	ሓመቕት	አማረረ	ሰኒፉቶም	ድንገት	ዝፎከሳ
ወንጀለኛታት	ሃስያቶም	ሐሚቖም	አማረረት	ሰኒፉተን	ድንገተኛ	ጠሉማቶም
ወደቐ	ክሃስያ	ሓሚቐን	አማረርና	ዝሰነፈ	ብድንገት	ዝጠለሙ
ወዲቐ	ክሃስዩ	ሕማቓት	ከማርር	ዝሰነፈት	ገበን	ዝጠለማ
ወዲቕና	መህሰይቲ	ዝሓመቐ	ክተማርር	ሰአነ	ገበነኛ	ጥልመት
ወደቐት	ሃሰይቲ	ዝሓመቐት	መማረርቲ	ስኡን	ገበነኛታት	ጠላሚ
ወደቑ	ዝሃሰየ	ዝሓመቑ	ዝማርሩ	ስኡናት	ጉሁይ	ጠለምቲ
ወዲቖም	ዝሃሰየት	ዝሓመቕና	ዝማርራ	ዝሰኣነ	ጉሁያት	ጠሊመ
ወዲቓቶም	ዝሃሰይኩም	ዝሓመቕክን	አማሪሮም	ዝሰኣነት	ጎሃየ	ጠሊማ
ወዲቓተን	ዝሃሰይክን	ዝሓመቕኩም	አማሪረን	ሰክራም	ጎሃየት	ጠላም
ዝወደቐ	ዝሃሰይና	ብሕማቕ	ከማርራ	ሰከረ	ጉሒላ	ጥምየት
ዝወደቐት	ሃስዩ	ሕሱም	ምርባጽ	ሰከረት	ጉሒላታት	ጠመየ
ዝወደቑ	ሃሰየት	ሕሱማት	ተረበጸ	ሰካራማት	ዝጎሓለ	ጠመየት
ዝወደቓ	ሃሳስ	ሕሱምቲ	ተረበጸት	ዝሰከረ	ዝጎሓለት	ጥሙያት
ውድቐት	ሃሳሳት	ክትሓስም	ተረቢጽና	ዝሰከረት	ዝጎሓሉ	ጠሚይና
ውግእ	ክሃሱ	ክሓስማ	ክትርበጽ	ዝሰክር	ዝጎሓላ	ጠሚየን
ተዋግአ	ካሃሳ	ሓሳማት	ክርበጸ	ሰይጣን	ጉሒሎም	ጠሚዮም
ተዋጊኦም	ምህሳሰን	ክሓስም	ርቡጻት	ሰይጣናት	ጉሒለን	በደለኛ
ተዋጊአን	ምህሳሶም	ክሓስሙ	ርብጽቲ	ሰያጥን	ጉድለት	በደለ
ተዋጊአቶም	ምህሳሳቶም	ዝሓሰመ	ዝተረበጸ	ረሳሕ	ጎዶሎ	በደል

ተዋጊአተን	ምህሳሳተን	ዝሓሰመት	ዝተረበጸት	ረስሐ	ጎደሎታት	ተበደለ
አዋገአ	አህሱስ	ዝሓሰምና	ክርበጽ	ረስሓት	ዝጎደለ	በዲለን
አዋገአት	ዝሃሰሰት	ምሕሳሞም	ክርበጹ	ረሳሓት	ዝጎደለት	በዲላቶም
አዋግእና	ዝሃሰሰ	ምሕሳመን	ተረቢጸም	ዝረስሑ	ዝጎደሉ	በዲላተን
አዋጊአን	ሃሰሱ	ምሕሳማና	ተረቢጸን	ዝረስሓ	ጉድአት	ዝበደለ
አዋገእቲ	ዝሃሰሰና	ሕሱር	ምሽቓል	ክረስሓ	ጎድአ	ዝበደለት
ዘሕፍር	ዝሃሰስክን	ሕሳር	ተሻቐለ	ክረስሑ	ተጎድአ	ተቓወመ
አሕፈረ	ሓራቕ	ሕስረት	ተሻቐለት	ረበሸ	ተጎድአት	ተቓዋሚ
አሕፈረት	ከሕርቑ	ዝሓሰረ	ተሻቀልና	ረባሺ	ተጎዱኦም	ተቓውሞ
መሕፈሪ	ከሕርቓ	ዝሓሰረት	ተሻቓሊት	ረበሸቲ	ተጎዲአን	ተቓወምቲ
ዘሕፍሩ	መሕረቕቲ	ዝሓሰርና	ክትሻቐል	ረበሸት	ተጎዱኣቶም	ተካል
ዘሕፈራ	ዝሓረቐ	ሓሳር	መሻቐሊ	ረቢሽና	ተጎዱኣተን	ተካላት
ዘሰቅቕ	ዝሓረቕና	ሕሳራት	ዝተሻቐለ	ረቢሽን	ዝተጎድአ	ክትክላ
መሰቀቒ	ዝሓረቕክን	ዝሓሰረ	ዝተሻቐለት	ረባሺት	ዝተጎድአት	ተጸላኢ
ሰቀቀን	ዝሓረቕኩም	ዝሓሰረት	ዘሻቕል	ረባሾ	ዝተጎድኡ	ጸሌአ
ሰቐቓ	ዘሕርቕ	ብዜሓሰረ	ምዕማጽ	ራብሻ	ዝተጎድኣ	ጸላኢ
ዘስካሕክሕ	ዘሕርቑ	ሕሱራት	ዓመጽ	ዝረበሸ	ጌጋ	ጽልኢ
መስካሕከሒ	ሓርፋፍ	ሕርቃን	ዓመጸ	ተረበሸ	ጌጋታት	ጸላእቲ
አስከሕካሒ	ሓርፋፋት	ሓረቀ	ዓመጸት	ተረብሽና	ግጉይ	ተጻራሪ
ዘስካሕክሓ	ክሕርፍፍ	ሓረቀት	ዓሚጽና	ርጉም	ጎሓፍ	ተጻረረ
መስካሕከሕቲ	ክትሕርፍፍ	ሓረቑ	ተዓመጸ	ረገመ	ጎሓፋት	ተጻረርቲ
ዘቃጽል	ምሕርፋፎም	ሓሪቆም	ተዓመጸት	ረገመት	ጎዳኢ	ተጻሪሮም
ቃጸሎ	ምሕርፋፈን	ሓሪቃተን	ዝትዓመጸ	ረጊምና	ጎድአ	መርዚ
መቃጸሊ	ምሕርፋፍና	ከሕርቅ	ዝትዓመጸት	ዝተረገመ	ጎዲኦም	መርዛም
መቃጸልቲ	ሓርፈፈ	ክተሕርቅ	ዝተዓመጹ	ዝተረገመት	ጎድአት	ሃካያት
አቃጸለት	ሓርፈፈት	ከሓርቁ	ዝተዓመጻ	ርጉማት	ጎዲአን	ታህኪይት
አቃጸለ	ዝሓርፈፈ	ከሓርቃ	ዝተዘዓመጽና	ረጊሞም	ጎዱኣቶም	ተሃንጢዮም
አቃጺሎም	ዝሓርፈፈት	አሕሪቆም	ዓማጺ	ረጊመን	ጎዲኣተን	ተሃንጢየን
አቃጺለን	ሃንደበታዊ	አሕሪቀን	ቀናእ	ሸፋጢ	ጎዲእና	በዓለገ
ዝንጉዕ	ሃንደበት	አሕረቀ	ቀንአ	ሸፈጠ	ዝጎድአ	ባርያ
ዘንገዐ	ብሃንደበት	ሓሪቃቶም	ቀንአት	ሸፈጢት	ዝጎድአት	ሽጉራት
ዘንገዐት	ሃካይ	አሕረቀት	ቀንኢ	ሸፋጦ	ዝጎድኡ	መሽገርቲ
ዝንጉዓት	ተሃኪይና	አሕረቅቲ	ዘቅንእ	ሸፈጥ	ዝጎድኣ	ጠሊመን
ደመሰሰ	ተሃኪዮም	ዘሕርቅ	ቀናኣት	ሽጉር	ዝጎድአና	ጠሉማተን
ደምሰሰት	ተሃኪየን	መሕረቅቲ	ቅንኣት	ተሸገረ	ጎዳእቲ	ከሓደት
ለመነ	ህኩያት	ህንጡይ	ቀጢን	ተሸገረት	ጉድኣት	ኪሕደን
ለማኒ	ዝተሃከየ	ተሃንጠየ	ቀጣን	ተሸጊርና	ጠለመ	ክሕደት
ለማኒት	ዝተሃከየት	ተሃንጠየት	ቑጡዕ	ተሸጊሮም	ጠሊምና	ዝከሓዱ
ለመንቲ	ዝተሃከዩ	ተሃንጢይና	ቑጣዐ	ሽግር	ጠሉሞም	ዝከሓዳ

AnnexThree: Inflection of Tigrigna Sentiment Words

Word	Inflection	Analysis	Latin	Prefix	Stem	Suffix	Root	Root
ጽቡቕ	ጽቡቕ	ጽቡቕ	Sbuq	-	Sbuq	-	Sbq	ጽ-ብ-ቕ
	ጽቡቕቲ	ጽቡቕ-ቲ	Sbuq-ti	-			-ti	
	ብጽቡቕ	ብ-ጽቡቕ	b-Sbuq	b-			-	
	ጽቡቓት	ጽቡቕ-ኣት	Sbuq-at	-			-at	
	ጸቢቐ	ጸቢቕ-አ	Sebiq-e	-	Sebiq		-e	
	ጸቢቕና	ጸቢቕ-ና	Sebiq-na	-			-na	
	ጸቢቐን	ጸቢቕ-አን	Sebiq-en	-			-en	
	ጸቢቕኩም	ጸቢቕ-ኩም	Sebiq-kum	-			-kum	
	ጸቢቖም	ጸቢቕ-ኦም	Sebiq-om	-			-om	
	ጸቢቓቶም	ጸቢቕ-ኣቶም	Sebiq-atom	-			-atom	
	ጸቢቓተን	ጸቢቕ-ኣተን	Sebiq-aten	-			-aten	
	ጸቢቕክን	ጸቢቕ-ክን	Sebiq-kn	-			-kn	
	ክጽብቕ	ክ-ጽብቕ	k-Sbq	k-	Sbq		-	
	ትጽብቕ	ት-ጽብቕ	t-Sbq	t-			-	
	ክትጽብቕ	ክት-ጽብቕ	kt-Sbq	kt-			-	
	ጽብቕነት	ጽብቕ-ነት	Sbq-net	-			-net	
	ክጽብቑ	ክ-ጽብቕ-ኡ	k-Sbq-u	k-			-u	
	ክትጽብቑ	ክት-ጽብቕ-ኡ	kt-Sbq-u	Kt-			-u	
	ክጽብቓ	ክ-ጽብቕ-ኣ	k-Sbq-a	k-			-a	
	ምጽባቖም	ም-ጽባቕ-ኦም	m-Sbaq-om	m-	Sbaq		-om	
	ምጽባቓተን	ምጽባቕ-ኣተን	m-Sbaq-aten	m-			-aten	
	ምጽባቓቶም	ም-ጽባቕ-ኣቶም	m-sbbaq-atom	m-			-atom	
	ዝጸበቐ	ዝ-ጸበቕ-አ	z-sebaq-et	z-	Sebeq		-e	
	ዝጸበቐት	ዝ-ጸበቕ-አት	z-sebaq-et	z-			-et	
	ዝጸበቕካ	ዝ-ጸበቕ-ካ	z-sebeq-ka	z-			-ka	
	ዝጸበቕኩም	ዝ-ጸበቕ-ኩም	z-sebeq-kum	z-	sebeq		-kum	
	ዝጸበቕክን	ዝ-ጸበቕ-ክን	z-sebeq-kn	z-			-kn	
	ዝጸበቕና	ዝ-ጸበቕ-ና	z-sebeq-na	z-			-na	

AnnexFour: Sample Data of Sentiment Sentences

1. ካብ ዝሓለኹዎ ንላዕሊ እዩ
2. ካብትግራይጎበዜዘሓፍተይበጣምደስ በሃላይ
3. ብጣዕሚ ደስ በሃሊይ
4. እዚደርፌሰሚዐናይብሒቑእየደስ ዝብል
5. ብጣዕሚእዩ ደስ ትብሌወለድኺን ዓድኺን ተዘክር
6. ዳቆንጆ በርትዕ
7. ዋው ትክክለኛናይትግራይሙዙቃ!!
8. ደጊምካ ተሪኢኻዮ ዘይስሊቺ
9. አዜያመሃሪት ዝኾነት ፊልሚ
10. ደቂዒይደስ በሃሊ
11. ብጠዒሚመሃሪትዝኮነትፊሌምእያ።
12. ትግራይ ተለቪዥንብጣዕሚእየዝፈትወኩምቀፁለለ
13. ትግራይ ተለቪዥንጸማማትጥራሕእዩ
14. እቶምመራሕትናጸቢብኣተሒሳስባእዩ ዘለዎም።
15. ዝፀንሐ እዩ ግን ሓሪፍ እዩ
16. ዘፅልእ ዝግጅት እዩ
17. ዝቀርቡዝነበሩኩሎምትሕዝቶታትናይትግራትህዝቢቃልሲ፣ባህልንመነባብሮንዝመለከቱስለዝኾ ኑፅቡቅአቀራርባእዩ
18. ናይትግራይቴሌቪንንብዙሕዝተርፎእንኳእንተኾነሓዱሽብምኺኑአብፖለቲካአብማሕበረሰብንኢ ኮኖሚንክልል፣ሀገርንዓለምለኸንፅቡቅእዩነይሩ :: ናይሰብሓይሊብቅዓትከማሓይሽአለዎ:: ፕሮግራማትናብገጠርወሪድኻአብምስራሕክፍተትይረአይነይሩእዮ:: እዚይክመሓየሽአለዎ
19. መብዛሕቱአብፖሎቲካጉዳይዘድሀበስለዝነበረናይኢኮኖሚይማሕበረሰብጉዳያትብዝበለፀትኹረ ትይንእስከምዝነበረ::

   ከምእዉንአብሚድያትንተናህዝቢሙብዛሀትኡግዜተደጋጋሚሰባትመኺኖም።ንገጠርማ/ሰብገይ ርኺዘይምስራሕጉድለታትይረአዩነይሮምእዮም።መኻሊእመንገዲግንተመራፂሚድያእዩነይሩ።

AnnexSix: Questioner Tigrigna and English

ናይ ኮምፒተር ሳይንስ ትምህርት ክፍሊ

ናይ ትግራይ ቴሌቪዥን ተመልከቲ ዝኾኑ ና ዘይኾኑ ዝምላእ መጠይቅ

ናይ መጠይቅ ዓላማ

ብቅድስት ማርያም ዩንቨርስቲ ብኮምፒተር ሳይንስ ትምህርቲ ንምምሃር ንማስተር ዲግሪ ምርመራ ንምምላእ ዝኮውን ናይ ትግርኛ ቋንቋ መሰረት ዝገበረ ዘመናዊ ቴክኖሎጂ ብዝፈቅዶ መሰረት ኮምፒውተራይዝድ ዝኾነ ናይ ትግራይ ቴሌቪዥን ተመልከቲ ሪኢቶ መለለይ ሲስተም ንምስራሕ ግብአት ንምኻን እዩ::

ንዝሀብዎ ንምላሽ አቀዲምና ነመስግን

ሕቶ :- ትግራይ ቴሌቪዥን ተኸታታሊ ዲዮም?

እወ ☐ አይኾንኩን ☐

መልስኹም እወ እንተኾይኑ ስለ አቀራርቡ ዘለኩም ሪኢቶ ይግለፁ

____
____
____
____
____
____
____
____

መልስኹም አይኾንኩን እንተኾይኑ ምኽንያቱ ይግለፁ

____
____
____

Department of Computer Science

Complete questionnaire for Tigray TV viewers and non-viewers

The purpose of the questionnaire

St. Mary's University's Computer Science department is a complement to a master's degree in research in Tigrigna-based Sentimental analysis on a modern Tigrigna-language television viewing system.

Thank you very much for your inquiry

Q: Are you a viewer of Tigray TV?

Yes ☐ No ☐

If your answer is yes, let me know what you think or feel of the TV

____
____
____
____
____
____
____
____

If not, write down why you do not attend

____
____
____
____
____
____
____
____