BUILDING MACHINE LEARNING BASED AMHARIC LANGUAGE INTENT CLASSIFICATION MODEL

BY: LEOUL MESFIN

A THESIS SUBMITTED TO THE DEPARTMENT OF COMPUTING
SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTING

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

OFFICE OF GRADUATE STUDIES
ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

ADAMA, ETHIOPIA
JAN, 2021

BUILDING MACHINE LEARNING BASED AMHARIC LANGUAGE INTENT CLASSIFICATION MODEL

BY: LEOUL MESFIN

ADVISOR: TEKLU URGESA (PH.D.)

A THESIS SUBMITTED TO THE DEPARTMENT OF COMPUTING

SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTING

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

OFFICE OF GRADUATE STUDIES

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

ADAMA, ETHIOPIA

JAN, 2021

APPROVAL OF THE BOARD OF EXAMINERS

All, the undersigned, members of the Board of Examiners of the final open defense by **“Leoul Mesfin”** have read and evaluated his thesis entitled “**Building Machine Learning Based Amharic Language Intent Classification Model**” and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirement of the degree of Masters in Computer Science and Engineering (CSE).

______________________	______________________	______________________
Name of Student	Signature	Date
______________________	______________________	______________________
Advisor	Signature	Date
______________________	______________________	______________________
Chairperson	Signature	Date
______________________	______________________	______________________
Internal Examiner	Signature	Date
______________________	______________________	______________________
External Examiner	Signature	Date
______________________	______________________	______________________
Head of Department	Signature	Date
______________________	______________________	______________________
School Dean	Signature	Date
______________________	______________________	______________________
Post Graduate Dean	Signature	Date

DECLARATION

I hereby declare that this MSc thesis is my original work and has not been presented in any other university and that all sources of materials used for the thesis have been fully acknowledged.

Name: Leoul Mesfin

Signature: ________________

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

Name: Teklu Urgesa (PHD)

Signature: ________________

Date of submission: Jan. 22, 2021

ACKNOWLEDGMENT

First and foremost, I’d like to thank God next to God I’d like to thanks my adviser Teklu Urgesa (Ph.D.). Without his patience and understanding over the years, I would not have made it this far. His mentor ship has helped me not only succeed academically but has also taught me much in my personal life. I have been very lucky to have him as my advisor. I would also like to thank my teacher kibrom Tadesse for his support on each progress and also partner Lidiya Biruk, who has provided more hands-on help and collecting data set, all with a great sense of humor.

I would also like to thanks my parents, for their understanding, love, and support. I'm proud to have such wonderful people not only be my parents but also friends. My brother yishak they have always had my back, and been willing to listen to any rants i could dish out.

CONTENTS

LIST OF TABLE

LIST OF FIGURE

LIST OF ABBREVIATIONS AND ACRONYMS

CNN	Convolutional Neural Network
DL	Deep Learning
ML	Machine Learning
SVM	Support Vector Machine
ANN	Artificial neural network
CONV	Convolution
FN	False negatives
FP	False positives
P	Probability
LSTM	Long short-term memory
BILSTM	Bidirectional Long short-term memory
MNB	Multinational naive Bayes

ABSTRACT

Natural language processing is a science that explores how systems read and interpret people's language. In recent years, social networks have become extremely popular. Via various social media users are far more likely to express their everyday life, ideas or intentions. Intent analysis is an approach used to analyze user generated contents to a way that is important for decision making. This research benefits both company's or service provider and customer in terms of making a wise and effective decision, Major Benefits of Social Media for Businesses Improved customer insights, Better customer service. The experiments are conducted on data that collected from You-tube API during the simplicity of data scraping and filtering features. The aim of these research analyzing Amharic text and extract intentions behind a huge text data. And classify tokens into five classes (positive, negative, suggestion, wish, and question). In addition, investigate the impact of noise input data which misspelled and extra white space could affect proposed model.

Keyword: ***-****Intent Analysis, Sentiment Analysis, NLP, NLU, Opinion Mining, Machine Learning*

CHAPTER ONE

1 INTRODUCTION

1.1 Background of the study

Intention analysis is a computational process that analyzes people’s wants and state of mind by utilizing NLU process. Intention analysis is an open domain, a natural language understanding issue, and it is greatly difficult to handle . Natural Language Understanding (NLU) is a field of Natural Language Processing (NLP) that deals with the conversion of natural language into a semantic representation that a computer can interpret. User intent filtering, intent domain identification, intent parsing, and extraction are some techniques procedure for the general NLU process .

Amharic is written left-to-right in its unique script (inherited from the clerical Ge'ez and shared with Tigrinya) which lacks capitalization and in total has 275 characters (mainly consonant-vowel pairs) . It is the second largest Semitic language in the world (after Arabic) and is spoken as a first or second language by around 40 million people. . Amharic-based NLP researches like opinion mining classify text as positive, negative, or neutral is the most common labels for text data.

Although review of sentiment is beneficial, it is not a complete substitute for reading survey responses. The sum of user-generated data will be a basis for a variety of decision-making processes in various fields. This is because these statistics may have a valuable resource for interpreting people's emotions. Depends what the applications demand or require. For example, if a medical application only requires to test positive or negative for a disease, then binary classification is sufficient, but if the application also requires multi class classification social media platform or E-commerce platforms have further two class to express user felling and also identify user generated intention

Intention Analysis is a computational task that analyzes people’s desires, wishes, and attitudes from user-generated texts. That is why sentiment analyses is not enough to express user intent representation. “ባየሁት ቁጥር የሚያስቀኝ ፊልም” it classifies as positive “በጣም ያማል” this text also has negative content those kinds of classification achieved by sentiment analysis with different approaches and performance. Analyzing user intention in advance and deep analysis further being classified as positive or negative. "አሁን ለምን አቀረብክ?" or “ምን ልበል ?” These kinds of comment wide to classify by common opinion mining positive or negative.

This paper formulates the absence of intent classification model as a problem for Amharic language. This work will redefine the pre-defined classification model flexible and advance classification labels to demonstrate the effect of this work. This paper use machine learning technique to extract and identify the labels. As mentioned making deep analysis will give a more accurate representation of each sentence, for example, one car manufacturer company produce a new model car and promote on the official social media page, the car color is black so on the comment section the client most of them comment “መኪናው ያምራል ነገር ግን ቀይ ቢሆን የተሻለ ነው” Based on this comment, classifying this statement in sentiment analysis can’t represent as positive, negative or neutral. That’s why this work has to take an advance and deep analysis model to specifically and truly addressable representation of user comments in Amharic based intent classification model. Comment represented based on our classification model as **“suggestion”** this label more represents this individual intention comment. When it seen from both sides from the client and service provider, the service provider gets well and full representation of customer insight also client intention can have acknowledged and rated as use full information. On the previous research, most of the work related to this paper is done on sentiment analysis that classifies Amharic text into three basic and well-known classification labels.

Many pieces of research have been done for resource-rich languages, such as English and Arabic. The researchers followed different approaches such as the Machine learning approach . Previous Amharic text classification researches related with entity recognition, opinion mining . This thesis focuses on developing LSTM based intent analysis that studies the effect of intent analyses in Amharic text and investigates misspelled input or test data effect on an intent classification model that could lead vary between wrong class domain” ዛሬ ደስ ብሎኛ” example of misspelled text.

As mentioned earlier, knowing user intent is a very difficult NLP problem to overcome. However, in this work, its range has been condensed and restricted in order to solve it using current supervised machine learning techniques. Technically, as follows, many significant contributions have been made to this work. Develop intent analysis that studies the effect of address user intention using labeled text data, the use of advanced predictive machine learning model (LSTM) for this role was explored. And identify the effect of noise input data effect on classification model.

1.2 Intent analysis

Intent analysis is challenging and advance NLP classification that can classify user intention through different approaches specially addressed social media share information and express opinions in discussions.. Root cause the identification of user-generated content in social media allows to address why users don't want a product or service. However, the recognition of consumer insight has a much more useful piece of knowledge in marketing and customer service terminology . This kind of knowledge helps businesses to tailor their offerings and practices to consumer intentions. This work deals with those forms of speech acts that can be considered 'intentions' according to a standard dictionary meaning of the term 'intention,' i.e. 'purpose or strategy.' This paper focuses in particular on the five types of intention labels .(Cohan-Sujay & Madhulika, 2012; Purohit et al., 2015)

Most businesses use social networking websites on a daily basis to advertise new goods and services and post updates to consumers. On the other hand, consumers have a wide variety of ways to communicate their intentions about goods and services. Both in the science community and in the business world due to the extraordinary gains from marketing and financial prediction, the chance of catching consumer intentions raised increasing interest. The knowledge of consumer intentions will encourage web marketing for companies and advertisers. When applied to product review with well-structured words, syntactic analysis using word and manual models to extract user intentions has proved effective.

To perform intent extraction and investigation from content posts, its preferable utilize a long short–term memory(LSTM) , a deep learning model that has been extensively and effectively used for sentence level NLP tasks such as named object identification, language processing, as long-term dependency information can be implemented. By referring amount of previous studies, define Intent of consumer as “Users intent is an underlying term for a user attempting to communicate his/her intention of information or service need, a task-specific, predefined or latent definition, subject or knowledge-base” (Shang, 2017).

1.3 Statement of the Problem

User intent detection and for many implementations, comprehension has often been a key concern, such as information retrieval, text mining, and a suggestive or recommended framework. Today, internet resources are an integral component of modern human life. (Shang, 2017)

The gap that these research try to identify related with natural language processing specifically based on Amharic language text. There are several researches are existed that related to intent analysis but they implemented on English(Luong et al., 2017) (Nigam et al., 2019)(Cohan-Sujay & Madhulika, 2012)(Akulick & Sayed Mahmoud, 2017)(György, 2017)**.**The absence of Amharic language intent analysis model is the main problem try to tackle. There are only limited number of classification research works on Amharic sentiment analysis due to many reasons like morphological complexity, in-availability of labeled data in the field of NLP, less existence of resources, lack of texts available on the web and inaccessibility.

Based on different language the problem might be wider and bigger also viscera in terms of data structure and resource availability could be make a difference on accuracy and effectiveness of the model, for that specific language.

Mostly representing user intention through opinion mining positive, neutral, negative domains but by considering deep and accurate intention classification of users need this kind of intention classification model and the domain that represent to address user expression well and better form (Wondwossen Mulugeta, 2014) (J. Wang et al., 2015).

The proposed classification model could be ground for deep and specific mining's that related with text format so as gap it will discover and contribute Amharic based intent classification model this will improve the ability to detect the user intention by fetching text (posts) from the individuals on specific idea or product. By gathering the posted comment, it can classify the majority outlook on that specific product.

1.4 Motivation

When I started delving into this research topic, frustrated by issues with text data processing and simulation. I have found some fascinating things, such as methods, tactics and workflows, which can be used to solve different problems. **Analyzing public feedback** or **Improving predictions** to **aid decision-making** is really help full to build different digital frame works on different area from business also for different governmental projects based on digital era. In Ethiopia there are some startup movement on AI systems it really motivates different researchers to contribute their own insights. Understanding something is helpful to make wise and perfect decision. This paper is trying to observe and analyze especially on natural language processing (NLP) area and there is no attainment on Amharic intent analysis so it's become new and contribute basic classification labels on Amharic language. Based on those two reasons impressed on this NLP research topic.

1.5 Objectives

1.5.1 General objective

Deploy Amharic Language intent classification model using machine learning approach.

1.5.2 Specific objective

- ✓ Analysis of the general grammatical structure of Amharic sentences with the aim for identifying and determining intent
- ✓ Build and test Amharic LSTM based intent classifiers model.
- ✓ Analyze developed model could be affect by noise data.
- ✓ Develop a prototype to demonstrate the validity of LSTM model.
- ✓ Build web-based application for interact users to the model.

1.6 Research question

User intent mining will be to directly or indirectly use words, subjects, principles, or other representation approaches to model the desires of the user. It is helpful to better understand user behavior and offers valuable guidelines for the creation of user centric apps, such as customized information retrieval, suggestion, user profile development, etc.

In the context of previous debates, this study lift the concerns as follows.

- ✓ Which possible algorithm to perform a better Amharic intent classification model?
- ✓ What is the effect of miss spelled and extra white space inputs data on proposed model?

1.7 Scope

Intent analysis is a modern and dynamic discipline of study that involves productive analysis and management subjective texts. This research includes the design and development of intent analysis for Amharic Language comments and posts and is limited on textual data. emoji and pictures other regional and foreign languages not considered it mainly focus on specific categorization classification for subjective Amharic posts and comments. Intention polarity limited on the listed five labels, positive, negative, suggestion, question and wish.

1.8 Contribution

This research contribution for the intent analysis on Amharic by classifying intent as **positive, negative, suggestion**, **question**, and **wish** this research paper will have the following contribution

- There is no officially available data set for intent analysis domain this research collects and contribute Amharic based intent classification domain corpus for the next researches.
- Develop a LSTM based intent analysis classification model.
- The reviews collected and preprocessed in this paper will be additional resources for the next research done on Amharic based NLP researches. It will also give research direction regarding Amharic intent Analysis.
- Analyzing noise data could lead miss classification.

1.9 Significance of the Study

The study provides different benefits for different stakeholders and other researchers. After the evaluations of the result, it can be applied in different areas:

- The study opens a door and encourage different researchers in Ethiopia, who have the interest to work on NLP field with different languages.
- Governmental or private companies can also get benefit from the study by to analyze feed backs customer or employee currently the governmental and private companies deploy different digital systems and web site this study help to understand customer and employee intention

- Incorporating Intent Analysis into Business to Improve Customer Experience in terms of enhancing digital marketing knowing customer demand and intention about the gap and supply product is more effective and its help full for both sides.
- Performing detail data visualization and flexible result representation to the user on the web-based application to classify the user intention.

1.10 Thesis Organization

The remainder of this paper is structured as follows. Chapter Two discuss previous research works on both resource-rich and Amharic language and state of the art approaches for intent analysis. Chapter Three discuss methodologies, Amharic language challenges, Data properties. Our proposed approach for intent analysis is elaborated in Chapter Four. The experimental setups, procedures, evaluation metrics, Results and debate in Chapter Five. Last but not least, Chapter Six discusses the conclusion and future research direction on Amharic intent analysis.

CHAPTER TWO

2 LITERATURE REVIEW

This chapter describes the related and empirical literature applicable to the key aspects of this study. The key facets of this research are machine learning systems for processing and classification strategies of natural languages. Especially Ethiopian languages NLP researches and foreign languages named intent, entity-recognition and sentiment analyses researches from foreign and local languages reviewed, state-of-the-art approaches are briefly discussed.

This section briefly reviews previous studies on intention analysis and related topics based on conceptual and related literatures.

Intent analysis is a process of find and extract different user data from different source. This type of work is related to NLP and those different papers they use different mechanism to solve natural language processing problem, so it categorizes in different groups based on the approach they follow to solve the problem.

2.1 Amharic Language

Amharic is written in its own special script from left to right (inherited from the clerical Ge’ez and shared with Tigrinya) lacking in capitalization and with a total of 275 characters (mainly pairs of consonant vows). Ethiopia is a vast country located in the horn of Africa with more than 80 nations and nationalities that have their language and identity. Amharic which it's a working language of the country is spoken by more than half of the population. The language uses its alphabet, ፊደል/ final, inherited from the Geez (Ethiopic) language. (Kelemework, 2013) (Eyassu & Gambäck, 2005) (Sikdar & Gambäck, 2018) (Addis et al., 2018).

TABLE I. SAMPLE ETHIOPIC SCRIPT ALPHABETS

	1st Order (ä)	2nd Order (u)	3rd Order (ī)	4th Order (a)	5th Order (é)	6th Order (i)	7th Order (o)
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
h	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
M	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
s	ሠ	ሡ	ሲ	ሳ	ሴ	ስ	ሶ
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
sh	ሸ	ሹ	ሺ	ሻ	ሼ	ሽ	ሾ
	...	...	...	...	...	...	...
P	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ
V	ቨ	ቩ	ቪ	ቫ	ቬ	ቭ	ቮ

Figure 3.1: sample Ethiopic Script Alphabets (Addis et al., 2018)

2.1.1 Punctuation Marks of Amharic Language

Punctuation marks are conventional signs that help for easy reading and understanding from a written text. The language in Amharic has about ten punctuation signs. However, mostly used signs are the end of sentence called AratNetib (“:"), question mark ("?"), an exclamation of sentence called "ቃል አጋኖ" ("!") and word separator Hulet Neteb(":") (Kelemework, 2013).

2.1.2 Challenges of Amharic Intent Analysis

Applications like information retrieval, text classifications or document filtering could benefit more by the existence and accessibility of basic instruments like stemmer, morphological analyzer, and POS taggers. However, know its little explained about its effect on classification performance or retrieval accuracy due to a little number of researches on the language (Sikdar & Gambäck, 2018). Methods cannot be directly implemented in the Amharic language due to some problems that make Amharic language challenging for NLP applications. The challenges are listed and illustrated as follows.

2.1.2.1 Character Duplication

Some Amharic Language characters have their form but having similar sound, there is no clear cut where to use these characters (Addis et al., 2018). For instance, the In Amharic, the word "ጸሀይ" ('sun') can be represented as ጸሀይ, ጸሐይ, ጸኀይ, ፀሀይ, ፀሐይ, etc. (Kelemework, 2013), there is no clear rule to use ሀ, ሐ or ኀ and ጸ or ፀ by the people use such characters for semantically same words interchangeably. The following table shows the characters that have different form with the same sound (Eyassu & Gambäck, 2005) (Wondwossen Mulugeta, 2014) (Kelemework, 2013)

Sound pattern	Matching Amharic characters
/sə/	ሰ, ሠ
/ɪə/	ጸ, ፀ
/hə/	ሀ, ሃ, ሐ, ሓ, ኀ, ኃ
/iə/	አ, ኣ, ዐ, ዓ

Figure 3.2Example of Character Redundancy (Eyassu & Gambäck, 2005)

2.1.2.2 Absence of Abbreviation Rule

There is no a clear-cut rule in abbreviating Amharic words. For instance, the word "ዶክተር" can be abbreviated as ዶር, ዶ.ር, and ዶ/ር , 'ዓመተምህረት', meaning 'AD', can be abbreviated as ዓም , ዓ.ም., ዓ/ም (Jabbar & Khan, 2015).

2.1.2.3 Spelling Variation

The same word can be written of multiple ways. The word, for instance computer can be written as ኮምፒዩተር or ኮምፒውተር or, 'ሰምቶአል' ('he hears') can be written as Amharic ሰምቶአል, ሰምቷል, ሰምትዋል, etc. Spelling differences may also occur as foreign words are converted into Amharic. For instance, the word 'ቴሌቪዥን' ('television') can be written as ቴሌቭዢን, ቴሌቭዥን, ቴሌቪዥን, etc. (Kelemework, 2013).

2.1.2.4 Lack of Resource

For instance, there is no collected data set for Amharic language analysis and other NLP application fields this makes data collection and pre-processing more challenging tasks. There are not adequately prepared and structured corpora available to be used for research purposes.

2.2 Conceptual literature's

2.2.1 Machine Learning based Intent Analysis

There are many intent analysis and sentiment analysis researches use machine learning approach to fulfill the researches of different languages. This study observe and analyze those machine learning algorithms, most of they are redundant and well known and effective machine learning models, like Recurrent neural network(RNN), Support Vector Machine (SVM), Naïve Bayes (NB)(Agarwal & Sureka, 2017) (Shang, 2017) (György, 2017) (Wondwossen Mulugeta, 2014). Some of the researches merged two or more machine learning models to demonstrate the effectiveness of the sequential segmentation, and also different algorithms to evaluate their method.

Most of the researches based on the three basic machine learning models, supervised, unsupervised and semi supervised machine learning models. Supervised machine learning models uses labeled data set to train and evaluate their models effectively. unsupervised machine learning models used unlabeled data set to evaluate the model and last one is semi supervised machine learning models this kind of machine learning model are using both label and unlabeled data to evaluate the model. (Luong et al., 2017) (György, 2017) (Wondwossen Mulugeta, 2014) (Shang, 2017) (J. Wang et al., 2015), by observing statistical review it can be said that all of them follow supervised machine learning models to classify and to build a well and effective model for both sentiment analysis and intent analysis researches.

Paper title	Data set	Model used	Result	Limitation
Intent Detection through Text Mining and Analysis (Akulick & Sayed Mahmoud, 2017)	358 review posts from banking institutions and 100 review posts from Amazon.	N-grams, Part Of Speech(POS),Support Vector Machine(SVM)	N-grams have better accuracy than POS and SVM	Algorithms should be tested on a larger sample of data with reviews and posts regarding multiple industries.
Intent Extraction from Social Media Texts Using Sequential	Collected totally 712 posts for real estate and 1500 posts for	LSTM–CRF	LSTM-CRF (Char+Pre+Drop) have better accuracy	This paper mainly considers extract intent information from

Segmentation and Deep Learning Models (Luong et al., 2017)	cosmetic & beauty.			online social media texts. Intent classification on the domains not consider.
Intention Analysis for Sales, Marketing and Customer Service	Only 4113 sentences belonged to categories related to opinion (praise, criticize and compare)	Naive Bayes classification, maximum entropy classification, and support vector machine classification	Maximum entropy have better accuracy than others	performed effectively on short texts long text not considered during the model they apply

Table 2.1 Table Literature collection on intent analysis papers

2.2.2 Deep Learning based Intent Analysis

Neural networks have recently developed a diverse approach for a range activity, such as image analysis, price forecasts, natural language processing. Neural networks for speech recognition, machine translation, description of text, identification of the object, texts generation and many other tasks are used during natural language processing. Deep learning approaches typically need a large number of training data for the particular model to be successful and performing better(Balodis & Deksne, 2019).

Several studies on intention recognition using deep neural network technique (Luong et al., 2017). However, by classifying or inferring intention from the spoken language using semantically enhanced word embedding, they directly target queries.

Deep learning model that has been widely and successfully used for sentence-level natural language processing tasks such as entity recognition, language modeling since they can integrate long-term dependency information. (Luong et al., 2017)(Purohit et al., 2015). There are several deep neural networks such as CNN (Convolutional Neural Network) CNN is one of the deep learning algorithms and it integrates the automatic feature extraction, feature selection and pattern classification in a single architecture., RNN (Recurrent Neural Network), Long Short-term memory (LSTM), Bidirectional Long Short-term Memory (BI-LSTM), most of those algorithms well known and effective they apply on this researches (Nigam et al., 2019) (Sikdar & Gambäck, 2018) (Addis et al., 2018) (Shang, 2017).

Deep neural networks have demonstrated that multiple language processing functions such as part-speech tagging, emotion recognition and name entity recognition (NER) have been successfully overcome. Another version of LSTM , Bi-directional LSTM deep learning neural network model better performance was achieved after merging the different features vector with word embedding's generated with word2vec (Sikdar & Gambäck, 2018)

Identification of query intention as a multi classification task with features of extracted query vector representations. It indicates the usefulness of indifferent experimental settings.

The findings show that the query vectors derived from the CNN models can be detected easily and are identical compared with bag-of-word features. In addition, they are automatically learned from training data, they far outperform the carefully designed rule-based baseline. (Hashemi et al., 2016)

Paper title	Data set	Model used	Result	Limitation
Intent Detection and Slots Prompt in a Closed Domain Chat bot (Nigam et al., 2019)	vocabulary size is 4229 words, Training Set Size 10980 , Validation Set Size 2353, Testing Set Size 2354	For subcategories using RNNs, and find entities using their Named Entity Taggers in multiple stages	The intent accuracy in on their model came out to be 75.07% and the accuracy of the baseline model turned out to be 63.97%. For slot-tagging task, the F1 score was 82.24% for our model and 42.71% for the baseline model.	Base model they use traditional RNN. And this model not test on other public data sets.

Named Entity Recognition for Amharic Using Deep Learning (Sikdar & Gambäck, 2018)	SAY corpus is roughly half the size of the WALTA corpus (having 4,237 sentences and 109,676 tokens), but all of it is annotated with named entities (with 5,480 of the tokens being annotated as names),	**Model-1** uses randomly initialize word vectors to train the LSTM network **Model-2** uses word2vec to generate a word vector for each unique word and train the LSTM network classifier on these word vectors. **Model-3** takes as input word embedding's developed using word2vec along with the language independent features (suffix and prefix, POS, frequency and digit-check).	average precision (77.2%) and recall (63.4%), for an overall 69.7% F1-score. **Model-3** is better result from others	The size of the data set and LSTM architectural configuration is reduce the performance of the proposed model

Studies on User Intent Analysis and Mining (Shang, 2017)	Manually collect possible queries from crowdsourcing platform. Basically, generally six categories of user intent, which are Call, Command, Music, Message, Navigation and Parking.	apply a classifier using convolutional neural network to classify user's intent, and it outperforms several baseline models. For intent attribute recognition, They use Conditional Random Fields with carefully feature engineering	Methods Accuracy KNN Classifier 0.5556 Decision Tree 0.8889 Random Forest 0.8519 Nave Bayes 0.5926 Logistic Regression 0.7037 SVM 0.7407 Proposed Method 0.9629	Model works on data with timestamps and response categories to predict the user's future behaviors, without considering the content of responses.
Query Intent Detection using Convolutional Neural Networks (Hashemi et al., 2016)	#of queries 10,000 #of low-level intent classes 125 #of high-level intent classes 14	They Using CNN model for standard classification measures and compare it with the contrastive baselines. And for the clustering task, they use the extracted query representation as features to group queries.	Query intent detection results with 125 low-level intent classes CNN 81.6 its better than others Also for the Query intent detection results with 14 high-level intent classes CNN 90.3 CNN have way more effective than other in terms of both features,	They are not use better memory management algorithm for query vectors such as recurrent neural networks with long short-term memory networks (LSTM) cells.

Table 2.2Table Literature collection on deep learning intent analysis papers

2.2.3 Rule Base Intent Analysis

There is a lack of existing rule base intent analysis research papers to show the effectiveness of this model, compare and list out the best method is difficult so review found some and particularly try to review. Arijit De uses rule based intent identification the construction of the basic rule is counterproductive to the efficiency of the short query intention recognition method. The rule base is a mapping of tag combinations to all domains with some weights. (De & Kopparapu, 2010) To build the rule-base, they used two distinct methods, the first one

2.2.3.1 System Generated or Semi Supervised Rule Base

The approach they have developed is based on the assumption that if such tags appear in a given domain, they are more likely to be tag-combined. It's less likely to be a human mistake. (De & Kopparapu, 2010)

2.2.3.2 Hand Crafted Rule Base

The system can work better using a hand designed rule-base. Handcrafts, however, require tedious human interaction and are vulnerable to human error. if the domain value is 1 while all other domains value 0. Domain confidences are not absolute in system-generated rule bases. For each domain, the trust of the domain can differ between 0 and 1. (De & Kopparapu, 2010)

2.2.4 Literature review on Sentiment Analysis

Sentiment analysis attracts growing interest from natural language processing experts, from organizations seeking to track consumers' views on their services and goods, and from the general public in order to gain a viewpoint on subjects of concern. The purpose of Sentiment Analysis or Opinion Mining is to decide the attitude of a speaker, writer or other subject to a particular subject or case. Its applicable for different world wide languages we can use different machine learning, deep learning, lexicon, rule based models to evaluate the implementation for some specific language and domain of each different reviews.

Maha Heikal wodk a sentiment analysis based on Arabic Language this paper, used an ensemble model, the combination of to forecast the sentiment of Arabic tweets, Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM) models. The algorithm they used reaches 64.46 percent of the F1 score, which beats 53.6 percent of the state-of-the-art F1 model for the Arabic Opinion Tweets Dataset (ASTD). Consisting of 10,000 tweets spread across 4 groups (positive, negative, neutral and objective). They boost the previous best model in this model (RNTN). This model operates on the model for the study of sentiments in Arabic. No features or dynamic components are used in the proposed model to extract unique features.

There are standard techniques to classification of sentiments such as machine learning, deep learning of lexicon and hybrid methods that combine machine learning with lexicon to improve the accuracy of classification of senses. (Heikal et al., 2018) (Jurek et al., 2015) (Mihret & Atinaf, 2019)

2.2.5 Lexicon based Sentiment Analysis

A piece of text message is described as a bag of words based on lexicons. Following this message representation, all positive or negative terms or phrases in the message are given sentiment values from the dictionary. To make the final prediction about the general feeling for the letter, a mixture feature such as total or average is used. (Jurek et al., 2015)

Lexicon based sentiment analysis is the unsupervised technique. This technique involves collecting and building a set of lexicon words that have positive and negative polarities. The collected combine with some rules of the language to determine the sentiment of a given sentence.

Anna Jurek proposed A new sentiment analysis algorithm based on lexicons, which has been designed with the main focus on analyzing Twitter content in real time. The algorithm consists of two main components, namely sentiment normalization and evidence-based combination function, which have been used to estimate the strength of the feeling rather than positive/negative mark and help the process of classification of mixed sentiment and enhance the algorithm's efficiency in cases when a mixed feeling occurs in a message. (Jurek et al., 2015)

They compared performances of five sentiment analysis algorithms, namely algorithm:

L Technique of general lexicon focused on negation and intensification. Instead, the value of both positive and negative terms in a sentence is easily summed up. In addition, it classifies messages as positive, negative or neutral whether the value received is positive, negative or equal to zero.

LN Performs an analysis in the same direction as L, but instead of summing it refers to adding the meanings of positive and negative emotion.

LNS executes LN for each phrase in the message and measures the total positive/negative sentiment of a phrase as an average of the values for all phrases in the message.

LNW It functions like an LN, but it uses the evidence-based function to identify the message as positive or negative in the event of mixed opinion within a message.

LNWS Runs LNW inside a message for each statement. For each of the words, the procedure is repeated. The final sentiment is determined for all the sentences as an average of the values received. (Jurek et al., 2015)

2.2.6 Rule-based Sentiment Analysis

Most researches that ruled their model to get that the specific objective on the labeled data. that progressively apply different rule to improve the accuracy those rules might be effective on some researches also will not for others so they apply different rule for specific objectives model and domains.

This rule-based approach is difficult because it requires analyses and deep understanding for each rule to be effective and the rules developed by humans so it’s more consume time to develop each rules.

The rule-based approach to sentiment analysis facilitates a detailed analysis of the review's opinion material, ensuring that not only the general sentiment of the review can be identified, but also each clause's separate sentiment. This offers ample knowledge for a person or incident to distinguish various positive and negative points. Therefore, such an overview offers more than a general feeling about the review or the rate of the social networks of a certain entity or individual. (Romanyshyn, 2013) Sujata Rani proposed adverb-adjective combination as a feature to improve the earlier only adjectives as a feature (Science, 2014).

2.2.7 Machine learning based Sentiment Analysis

As mentioned earlier in the literature review on intent analysis models they use the most common and well known algorithms like Nave Bayes, Support Vector Machine (SVM) and Maximum Entropy. Other features include parts of speech (POS) tags, n-grams, bi-grams, unigrams and word bags-of-words. Mostly they use commonly for both intent and sentiment analysis.

2.2.8 Deep Learning Sentiment Analysis

Deep learning has emerged as an effective technique in machine learning that studies several layers of data representations or features and delivers effects of state-of-the-art prediction. Deep learning is popularly used in the sentiment analysis of recent years, in addition to the growth of deep learning in several diverse applications. (Alemu, 2018) (Mihret & Atinaf, 2019)

Abdalraouf Hassan propose ConvLstm, a neural network architecture utilizing Long Short-Term Memory (LSTM) Convolutional Neural Network (CNN), merged these two algorithms in addition to pre-trained word vectors to minimize the loss of detailed local information and capture long-term sentence sequence dependencies. (Hassan & Mahmood, 2017)

Wenkuan Li, Peiyu Liu also propose sentiment feature enhanced deep neural network (SDNN). They are introducing a novel method of sentiment attention mechanism to select a key sentiment-word-relevant context word. Second, by combining bidirectional gated recurring units and a convolutional neural network, they try to enhance the neural network in order to extract sequential correlation information and local text. (W. Li et al., 2019)

2.3 Related works

In this section review sentiment analysis, Amharic based approach and some papers apply a different Ethiopian language called agew and they use different models, extraction method, and features. To achieve the highest result. She uses a face book to implement model and the domains positive negative, neutral.

Most researches on sentiment analysis is based on social media API for evaluate and use as a data set such as Face book, tweeter, YouTube. yodit teshome (TESHOME, 2019) also work on sentence level classification on Amharic language. This research is conducted based on data gathered from social media Facebook without restriction of domain and also comment and replay comment classify as positive, negative and neutral.

The model is tested extract annotated data and tested using 800 experimental annotated data. The results of the experiment show the performance of the method has 92% precision and 82%

recall for positive class, for negative class 94% precision and 96% recall whereas 90 % recall and 90 % precision for neutral is achieved in the determination of opinion word. (TESHOME, 2019)

Melese Mihret Present the SA model to the Awngi(Agew) the language is spoken in Ethiopia (Mihret & Atinaf, 2019), Use a supervised approach to machine learning. By collecting some 1500 posts from on-line sources, they build corpus. They tried to improve the calculation procedure for selecting and weaving features. They proposed a more suitable SA algorithm for extracts of features called CHI and weight calculations known as TF IDF.

The main focus on this research is the agewi language music review using machine learning technique to resolve the awngi music review. to give the polarities of positive negative and neutral for the data they collected positive and negative data collection and calculate sentiment score for each review. They also used well-known Vector Machines (SVM), Naïve Bayes (NB) and MxEn machine learning algorithms for the model evaluation.

Each post in the test set will be used one by one as an input post and its polarity will be returned. Because of the three learning algorithms was presented with sentiment words (Unigrams and Bigrams) as a feature. They achieved an average accuracy of 75% NB, 70% MxEn and 79 % SVM performances they obtained. (Mihret & Atinaf, 2019)

A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts is developed by Wondwossen Philemon and Wondwossen Mulugata (Wondwossen Mulugeta, 2014) using a supervised machine learning approach. The Author collected 608 social media product and marketing news from Facebook, Twitter, Dire Tube and Ethiopian reporter websites. The developed model has Accuracy of 43.6%, 44.3% and 39.5% for uni-gram, bi-gram, and hybrid features respectively. The developed model has performance limitation due to less number of training data.

CNN based Amharic sentiment analysis (Alemu, 2018) is developed by Yeshiwas Getachew and Abebe Alemu. The Author collected 1602 reviews from Fana broadcasting Facebook page using face book API. The model produces different results depending on the amount of test and rain-data size. Both models use stemming as part of preprocessing.

Literature evaluations for other languages have indicated that stemming does not have a positive effect on the results of the sentiment analysis model. Hence, it is necessary to investigate the effect of adding stemming as part of preprocessing on Amharic sentiment analysis The corpus used by both models is annotated by humans which takes much developing time and results in disagreement between annotators.

Selama G/Meskel (Gebremeskel, 2010) has developed a Sentiment mining model for opinionated Amharic texts. In this research, the author used a term counting approach to determine the sentiment of comments by checking each word from the prepared lexicon. The author developed 411 positives,544 negative generals, domain-specific lexicon terms and contextual valence shifter terms. In the paper, each positive lexicon is tagged with +2 value and each negative lexicons are Tagged with -2 value but the final polarity of the review is determined by considering the existence of valance shifter, intensifiers, and negation terms. Even if the paper prepares opinion terms, the prepared opinion terms are not sufficient because inflected Amharic sentiment words are not handled in this research work. Besides, the paper only considers textual labeled lexicons for determining the polarity of reviews.

Emojis which are increasingly used in social media contents to express one's feeling has not been handled. The other work on the Amharic language is a feature level sentiment analysis by Tulu Tilahun entitled “Linguistic Localization of Opinion Mining from Amharic Blogs” (Tilahun, 2014). The Author collected 484 comments from a hotel, university, and hospital as 578 negative and 423 positive words. The Author extracted features using some rules and determine the polarity of the review by considering the polarity of adjacent words of the feature word.

The effectiveness of the system depends on the existence of valance shifters and the existence of adjacent left and right adjective features. The other research paper is "Trend Detection and sentiment Analysis for social media” (Arega, 2019) by Yosef Arega which uses Posts from the Facebook page of Dire Tube, EthioTube, VOA Amharic,mereja.com, and Yehabesha. The model developed by Yosef uses Rules and Lexicons used by (Tilahun, 2014).

TF-IDF and Latent Dirichlet Allocation approaches are combined for the trend detection system. The model gives different results depending on the topic selected and the existence of valance shifter Terms. All works consider only textual words for determining the polarity of the review.

However, Emoji's are increasingly used in social media content to express one's feelings on a certain event. Hence, a framework that takes must be built Emoji's for labeling purpose. The work by Tulu Tilahun used adjectives only as sentiment words to determine the opinions of the review sentences, but sentiment words are not only adjectives but also include adverbs, verbs, and nouns. In addition to this, the sentiment words are not sufficient.

Paper title	Data set	Model used	Result	Limitation
Improved lexicon-based sentiment analysis for social media analytic(Jurek et al., 2015)	They use two main data sets first one is IMDB 25,000 movie reviews including 12,500 positive and 12,500 negative	LN with LNS and LNW with LNWS	L 67.5 LN 51.4 LNS 71 LNW 60 LNWS 74.2	presented a approach is a lexicon-based sentiment analysis during time and labeling its more sophisticated.
Deep Learning Approach for Sentiment Analysis of Short Texts(Hassan & Mahmood, 2017)	**SSTb** Train 8544 Dev 1101 Test 2210 **IMDB** Train 25k Dev 198.8 Test 25k	ConvLstm	Accuracy on SSTB data set for fine grained and binary prediction at the sentence level Fine-Grained Binary 47.5 88.3%	They mainly consider short text representation of user data

An Improved Approach for Text Sentiment Classification Based on a Deep Neural Network via a Sentiment Attention Mechanism(W. Li et al., 2019)	movie review (MR) data set has 5331 negative samples and 5331 positive	SDNN	Accuracy F1 83.7 84.9	Need improvement on the data set with fine grained sentiment labels. And analyze the actual emotions of specific users might not consider.

Table 2.3 Table summarized sentiment analysis literature

Paper title	Data set	Model used	Result	Limitation
Sentence level opinion mining for Amharic language (TESHOME, 2019)	They conduct from Facebook API total amount of 800. 600 reviews are politics, 100 reviews are drama, 50 reviews music, 50 reviews guest program	Rule-Based Model	0.92 precision, 0.82 recall and 0.87 F-Measure in for positive class, for negative class 0.94 precision, 0.96 recall 0.94 and F-Measure and for neutral 0.9 precision, 0.9 recall, and 0.9 F-Measure	Mainly on this work they apply an Rule-Based method during the current approaches its preferable use machine-learning and deep-learning models
Sentiment Analysis Model for Opinionated Awngi Text: Case of Music	They collecting around 1500 posts from online sources	They present a SA model for the Awngi language	Average accuracy of 75% NB, 70% MxEn and 79 % SVM performances they obtained	The data cleanness and small size of data and lack of memory management in

Reviews(Mihret & Atinaf, 2019)				terms of selecting algorithms.
A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts(Wondwossen Mulugeta, 2014)	They collecting around 600 posts from online sources	Naïve Bayes	performance accuracy of 43.6%, 44.3% and 39.5% for unigram, bigram and hybrid language models	Multi scale representation still further data limitation and quality its consider only the class of traditional label's as more positive
Trend detection and sentiment analysis in Amharic social media.(Arega, 2019)	Face book total data set 42,762	T-LDA for order the topics T-LDA is a combination of TF-IDF and LDA	F-score 89% for positive, 79% for negative sentiments 99% for neutral sentiments	Detecting the current popular topics on current situation and identifying most popular issue on social media. But not detect user intention

Table 2.4 Table summarized Amharic sentiment analysis literature

2.4 Recent Literature

In this section try to review, recently literature on opinion mining that related with covid-19 This paper analyzes Twitter posts for the disease since January 2020 in statistical form. There have been two kinds of observational research. The first one involves word frequency, and the second the emotions of each tweet (Rajput et al., 2020). The outcomes have been approved by Sum of Square Error (SSE), R2 and Root Mean Square Error (RMSE). High estimations of R2 and low estimations of SSE and RMSE lay the justification for the decency of this model.(Rajput et al., 2020)

Dr. Murthy Study on sentiment analysis using Long Short-Term Memory (LSTM) for text evaluation. Recently, neural networks have had strong results in text recognition with the potential to manage vast volumes of information. LSTM networks, in particular. DL techniques such as LSTM demonstrate improved sentiment classification efficiency with 85 percent accuracy when more training data is available. The cumulative number of data for Amazon 50,000 (from outlets such as Amazon and IMDB) and also for the use of IMDB 50,000 data to train and validate the model(Dr. G. S. N. Murthy et al., 2020).

XiaodiWang work on The study of aspect-level sentiment is to classify the polarity of sentiment of a given target word in sentences. By integrating multilevel interactive bidirectional Gated Recurrent Unit (GRU), focus processes, and position characteristics, this analysis proposes an improved classification model (MI-biGRU) (X. Wang et al., 2020)

Hiwot Wonago apply the concept of sentiment analysis on Amharic text on social media and presents a comparative study on machine learning algorithms. The created social media content filtering system has been tested on Facebook posts of each class, and it has been observed that SVM with word2vec. In this paper, they proposed Amharic text sentiment-based social media content filtering approach for the filtering or hiding of offensive/in appropriate content. The classifier that they developed based on Word2Vec with SVM gives us accuracy (0.72) when applied to a sample of 4 classes. (Kululo, 2020)

Summary

In this chapter, textual information such as, opinion mining, and intention analyses are explained. Levels of sentiment mining, intent analysis main tasks and components of intent analysis are also described in this chapter. In addition, the different techniques of sentiment and intent mining’s that Machine learning, deep learning, analysis of natural language, lexicon based and rule based are discussed. The general steps in lexicon based sentiment mining are highlighted where the high level steps are: text collection, pre-processing, polarity words detection, weight assign and propagation, and classification. The different approaches for building an intent lexicon that includes manual approach and corpus based approach are also described in this chapter. Finally, the basic rules types for both sentiment and intent analysis are also described.

Generally, for this research both local and international researches reviewed and observe different methodologies, techniques, applications to achieve natural language processing. Based on our observation data is most required and deiced the performance of the model. Collect data is big issue on different NLP topics. Thus this give lesson about different data collection technique, preprocessing, data balance, cause over-fitting and under-fitting and avoid technique, different classification algorithms with their features are gather from this literature review to achieve our objective.

CHAPTER THREE
3 RESEARCH METHODOLOGY

3.1 Overview

The key goal of this chapter is built the methodology to carry out the study. This study followed experimental research design and applied different methods starting from problem identification, define the objectives for a solution, Design and prototype development, testing and evaluating the performance. This is the procedure used in evaluating the various predictive machine learning approach and algorithms using the imported custom data sets from social media. The chapter also deals with the Explanation of the software tool used to apply various algorithms to data preprocessing, feature selection classification, and prediction. Also observation and analysis of Amharic language raw data, back ground and limitation case study natural language processing outlying done.

This research studies specifically the gap that refrained user's noticeable intention text based posts attached to some channels or groups. The main issue is identifying the digital text representation, using customer intention comment as a data, analyze and detect using our frame work and identify Amharic based user intention labels well classified and structured representation for individual user intention. Finally gives us the format easily identify also organized output.

Understanding the intentions of internet consumers has now become a critical necessity in many various business fields, such as manufacturing, finance/banking, real estate, travel, culture, sport, e-commerce and online marketing.(Luong et al., 2017) data analyses and classification using some framework to enhance the customer participation on modify better service experience details on the user's activity help out quality and digital environment is enough to communicate and interchange the intentions in an easily recognizable format.

In this research iteration requires several data collection instruments or strategies that range in complexity, design, etc. interpretation, and administration; each technique is suitable to get certain type of information (Rahi, 2017). Data collection techniques such as interviews, questionnaire, direct observation, and document analysis have been used in different studies.

Those common approach help us to collect sufficient and effective data sources. Mostly collected from public service provider, organizations, digital platforms, are the popular once. (Rahi, 2017) (Wondwossen Mulugeta, 2014) (Mihret & Atinaf, 2019)

For this study data collected from YouTube API using YouTube comment scraper to collect structured and full text based Amharic comment that posted from each individuals including filtering feature to filter out unnecessary strings from comment.

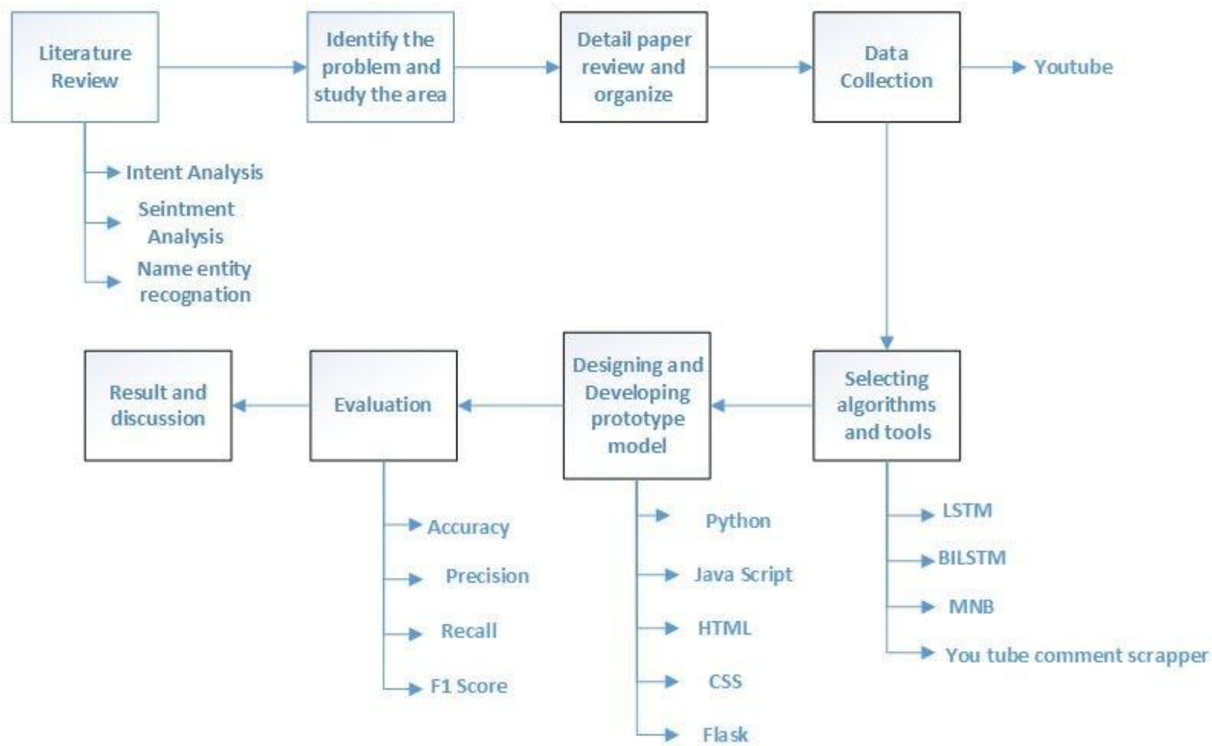


Figure 3.3 Fig: Research Methodology

3.2 Research Design

The research design relates to the overall methodology to coherently and logically combine the different components of the analysis. In order to meet the general and specific aims of the research, and to solve the problem by answering the research questions, different methods are used. After deciding the area of interest, the first step was performing literature review to get previous works intent analysis on the area of different languages and methods for both intent and sentiment analysis study.

3.2.1 Understanding the Problem domain

Currently, various Amharic text lexicon data are not available, organized and support for NLP. The most challenging of this study is filtered selective raw data in form of proposed model, which could be effective and answer the research questions. To solve these problems, the requirement or data collection was the basic step to design the proposed framework after identifying the problem. In this study, the required data and user needs were collected using a scrapping user comment from YouTube API.

3.2.2 Understanding the data

Data interpretation involves a closer look at the data. This move involves gathering sample data and determining which data depending on form and scale would be necessary. In terms of the models it attempts to illustrate for this analysis, and check the usefulness of the results required.

Data needs to be tested for completeness, redundancy, missing values, plausibility of attribute values. By the conclusion of the data understanding process, its better if the observations made during the project understanding phase are justified in terms of representativeness, formativeness, data consistency and the existence or absence of external variables.

3.2.3 Preparation of the data

Good data planning makes it easier to manipulate raw data, which is often unstructured and chaotic, into a more organized and usable form that is ready for further study, for accurate analysis, limits errors and inaccuracy that can occur to data during processing. The entire planning process consists of a number of main operations (or tasks) including data profiling, cleaning, integration, and transformation, and makes all processed data more formal and easy to achieve prepared model targets.

One of the main advantages of establishing a systematic method for data preparation is that consumers can waste less time identifying and structuring their records. This is the main step on which the progress of the whole method of exploration of information depends; it typically absorbs about half of the entire effort of study. It can entail data sampling, running, data cleaning, such as testing the completeness of data records, eliminating or correcting noise, etc. meeting clear input criteria to be included in our model as expected.

3.2.4 Data Source

YouTube is the world's most popular streaming video website, with people viewing video every month for 4 billion of hours. YouTube launched in February 2005 and was founded by Chad Hurley, Steve Chen and Jawed Karim, who called it "YouTube."

People began to create a video-sharing website via the YouTube network on which people could upload, download, and watch videos. (Alias et al., 2013) The data set was collected from YouTube API, user comment on different social media including you tube have user comment in different formats such as Text , Emoji's , GIF ,Pictures .during the amount of user and the influence on digital marketing by integrating payment system for user its popular to promotion and people use to publish and advertising product output. During those features and data filtering features of you tube comment scrapper tools its preferable to apply our data collection process on you tube regarding Facebook, Tweeter. Different standard guidelines and general practitioners. The data collected from YouTube API. To specify the domain area, data on intent analysis. The type of data set collected from YouTube API using you tube comment scrapper as a tool.

3.2.5 Data Collection Techniques

In a variety of respects, analysis varies, but they share certain similarities. The need to gather data is one of the common elements. This study dives and extracts data about videos across **YouTube** channels. With a URL or search term as the input, you can get video information, including channel name, likes, number of views, and number of subscribers. The results are provided in several formats, such as XLS, JSON, HTML.

3.2.6 Software Tools

Mainly this research use python, is a high-level programming language that is interpreted by a python interpreter to deliver output. There are many features that makes python the best choice of language for machine learning and Artificial intelligence. textual analytics It is open source and efficient in solving machine learning criteria such as the handling of massive data sets and statistical computing for user interface apply a flask frame work with JS, HTML and CSS.

3.3 Framework Design

The design of the intent analysis model focused on the elements or components from the listed labels to achieve the goals under the given constraints and limitations. To design the proposed model, reference architectures, models, different NLP architecture on deep learning or machine learning based neural network resources were used. In this section Amharic language and LSTM networks identified and analyzed.

3.4 Word Representations Approaches

The aim of language modeling is to learn a common probability of a given word series occurring in texts. The preprocessed data should be changed to a format that is suitable to the neural network classier. Mainly, there are most common-word representation approaches; like bag of words and word embedding.

3.4.1 Bag of Words Model

The bag-of-words model is very easy to understand and execute and provides a lot of versatility to configure your unique text details., words in this process Described by dense vectors, where the vector represents the projection of the term onto a continuous vector space, takes raw data as input and counts the Amount of incidences of each word in document as output. In Bag of words, word lists are paired with their amount of occurrence per document (György, 2017) (Alfina et al., 2018). Bag of words may use Unigram, Bigram, Trigram or N-gram for creating a vocabulary of a document that is used to classify the document (Krawczyk, 2016).

3.4.2 Word embedding

Word embedding's they are essentially a type of word representation that bridges the human comprehension of language to that of a computer. Word embedding is a dispersed representation of text in an n-dimensional space. These are important for the resolution of most NLP problems. (Hashemi et al., 2016) Word Embedding often referred to as distributed semantic model or distributed represented or semantic vector space or vector space model. When you read these names, you come across the word semantic, which means that related words can be categorized together. (Sikdar & Gambäck, 2018) There are well known word embedding algorithms like Word2Vec. (Sikdar & Gambäck, 2018) (Hashemi et al., 2016)

3.4.3 How Word2Vec Works?

For improved word representation, Word2vec is the technique/model to build word integration. It catches several exact syntactic and semantic words. The idea of word2vec method is that appear in similar ways have related implications. It takes feedback from large corpora and produces a word vector for each word. embedding's: continuous-bags-of-words (CBOW) and skip-gram models. (Sikdar & Gambäck, 2018)

In the representation of uncommon terms or phrases, the Skip Gram model is more accurate and works very well with a limited amount of details. To create a model such that the word in the middle is given "jumped", the model would be able to predict or produce the terms surrounding it. "The", "cat", "over", "the", "puddle". Here, the word "jumped" is considered the context. This sort of model called a Skip Gram model. (Sikdar & Gambäck, 2018) (Hashemi et al., 2016) Every word has been trained against the meaning in the CBOW template. One way is to treat {"The", "cat", 'over", "the', "puddle"} Be able to forecast or produce the central word as a context and from these words "jumped". This type of model is called the Continuous Bag of Words (CBOW) Model.

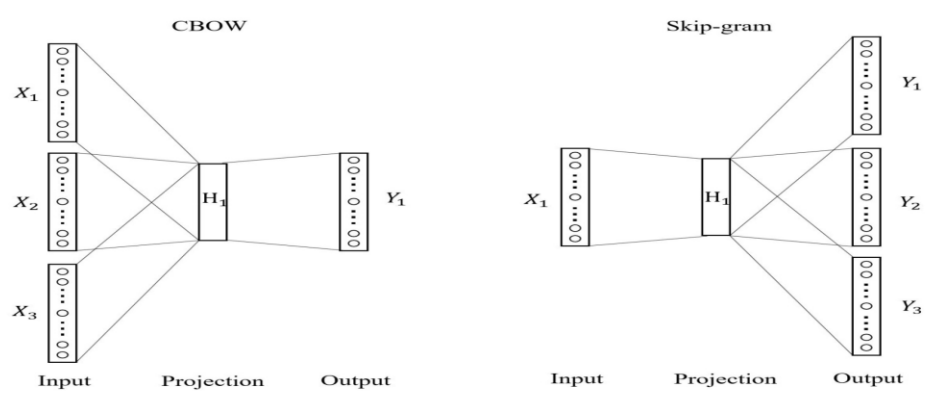


Figure 3.4. Skip Gram model CBOW model (Argaw, 2019)

Depending on the amount of training data Skip-gram models are efficient for a limited number of training details whereas CBOW is efficient with a huge volume of data for training.

3.5 How Long Short-Term Memory (LSTM) Works

3.5.1 Introduction of ANN (artificial neural network)

Artificial Neural Networks (ANNs) are machine simulation tools that have recently appeared and are commonly recognized in many fields for modeling complex real-world problems. ANNs can be characterized as structures consisting of densely interconnected adaptive basic processing elements (called artificial neurons or nodes) capable of performing massively parallel data computing. Basically, the ANN was influenced by the human nervous system. It is based on hypotheses of vast interconnection and parallel processing architecture of the biological system human neurons have various types of phases, method, even stages of producing certain observable output, beginning from observing and extracting information through the neural process and eventually determining what the output will be. An ANN model is a data-driven mathematical model that has the potential to solve problems with machine learning neurons. (Le et al., 2019) (Fischer & Krauss, 2018)

Artificial neural networks (ANNs) are relatively modern computing methods that have been commonly used to solve many complex real-world problems. The attractiveness of ANNs derives from their exceptional information processing characteristics, which are primarily related to non-linearity, high parallelism, fault and noise resistance, and learning and generalization capacities. ANN learning is done iteratively, as the network provides samples of instruction, close to the way to gain experience. Most researches mentioned ANN to properly learn If it is capable of managing.

1) Imprecise, blurry, noisy and probabilistic information without significant adverse effects on the level of response.
2) The mission it had learned from unfamiliar ones was generalized.

A special aspect belonging to intelligent systems, biological or otherwise, is the capacity to learn. Training is used in artificial systems as the method of modifying the system's internal representation in response to external stimuli so that it can execute a specific task, there are three key layers in the typical form of ANNs, namely: **an input layer, a hidden layer, and an output layer**. In an ANN, the data layers for input and output are practically independent of each other. Very frequently, it is apparent that one or more hidden layers are often interconnected by weight matrices, biases, and multiple activation functions between the input and output layers.(Le et al., 2019) ANN models have their own limitations like sequential data problems, time series problem on text competition problem ANN models predict the next word is related to the positon this could be on the word or sentences in the previous position.

Based on this ANN model drawback recurrent neural network introduces to maintain the model RNN is also a class of ANN (Le et al., 2019)

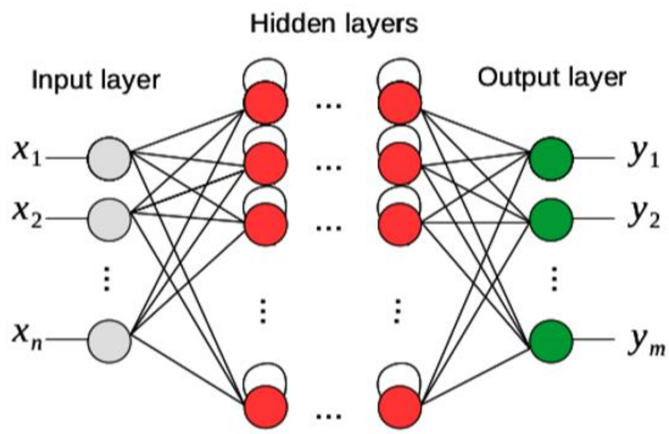


Figure 3.5: Typical Artificial Neural Network (ANN) architecture (Azzouni & Pujolle, 2017)

3.6 Recurrent Neural Network (RNN)

Application of machine learning have a lot of attraction in last few years there are a couple of big categories that able to magnify and being popular specially identifying pictures and sequence to sequence translation that could be speech to text or one language to another, most of the former are done in conventional neural network (CNN) and most of the latter are done In the recurrent neural network(RNN) especially long-term memory(LSTM) for this analysis, also this research try to deploy the LSTM neural network.

Neural network-based approaches have achieved a significant range of natural language workflows. Recently incredible improvement on different problems such as recognition of expression, language modeling, translation, captioning of images, and recurrent Neural Networks are a class of artificial neural networks that are applicable to different problems.

RNN has a "memory" that remembers all the details about what has been determined. It uses the same parameters for each input as it executes the same task on all inputs or hidden layers to create the output. This reduces the complexity of the parameters as opposed to other neural networks. Decisions made by RNN are dependent on what is learned in the past and things learn during training. It has an input layer, one or more hidden layers and an output layer in structure. RNNs have chain-like systems with repeated modules with the idea to store essential data from previous processing phases using these modules. (Nowak et al., 2017) (Le et al., 2019)

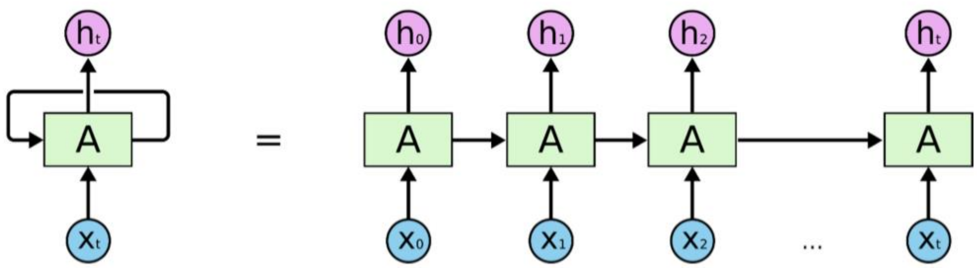


Figure 3.6 sequential processing in RNN (Le et al., 2019)

LSTM introduced by Hoch Reiter & Schmidhuber (1997) (Le et al., 2019) Long Short Term Memory is the basic persistent RNN architecture that can benefit from our experience of classifying, processing and predicting time series of uncertain time lags. LSTMs have proved more precise than traditional RNNs for modeling temporal sequences and long-range dependencies (Azzouni & Pujolle, 2017) (Le et al., 2019).

The back-propagation algorithm is the most widely used learning algorithm to train NNs. The principle of back propagation is to distribute the error from the output into the input, where the weights are periodically adjusted to a default value.

LSTM’s have the capability of learning long term dependencies, and this enables RNN's take a long time to recall the inputs. The LSTM contain in a memory of their information can be considered as a gated cell where the cell decides whether to save or delete information based on the importance the LSTM network assigns to the information. (Azzouni & Pujolle, 2017) (Nowak et al., 2017)

3.6.1 LSTM Architecture

The LSTM network can store, write, read, delete information on its memory. The memory of LSTM is similar to computers memory LSTM networks are embedded in a hidden layer(s) composed of so-called memory cells. There is a hidden process for every phase and feature to perform well with the neural network Each of the memory cells has three gates to retain and change its cell state to forget the gate, the input gate, and the output gate. (Fischer & Krauss, 2018)

There are some researches doing improve LSTM architecture [44] they try to maintain some part of LSTM model and called LSTM like architecture forget gate about the gate is substituted by a practical input-output gates. These LSTM improvements, now abbreviated LSTWM, are named with working memory [41]. There are different approaches applied to reshape LSTM architecture on different researches they also difficult and complex to manipulate when it compared to the base LSTM architecture so its prefer to continue use standard LSTM architecture. As it mentioned before LSTM contain different units called memory blocks. That have memory cell with each self-connection temporary information that still in progress with those gate list input gate, forget gate, output gate so each memory block contains those three objects to work effectively. The structure of a memory cell is represented as follow figure

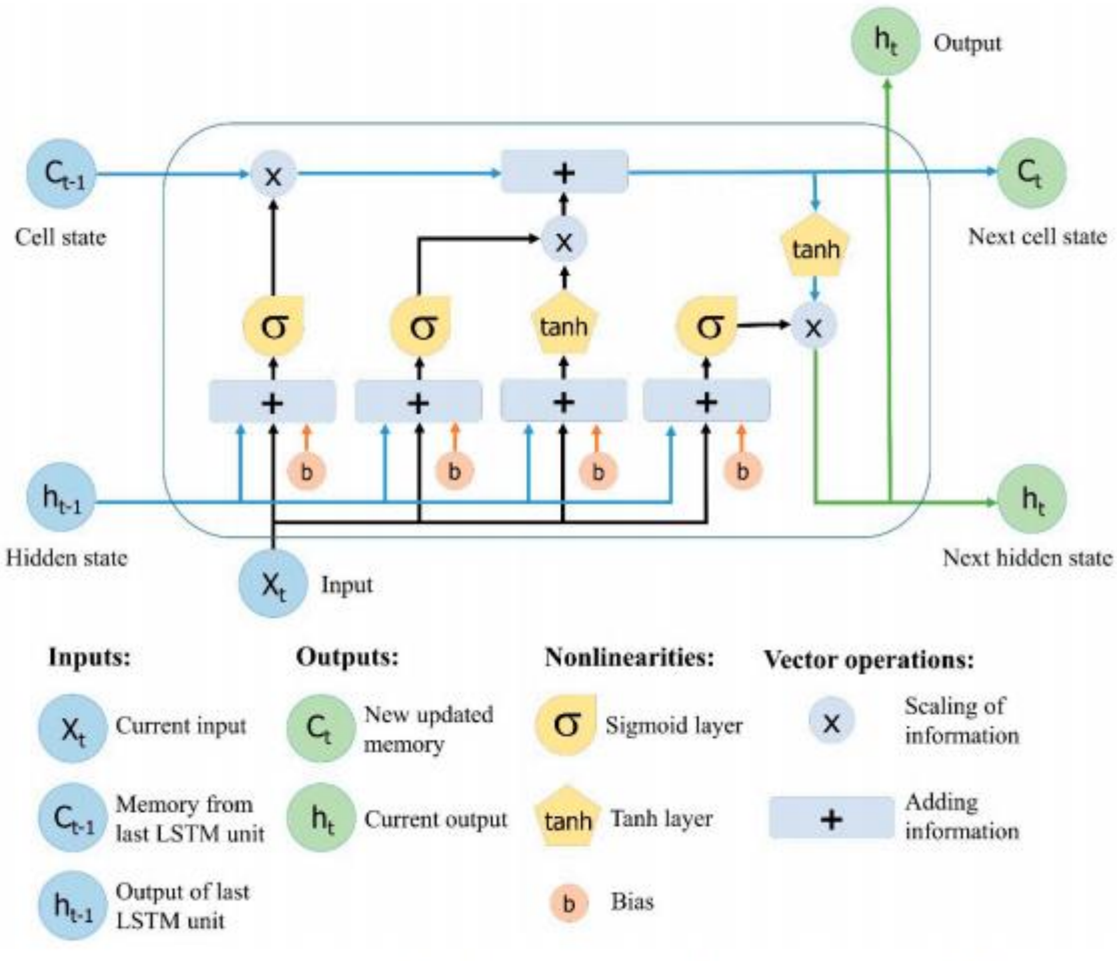


Figure 3.7: Structure of LSTM neural network (Fischer & Krauss, 2018) (Olah, 2015)

The first move will be building an LSTM network is to classify information that is not needed and will be excluded from the cell in that step. Sigmoid function is deciding and identifying also excluding data, that takes the last LSTM unit output (ht−1) at time t−1 and with the current input (Xt) at time t. In addition, the sigmoid function decides which portion of the old output can be removed.

σ (Sigmoid layer): sigmoid function decides which part of the old output would be eliminated or kept for current state. To perform this operation, it uses forget gate value range between 0 and 1 step in constructing an LSTM network

$$ft = \sigma(Wf\,[ht-1,Xt] + bf) \text{….(Le et al., 2019)}$$

1) **Decide the information delete**

The first step of LSTM is to determine which information can release from the cell state. This decision is made by the sigmoid layer, this layer looks at the value of ht-1 output and Xt input, determines the value in the range from 0 to 1 for each Ct-1 state. If the layer is returned 1, this means that this value should be left (remember) if 0 is excluded from the cell state. To use only relevant data and pass for the next state it matters wheatear the information deleted or kept for this specific task it uses sigmoid function namely called forget layer (or ft). Finally, the sigmoid function scales all activation values to the range between 0 (forgotten completely) and 1 (completely remember).

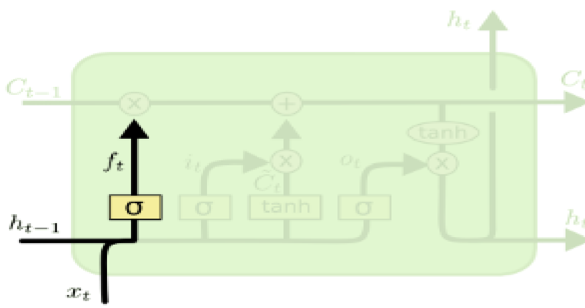


Figure 3.8: Forget layer of LSTM Network (Fischer & Krauss, 2018) (Olah, 2015)

$f_t = \sigma (W_f [h_{t-1}, X_t] + b_f)$ …………………. (3.1) (Fischer & Krauss, 2018) (Olah, 2015)

2) **Deciding what new information store**

The next step is to specify which new knowledge can store in cell state perform this firstly sigmoid function needed to decides which values are going to update when each iteration new information flow those connected networks. This operation it has two main sections of the sigmoid layer and second half of the tanh layer. Layer of Sigmoid this is called the "input gate layer “which determines which values we'll updated or ignored (0 or 1), and second, tanh layer (establish new values for candidates who should be applied to the network).

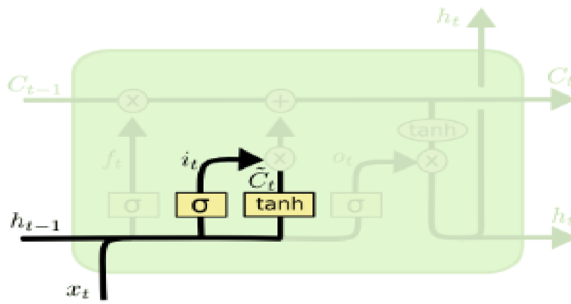


Figure 3.9 New information store

it=σ (Wi. [ht1, xt] + bi) ------------------------(3,2)

Ćt=tanh (WC. [ht-1, xt] + bC) -----------------(3,3)

3) Update the old cell state into the new cell state

Change the old cell state to the new cell state Update the old cell state by multiplying the new cell state with what you have forgotten before, and then adding the latest candidate cell state. The new candidate values are scaled by how much each state value wanted to change. (Le et al., 2019) (Fischer & Krauss, 2018)

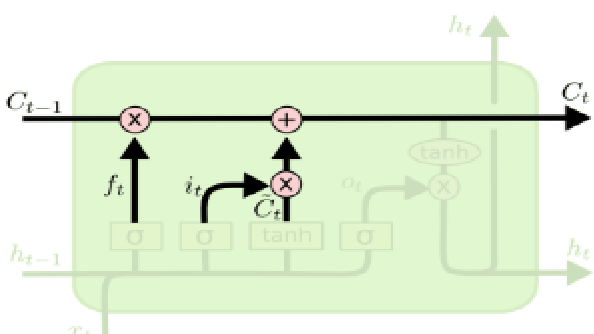


Figure 3.10: Memory cell update layer of LSTM Network

Ct=ft *Ct-1+it* Ćt--------------------(3.4) ………. (Le et al., 2019) (Fischer & Krauss, 2018)

4) Final step, decide the output values

The neural network output layer captures and transmits the information accordingly in the manner it has been programmed to provide. The output value (ht) step is based on the state(Ot) output cell, but is filtered. A sigmoid layer first specifies which parts of the cell state are formed by the cell to output, the output of the sigmoid gate (Ot) is multiplied by the new values created by the tanh layer from the cell state (Ct), with a value ranging between−1 and 1.

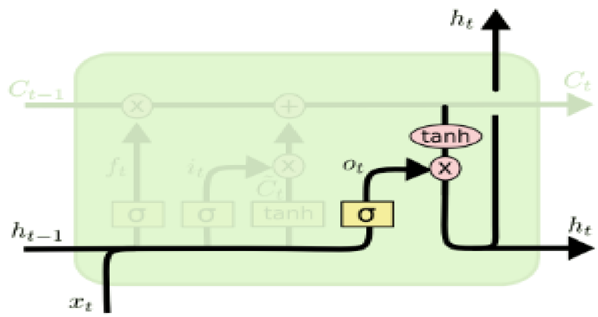


Figure 3.11: Output layer of LSTM Network(Le et al., 2019) (Fischer & Krauss, 2018).

$$Ot = \sigma\ (Wo\ [ht-1,\ Xt]\ +bo)\ \text{-----------}\ (3.5)$$

$$ht = Ot * \tanh\ (Ct)\ \text{----------------------}(3.6)$$

3.6.2 BI LSTM Architecture

The concept of BILSTMs derives from bidirectional RNN [18], which processes data sequences in both forward and backward directions with two different hidden layers. The two hidden layers are connected by BILSTMs to the same output layer. BI LSTMs have been seen to be particularly helpful in cases where the context meaning is required. It's really good for jobs like classifying emotions. In the unidirectional LSTM, information travels from the back to the forward. By comparison, bi-directional LSTM information not only flows backwards to forwards, but also forwards to backwards using two hidden states.(Cui et al., 2018)

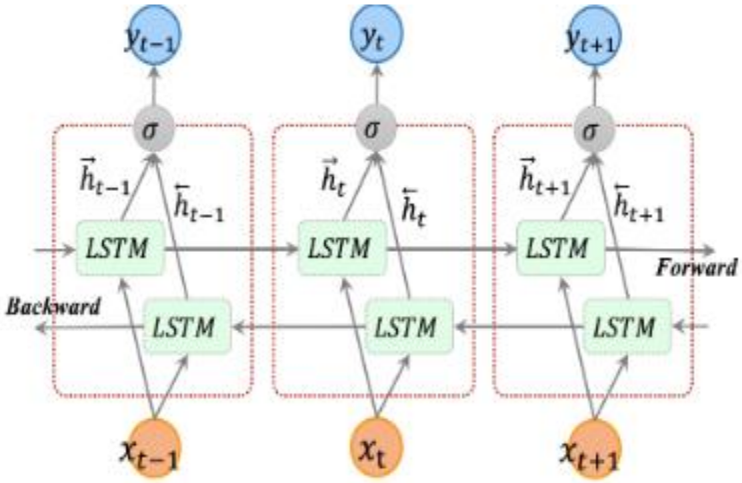


Figure 3. 12 BILSTM Representation (Cui et al., 2020)

Fig.3.12 The forward layer output sequence, $\boldsymbol{h}$ {{$^{\rightarrow}$, is iteratively calculated using inputs in a positive sequence from time $T-n$ to time $T-1$, while the backward layer output sequence, $\boldsymbol{h}^{\leftarrow}$ {{, is calculated using the reversed inputs from time $T-n$to $T-1$. (Cui et al., 2018)

One specific detail that differences when comparing the LSTM and BILSTM methods is that the consistency is achieved at a slower rate in the case of the BILSTM models compared to the LSTMs. In fact, when referring to the BILSTMs, two LSTM networks are applied to input data as follows: first, the input sequence is processed via the first LSTM network, in the so-called "forward layer" while, second, the inverted input sequence is entered through another LSTM model, thus receiving the "backward layer".

Summary

Generally, on this chapter define research methods data representations, process model, Amharic language properties and complexity and different neural network.

RNNs have a big setback called a vanishing gradient; in other words, it is difficult to learn long-term dependency (relationship between entities that are several steps apart). Theory indicates that RNNs are entirely able to cope with such "long-term dependence" and that people should deliberately choose parameters to solve toy problems. Sadly, RNNs appear to be not in a position to learn them in reality.

The problem was explored in depth by Bengio, et al. (1994), who found some fundamental reasons why it might be difficult. A special type of RNN called Long Short-Term Memory Cell (LSTM) was created to solve this problem. Long Short Term Memory networks generally referred to as "LSTMs" are a special form of RNN capable of understanding long-term dependencies.

- LSTMs They are deliberately designed to eliminate the issue of long-term dependence. Remembering knowledge after a long time is basically their default action, not something they're trying to learn.
- All recurrent neural networks have the shape of a chain of recurring neural network modules. In regular RNNs, this repeat module would have a very basic structure, such as a single tanh plate.
- LSTMs also have a structure like this chain but a different structure refers to the repeated module. There are four, communicating in a very unique manner, instead of having a single neural network layer.

CHAPTER FOUR

4 PROPOSED INTENT CLASSIFICATION MODEL

The key contribution is this thesis is to be established LSTM based intent Analysis for Amharic sentences and investigating the effect on comments from You-Tube, the procedures and methodologies followed are discussed in this chapter.

4.1 Data Collection

To develop an intent analysis, data collection is the first and fundamental step that encompasses collecting user-generated contents from blogs or social media like Facebook, YouTube, Twitter, and others. Scraping is the process of collecting user-generated content from social media and blogs. This research paper uses You Tube Comment Scarper for collecting raw data in the form of unstructured data because they are expressed in different vocabularies, ways of writing and others. And simple random sampling technique for sampling our data. Total data collected from YouTube is around 15,321 raw data without emoji's and graphical representations from 18 Amharic videos. Data attributes of the collection with the time interval on 2019 January to 2020 march. Mainly focused on Amharic Video to collect Amharic based content movie, music, Spiritual, Electronics videos are addressed on data collection.

የምር ቀሽት ሙዚቃ ደጋግሜ እየሰማሁት በመዳም ኩሽና ፈታ….
የምር ቀሽት ሙዚቃ ደጋግሜ እየሰማሁት በመዳም ኩሽና ፈታ….
ጌታን ኮመንቶች ጦሽ አረጋቹኝ ወይኔ የት ሄጂ ነው ወሬ ያመለጠኝ
አስጌ ለመጀመሪያ ግዜ አራዳ አራዳ ሲጫወት ግን አለምደው ያለ ይመስላል ጥርሱ ጠፋ
ምድር በልቃጣ እጭ ሀበሻ መድህን አስጠላኝ ሴት ልጅ ገላዋን እያሴየች ቱቱቱቱቱ
ናቸው።
ይህም መልካም ስብእና እና የሰከነ መንፈስ የተላበሱ፤ በማህበራዊ ህይወታቸውም የተሳካላቸው መሆናቸውን ያሳያል። 3. ለእነሱ መልካም በሆኑ ሰዎች የተከበቡ እና ወዳጅነት ለመፍጠር የማይቸገሩ ሲሆን፥ በጥሩ የማይመለከቷቸውን ለማራቅም አይዘገዩም። 4. ስሜታዊነት(ግልፍተኝነት) እያጠቃቸውም፤ የሚያስቀይም ትችት ለመስጠትም አይፈልጉም። 5. በጠላቶቻቸው ከመደናቀፍ ይልቅ የመነቃቃት እና የመነሳሳት ባህሪ የተላበሱ ናቸው።
6. ሰዎችን ለማስደሰት ሳይሆን ጠቃሚ ነው ብለው ላመኑበት ነገር ትኩረት ይሰጣሉ። 7. ግለሰባዊም ሆነ ሙያዊ ምራላቸው ከፍ ያለ ነው። 8. የአካል ብቃት እንቅስቃሴ ያዘወትራሉ። 9. የተመጣጠነ ምግብ ይወስዳሉ። 10. አዘውትረው ያነባሉ፤ "ማንበብ የህይወት መስመርን ለማስተካከል ሁነኛ መሳሪያ ነው" ብለው ያምናሉ። 11. ችግርን ለመጋፈጥ አይፈሩም፤ ለፀፀትም ቦታ የላቸውም። "ህይዎት ችግርም መልካም እድልም አላት፤ ብዙ ፈተና በርካታ እድሎችን ይዞ ይመጣል" ባይ ናቸው። 12. አያማርሩም፤ ዛሬን በደስታ ለመኖር እንጅ ያለፈውን እያብሰለሰሉ ለመቆዘም አይሞክሩም። 13. ጭንቀት አያበዙም። 14.🔥🔥🔥🔥👏👏👏👏
ዋው👏👏👏👏
ዋው. እዴትደስይላል. ይመቻችው. ወሎከወላይታ ሲቆራኝ እድ ነው. ቤሄራችን. ፍቅርማለትይሄነው
ኑሮ በአመርካ ከኛ ጋር መቃኘት ይፈልጋሉ? ኑሮ በአመርካ ከህ እስከ ፐ ለመቃኘት ዝግጅት ጨርሴናል:: እንግዳውስ ስለ ኑሮ በአመርካ መረጃ ለማገኘት ከፈለጉ ቻናላችን Like, Subscribe ብእማድረግ ይጠብቁን::

Figure 4 13: Row sample data set

4.2 Data Preparation (Preprocessing)

Data preparation is a major step in data analysis that encompasses many techniques mainly data transformation, data cleaning, and data reduction to make the collected data clean and suitable for analysis. In data preparation, non-textual contents and contents that are unimportant for the analysis are first identified and eliminated.

Natural Language applications use consecutive data preparations approaches due to many reasons, for example, real-world data is full of impurity and incompleteness, data preparation generates a small amount of quality data set than the original one which improves the efficiency of data mining algorithms and results in high-performance mining systems, this research paper come up with the following approaches for cleaning the collected data.

- Eliminating non-textual contents and symbols like hashtags, URLs, links and others (György, 2017) (Watanabe et al., 2018).
- Removing non-Amharic words
- Spell correction has been done.
- A consistent tokenization rule should be applied to split sentences into words and treat each punctuation as a separate token word tokenizer from NLTK is used in this thesis. Also this research assumes that punctuation elements, including capitalization, the presence of question marks and exclamation marks, etc., tend to identify hateful expression, and cannot simply be dismissed.

4.2.1 Tokenization

Tokenization is one of the preprocessing steps in NLP. Tokenization is splitting the text into smaller units and each unit is called tokens. Tokenization mostly is the first step in pre-processing of the input review. The input for this activity is the actual review which is going to be categorized. This activity reads a sequence of characters as a string and tokenizes them using predefined list of delimiters such as new lines and space (Mihret & Atinaf, 2019).

Separate tokens are properly separated, such as punctuation marks, words, email addresses, links, numbers, abbreviations, etc. This process detects the boundaries of a written text, Tokenizing of a given text depends on the characteristics of language of the text which it is written. The Amharic language has its own punctuation marks that demarcate words in a stream of characters. The tokenization of text on this component uses Amharic punctuation marks and white space.

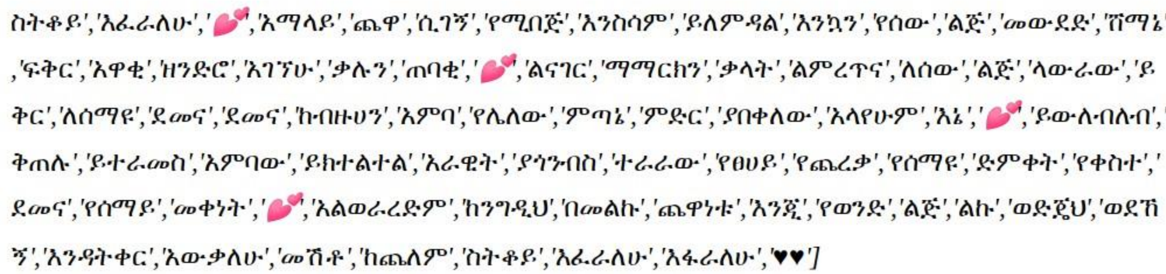

ስትቆይ','እፈራለሁ','💕','አማላይ','ጨዋ','ሲገኝ','የሚበጅ','እንስሳም','ይለምዳል','እንኳን','የሰው','ልጅ','መውደድ','ሽማኔ'
,'ፍቅር','አዋቂ','ዘንድሮ','አገኘሁ','ቃሉን','ጠባቂ','💕','ልናገር','ማማርክን','ቃላት','ልምረጥና','ለሰው','ልጅ','ላውራው','ይ
ቅር','ለሰማዩ','ደመና','ደመና','ኩበዙህን','አምባ','የሌለው','ምጣኔ','ምድር','ያበቀለው','አላየሁም','እኔ','💕','ይውለብለብ','
ቅጠሉ','ይተራመስ','አምባው','ይክተልተል','አራዊት','ያጎንበስ','ተራራው','የፀሀይ','የጨረቃ','የሰማዩ','ድምቀት','የቀስተ','
ደመና','የሰማይ','መቀነት','💕','አልወራረድም','ከንግዷህ','በመልኩ','ጨዋነቱ','እንጂ','የወንድ','ልጅ','ልኩ','ወድጄህ','ወደኸ
ኝ','እንዳትቀር','አውቃለሁ','መሸቶ','ከጨለም','ስትቆይ','እፈራለሁ','እፋራለሁ','♥♥']

Figure 4.14 sample of tokenize data set

4.2.2 Normalization

Preprocessing tasks such as normalization, tokenization, stop word and number elimination, stemming, weighting words, and dimension reduction are performed (moving varying Amharic characters with similar sound to a standard type, changing punctuation marks to space). After all these preprocesses, datasets are prepared in a matrix form, which is accompanied by the preparation and testing of training and evaluation data sets. A model (classifier) is developed from the training data set.

Normalization is one of the steps used to get clean data from unstructured textual data. After tokenization, it is normalization of homophones that is followed. Amharic writing system has homophone characters, characters with same pronunciation but different symbols; for example, it is common that the character ስ and ሥ are used interchangeably as ስራ and ሥራ to mean work, also ሀ, ሐ or ኀ should be changed to the same form because they have the same punctuation all represent one alphabet called "ha". Such types of inconsistencies in writing words are handled by replacing characters of the same sound by a common symbol. The normalization handles:

Substitution of Amharic alphabets with the same spelling and use, but different representation with common alphabet. Short forms of characters that are usually written using forward slash "/") and period ("."), for example, ጠቅላይ ሚንስተር can be written as ጠ/ሚኒስተር አዲስ አበባ as አ.አ and ዶክተር as ዶ/ር. There are two kinds of Normalization the first one is character level Normalization, and the second one is word-level normalization. Normalized characters and words.

4.3 Data Annotation

4.3.1 Stop Word Filtering

Removing stop words helps to reduce the dimensionality of a term space in Amharic, there are common stop words which are used for grammatical purposes, such as ነዉ, ነበር, ሆኖም, እና, ነገርግን, etc, to classify records which are non-informative. In addition to the standard stop words, there are also news relevant stop words including ገለፁ, ዘግበዋል, አስታወቀ, etc.

Those reductions help a lot in terms of dimensionality but on sentiment analysis researches apply stop word removal more than intent analysis because when it see based on our context there are some additional label like question, when, why, where those words are not relevant and its removed as stop word, because those words used for our corpus. The following figure shows steps followed for data collection, preprocessing and annotation

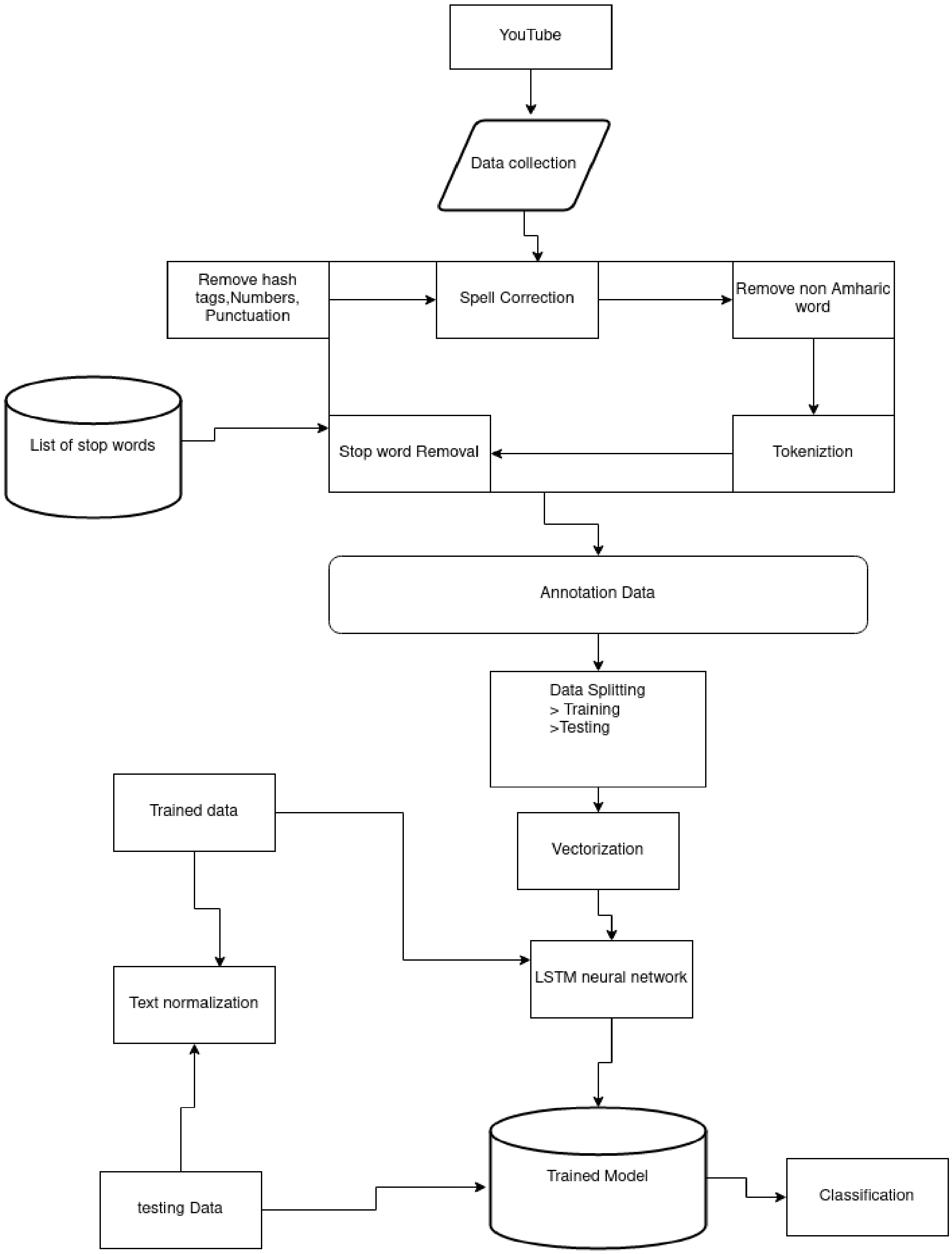


Figure 4.15 Data Collection and Annotation with development process

For example, argue, argued, argues and arguing will be reduced to argu (György, 2017). The trouble with stemming is that multiple words may be simplified to the same word, e.g. an end cat scan of the same root cat based on the heuristics used to produce these similar roots. Amharic is morphological rich language. Amharic can have many words with the attachment of different affixes to a stem. Stemming can be applied to both inflectional morphology and derivational morphology or on either of the two. Derivational morphology usually results in a change in class of word which may result in some loss of semantic. This semantic loss of a word may create a negative effect on the performance of system. For example, the Amharic word ዳኛ (a judge or an arbiter) and ዳኝነት (judging) has the same stem ዳኝ. But these two words have different meanings. Amharic language includes some prefixes and combination of prefixes and suffixes which create negative meaning when they are applied to a given stem. This kind of semantic loss is not usually witnessed during inflectional morphology, which usually involves grammatical features such as, singular/plural, tense.

4.4 Intent Detection Approaches

In order to detect the emerging of intent topics, finding the topics from the collection of documents are the basic step to find and detect those topics is really important to select the best modeling process to get effective result. One of the pointed process is convert high level text in to a vector form *word 2 vector* is algorithm used for this task.

4.5 Neural Network Building Process

Neural net is a machine learning system that uses a network of functions, normally in another form, to interpret and convert a data input from one form into a desired output. Neural Networks is a technique that aims to optimize certain weights, the neuron body, that are multiplied by the dendrites, the vector of characteristics.

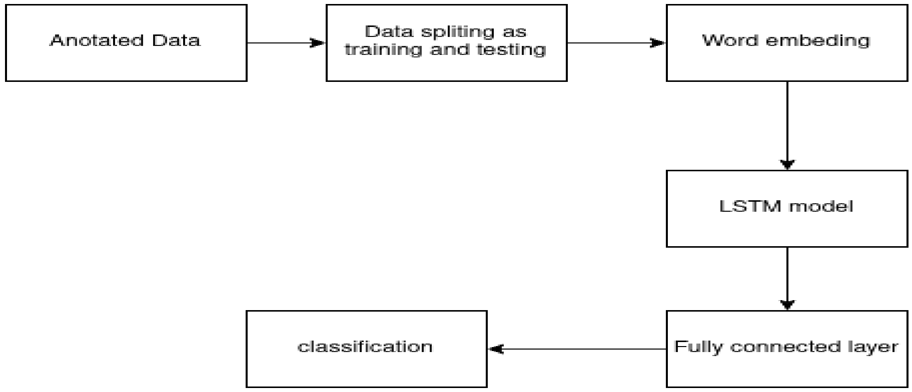


Figure 4. 16: LSTM network development steps(Azzouni & Pujolle, 2017)

4.5.1 Development Steps

Using neural networks or matrix factorization, the learning of word integration can be achieved. One widely used word embedding method is Word2 Vec, which is basically a computationally efficient neural network prediction model that learns word embedding from text.

4.5.2 Data Splitting

One of the most important facts in machine learning is that algorithms store the data and it has low performance in order to use it with new data, this behavior is known as over fit. To prevent this issue. Thus management is required to get effective result data set which divided in random parts in this step, the whole data is separated into training and test with some ratio (90%:10%,80%:20% or 70%:30%)

4.5.3 Word Embedding

The word embedding is a weight matrix, with a row for each word/token in the vocabulary, and the deep learning process is split into two parts: word embedding and LSTM (Long Short Term Memory). In this work, the word embedding's were generated in word2vec. In the random vector setting, all the words in the corpora are initialized with random values. These random vectors are fed to the LSTM model to classify the words into the defined labels.

The fundamental theory of the word2vec model is that words that appear in a common way are related. There are two flavors of the Word2Vec algorithm: the continuous bag-of-words (CBOW) model and the skip-gram (SG) model. Given a word context, the continuous bag-of-words model predicts a target word, while the skip gram model flips the CBOW architecture by predicting the context of a given word. (Sikdar & Gambäck, 2018).

4.5.4 LSTM

LSTM network is learned using these word embedding's and classifies the words into the pre-defined label categories as it mentioned model was utilized to classify the words based on the word vectors generated by word2vec, LSTM is really effective on capturing long and short term memory, takes vectors of word2vec as input from the dense layer. LSTM layer has a memory unit that helps in adding and removing previous information. LSTM model, the network is learned by using training data and tested data. The network consists of an embedding/input layer, two hidden layers, and an output layer [2].

4.5.5 Evaluation

After classification has finished, the performance of the classier is expressed using different performance metrics like (recall, precision, F-score and accuracy) are displayed using Scikit-learn python library. Training and validation ac curacies and losses are also used as evaluation metrics.

CHAPTER FIVE

5 Experiments and Result

In this chapter, the results of active learning experiments will be presented. Basically related with our research questions. Also the data representation and prepossessing steps are explained, With data properties. Afterwards, the experimental settings are explained in order to introduce the rules of experimentation for a better overall understanding. of steps taken and the process of reasoning. Later, evaluation and extensive results are presented in the same chapter.

5.1 Data Description

For this research, the corpus collected and prepared Amharic language, which manually annotated by humans. Data set created that containing only text comments written in Amharic language. The data set contains 12,712 comments collected from YouTube API using YouTube comment scrapper and python program to scrape the comment.

5.2 Human Annotation Task

This research investigates alignment of comment labels that are annotated manually by humans and all comments from the total data set are annotated by three human annotators. To investigate annotation disagreement by two human annotators and analyses the possible way to become agree on the majority annotation disagreement. Possible annotator disagreement could occur due to the following reasons: i) absence of a lexicon in the lexicon database. ii) text annotation is subjective. A positive text for one person may be negative for the other, and iii) Contextual expression of a sentence. One can express a positive felling using a negative word and vice versa. From the results, most comments that are not correctly classified are context-based expressions. For example, "መቼ ነው ቁጥር ሶስት የሚወጣው" express one's positive excitation to see the next episode of certain film or action, "ጎበዝ እንባዬ አልቆም አለኝ" express one's negative feeling about certain event using positive word. It is to mean that "I cannot control my tears" these and other kinds of expression are context-based expressions.

There are incorrectly labeled multi-aspect comments that lower the automatic labeling system. For instance, "ነፃነት ግን መች ነው ገፅ ባህሪውን የሚቀይረው" and "ከብዙ አመት በኋላ አየሁትግን አይሰለቸም" are example of incorrectly labeled comments.

No	Comments	First annotator polarity	Second annotator polarity	Third annotator polarity
1	አብይዬ አላህ ይጠብቅህ	Positive	Wish	Positive
2	አረ የት ሄጄ ልሣቅ ጉድ ነው	Negative	positive	Negative
3	መጀመሪያ የሰው ጭንቅላት ላይ ነው መሰራት ያለበት ግን ይቅርታ አስተያየቴ ከከፋቹ ግን ትክክል ነኝ	positive	negative	positive
4	አገር በስራ እንጅ በወሬ አይገነባምና	suggestion	Negative	Negative
5	ብሎክ አታድርጉት	suggestion	positive	positive
6	ሸህ ዓመት ይንገሱ	positive	wish	suggestion
7	ኢትዬጵያዊነት ለዘላልም ይኑር ሰምተኻል ማነኸ እዛጋ ምድረ ቅጥረኛ	positive	negative	Wish
8-	እውነት ለመናገር አንተ ሒሳብ ፊዚክስ ኬሚስትሪ ብታሰተምረኝ ሰቃይ ነው ምሆነውአንተ አስረድተከኝ አለመረዳት ከባድ ነው	positive	Negative	positive
9	ግን ትንሽ ማስተካከያ ብታረጉ አሪፍ ነው ትንሹ እዮብ አሪፍ ነበር	positive	suggestion	Negative
10	ግን ፊልሙ ፈተናው በዝቶበት ነበር ለበጎ ነው	negative	positive	suggestion
11	ምነው ምጣኔ የሚባል የለም ሰአት ማን ስራ ፈቶ ያያል ?	Question	negative	positive
12	እባካቹ ግጩኛ ?	Negative	Question	suggestion
13	እልል በናታችሁ እንዳታነሱ ?	Suggestion	Positive	Positive

Table 5.5 Sample inter-annotators disagreement

This paper generalizes there is no a total agreement between the inter-annotators in all comments they are assigned to label. However, most disagreements are resolved through discussions. This corpus contains Amharic words in their context with appropriate intent analysis aspect Table provide information about the data set before and after applying prepossessing on the data set. The information includes: the total number of sentences and words.

Before and after pre process	Total number of sentences	Total number of words
Before	15,321	134,251 words
After	12,712	115,776 words

Table 5.6 data set characteristic

5.3 Evaluation Matrix

As the name suggests, the evaluation matrix gives us a matrix as output and defines the model's maximum results, well-known evaluation metrics such as precision, accuracy, recall and F score. (Mihret & Atinaf, 2019)[22] [32] [33]. The most commonly used performance measures in text classification specially sentiment analysis by precision, recall, accuracy and F score (Wondwossen Mulugeta, 2014).

TP (True Positive): predicted positive and the actual output was also positive measures

TN (True Negative): predicted negative and the actual output was negative

FP (False Positive): predicted positive and the actual output was negative.

FN (False Negative): predicted negative and the actual output was positive.

Precision**:** It is the number of correct positive results divided by the number of positive results predicted by the classifier [29 35].

Precision= TP/FP+TP………………………………. (5.1)

Recall: It is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive) [29 35].

Recall= TP/FN+TP ………………………………… (5.2)

F1 score - F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. [29 35].

F-Measure = (2* Precision * Recall) / (Precision +Recall) …………. (5.3)

Accuracy - Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. [29 35]

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(5.4)}$$

5.4 Experimental setup

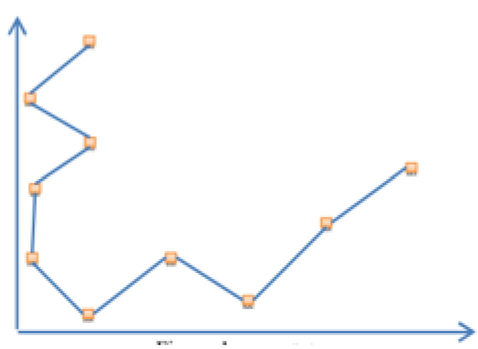
This paper uses one personal computer for all experiments. Table 5.8 shows the hardware and software specification of the machine used in all experiments

Manufacturer	LENOVO
Model	G-50
Processor	Intel® Core™ i7-4510U CPU @ 2.00GHz × 4
Memory(RAM)	8 GB
Operating System	Ubuntu 20.04 LTS

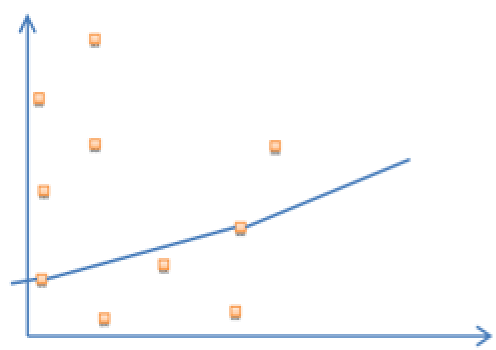
Table 5.7 Experiment setup

Model Configuration: In addition to hardware and software specifications, configure the LSTM network required for developing intent analysis model. In the experiment,

LSTM network is configured with the following parameters: one LSTM layer with 8 neurons, 20 epochs, 512 batch, 8 NB_WORDS (Embedding size), 'Adam' optimizer, 90% train size, and 10% test size. main problem in training neural networks is the over-fitting and under fitting (Jabbar & Khan, 2015). Over fitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. Generally processing on training task it’s not improving to increase the performance, but its adding and continue randomly learn training pattern

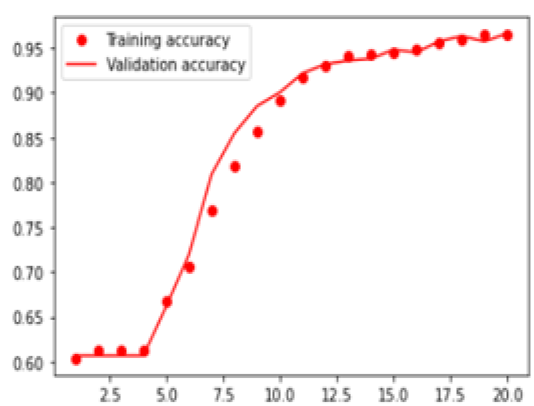


sample of over-fitting model

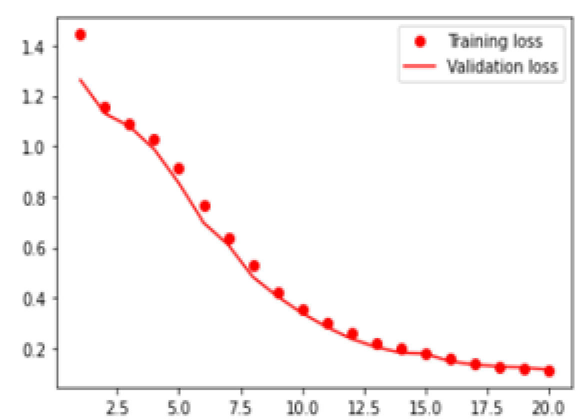


sample of under fitting model

Figure 5.17 sample of over-fitting model and sample of under fitting model (Jabbar & Khan, 2015)



Training and validation accuracy LSTM



Training and validation loss LSTM

Figure 5.18 training and validation accuracy LSTM and Training and validation loss LSTM

The inverse of over-fitting is under-fitting. This arises because the model is unable to capture the data uncertainty. The effect of sampling noise would be small training data and complicated relationships, so they will appear in the training set but not in actual test data, even though it is taken from the same distribution (Jabbar & Khan, 2015)(Srivastava et al., 2014) . In order to choose an appropriate epoch number, this paper has used training versus validation accuracy plot of the model. To identify occurrence of over fitting from a plot the following conditions will occur. Figure 5.18 shows sample training vs validation accuracy and loss plot of the model using our data set.

Our data set is small and unbalance in comparison to what deep neural networks require, so it is necessary to use techniques such as early stopping and dropout to avoid over fitting and balance data Modify the training set to make it optimal for a regular learning algorithm. (Jabbar & Khan, 2015) (Srivastava et al., 2014)

5.5 Handling imbalanced data

Three different core approaches to learning from imbalanced data:

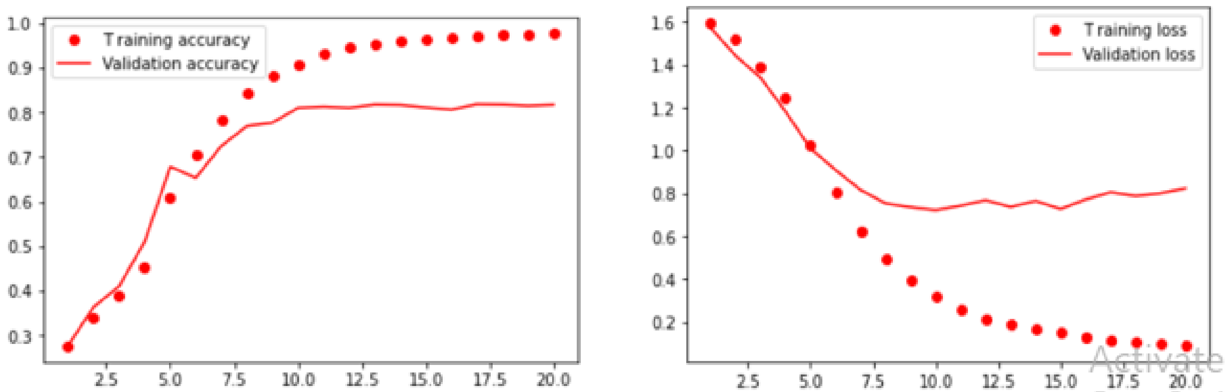
First method for handling over-fitting model dealing **data-level** methods that modify the collected data set to balance distributions by remove or maintain difficult samples also balance corpus amount of each class numerical properties.

The second one dealing **algorithm-level** approaches that retain specifically alter current learning algorithms to eliminate bias against majority artifacts and to apply them to mining data with different distributions.

Last method applying both the above two method or hybrid method. Also there are plenty of methodologies Used to prevent over-fitting and under-fitting and **early-stop approach** and **dropout.**

Early Stopping Configuration: - Early stopping is a way to define and end training times arbitrarily when the performance of the model stops to increase in the data collection.(Lewis, 2016) (Jabbar & Khan, 2015).

Dropout: Dropout helps to prevent over fitting by randomly dropping some of the unit sand their connection from the neural network during training (Lewis, 2016) (Jabbar & Khan, 2015). In this paper the optimal choice of dropout is 0.2 specifically between 0.2 and 0.5.



Training and validation accuracy LSTM *Training and validation loss LSTM*

Figure 5.19 Training and validation accuracy LSTM training and validation loss LSTM

In data for machine learning or deep learning classification scenarios, imbalanced data is typically seen and refers to data that includes a disproportionate proportion of observations in each class. This disparity will lead to a wrongly assumed positive influence on the accuracy of a model, since the input data has a bias against one class, resulting in the trained model imitating the bias.(Jabbar & Khan, 2015)(Srivastava et al., 2014) The above figure show after maintain data set by balance of each class data to decrease over-fitting occurrence. As it seen in the above figure after applying balance on each class it shows and give as a better plot and better when perform classification task

5.6 Development Tools and packages

These section reveal development tools used in this research. python programming language commonly building, compare, evaluate, graphical representation, data processing is performing by using python programming language. Python **is Interactive**, **Interpreted** Consistency and simplicity are some beneficial properties. Also reduce development time, because it contains many libraries and framework for machine learning and deep learning architectures.

5.6.1 Libraries and frameworks

Scikit-learn is suitable for processing basic algorithms for machine learning, such as clustering, linear and logistic regression, regression and classifications for small-to medium-sized data sets for feature engineering and classical ML simulation. (Raschka et al., 2020) evaluate deep neural classifiers by calculating A confusion matrix, which is a table frequently used to explain a classification model's output.

Pandas is suitable for advanced data structure and analysis, allowing data to be merged and filtered, and data collected from other external sources such as Excel. (Raschka et al., 2020)

Keras Main focus is enabling implementation of deep learning models for research and development as quickly and easily as possible, suitable for deep learning, enabling fast calculations and prototyping. Because the software library uses a GPU in addition to the computer's CPU. (Raschka et al., 2020)

Tensor Flow is suitable for deep learning by setting up, training, and using large data sets of artificial neural networks. (Raschka et al., 2020) Tensor flow is used as a back end for Keras library which is used for development of deep neural network classifiers.

Matplotlib is suitable to construct 2D plots, histograms, graphs, and other visualization types operations to easily and cleanly data and progress representation based on give para meters. (Raschka et al., 2020)

NLTK It is a tool used to create Python programs to work with human language data. Suitable for computational linguistics and the identification and processing of natural languages.(Raschka et al., 2020)NLTK is used for many tasks like tokenization, lemmatization, stemming, parsing and POS tagging. NLTK is developed especially for languages that are rich in resources.

5.7 Experimental scenarios

To answer the research questions that set by the researcher different experiments conducted. Generally, these experiments are grouped into three categories. The first group of experiments aims to show the evaluation of different classifier models with different feature. The three approaches are, using collected data as an input for three intent classifiers, LSTM, BILSTM and Naive Bayes.

These experiments answer RQ1: Which algorithm perform better performance intent analysis model? The second group of experiments aims to identify the impact of noise and normalized test data related intent classifier model. These experiments are conducted to answer RQ2: Examine variation of predicted class.

5.8 Results

Building and Measuring intent classification model performance

Most research made around intention analysis have specific scope like commercial based (Luong et al., 2017)(Cohan-Sujay & Madhulika, 2012), Domains intention of buying, not buying (Luong et al., 2017)(Cohan-Sujay & Madhulika, 2012). This paper prepares a model that would be common and general representation widely applying by human intention for that specific item identifying and knowing user intention help us to improve the contents that published, to followers or customer. To analyze user intention, the class classified as positive, negative, suggestion, question, and wish to see the labeled data set, and number of labeled comments in each intent class, and intention classifier performance.

Type of comment	Amount of comment
Positive Comments	3554
Negative Comments	2345
Suggestion Comments	2676
Question Comments	1979
Wish Comments	2166

Table 5.8 show using labeled class

The above table shows total number of collected comments combine all class with data properties. For instance, the word" ክፋቱ ምንድነዉ?" represent as question in terms of predefined label, "ደጋግመው ይስሙት" is classified as suggestion "ቴዲ የኒ ኑርልኝ" classified as wish "አልጠግበዉ

አልኩ ይህን ዘፈን” positive “ባንዳ እና ሆድ አደር” negative, those sample label with initialize user comment.

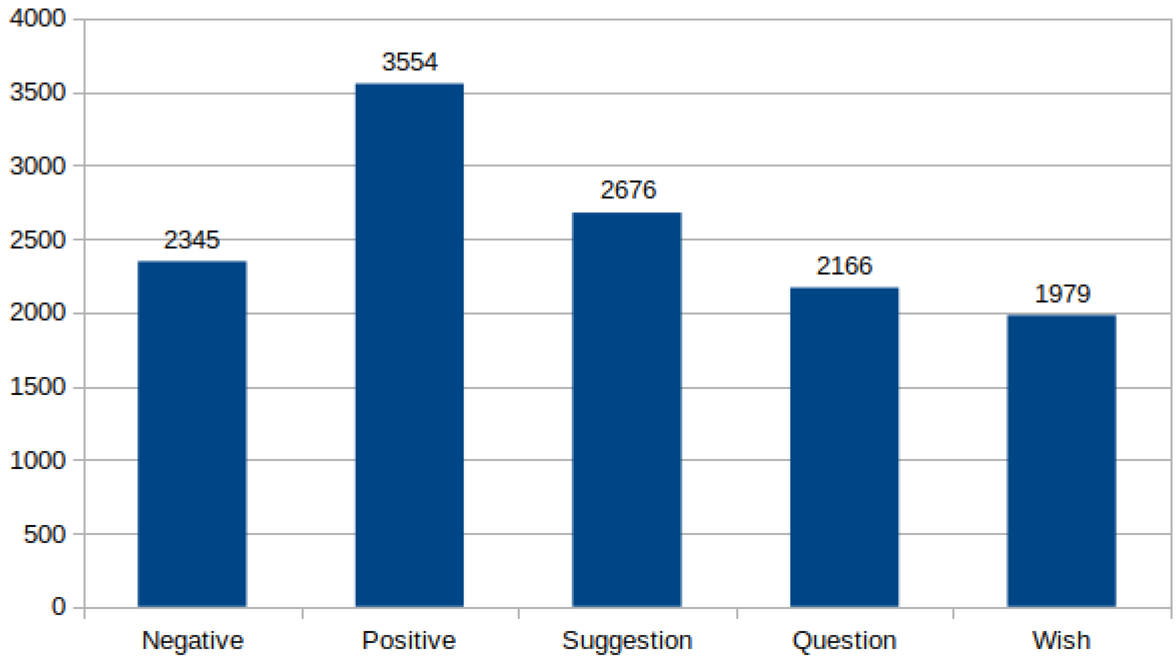


Figure 5.20 Amount of labeled comments in each intent class

Above Figure shows the amount of labeled comments as positive, negative suggestion, question, wish. Total data set 12,725 comments use, for those different domains to separate uniquely flagged each class by different number.

The amount of labeled comments using machine learning approach is lower by the amount of total clearly it labeled by sentence level, obviously if lexicon approach used it might have lot amount of data count. Also its known that those researches have lexicon based but it has its own challenging task specially labeling process. (Jurek et al., 2015)

Above figure shows data distribution not balance comparing each class. Data distribution imbalance applies primarily to different text number of different groups. Generally speaking, the class distribution imbalance is represented by the text number ratio of the small and large classes. To tackle this problem there are different methods applied bus most common way follows three approach to maintain this problem (Krawczyk, 2016). Based on level of branch the complexity of data engineering is more difficult. Binary data model faces binary imbalanced classification; Multi-class imbalanced classification occurs classes more than two or binary this level of branch is more difficulty. Thus this research used multiple class architecture to avoid data imbalance distinguish three major learning methods from imbalanced data:

first one is using **data level** by customize add or remove complex data from the data set. Second one by **Algorithm level** by improving algorithm by directly configure existed model. The last on is use **hybrid** by combining both the above method. (Krawczyk, 2016)

In this research, data distribution imbalance goes to between positive, negative, Suggestion, Question, Wish. When it combined it vary and clearly show imbalance property on our data set, on this research the experiment demonstrate this experiment and answering research question focuses on analyzing the domains that could improve and increase text comment addressing problem, Data level manipulation is also affect other class that contain small amount of data (Krawczyk, 2016) (Y. Li et al., 2010) Next table show the experiment that measure the performance of the model structured on tabular form.

Performance measure	**Domain of class**				
	positive	**negative**	**suggestion**	**question**	**wish**
Precision	0.71	0.88	0.95	0.90	0.93
Recall	0.90	0.67	0.92	0.86	0.86
F Score	0.79	0.76	0.94	0.88	0.89
Macro-Precision	0.88				
Macro-Recall	0.84				
Macro-F score	0.85				
Accuracy	0.83				

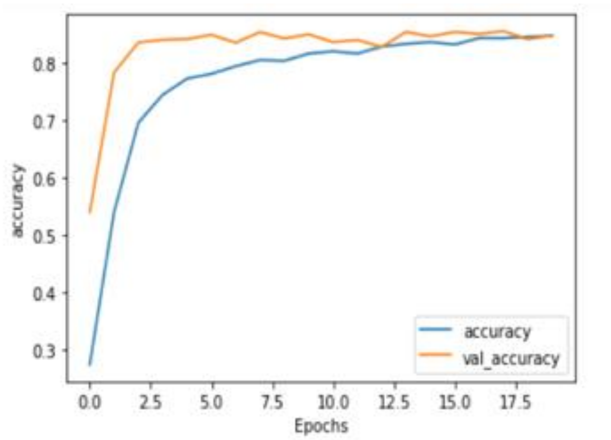
Table 5.9 LSTM performance with defined domains

LSTM network switch 64 layer with 5 dense defined layer for output. 0.2 drop A complete 5×5 misunderstanding matrix was created at the end of each trial, accuracy was calculated for the resulting network, and precision recall and f-measurement were calculated for each target neuron detection rate. The goal production with the highest numerical value was used for the purpose classification. Educated LSTM networks have benefited at least partly from all five intention groups. The networks were trained for up to 20 epochs. The performance of the trained network was measured at 20 epochs check by adding more epochs above 20 to check to increase performance 50, 70,100 epochs result could not change training and validation test above 20 epochs. Every experiment contained 5 trials. To check the variation and take better performance measurement.

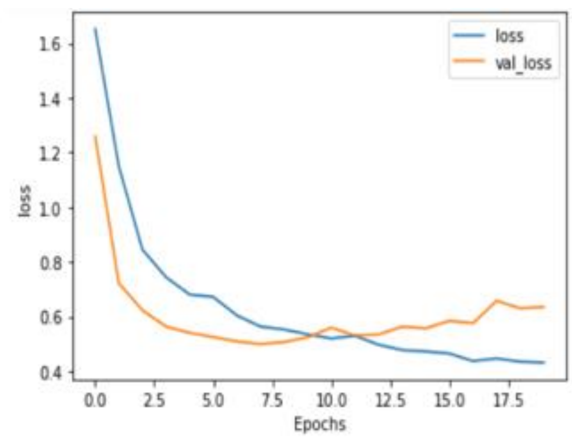
Performance measure	**Domain class 30% test, 70% training data**					**Domain class 20% test, 80% training data**				
	P	N	S	Q	W	P	N	S	Q	W
precision	0.62	0.92	0.92	0.89	0.92	0.71	0.87	0.91	0.80	0.93
Recall	0.83	0.62	0.84	0.78	0.84	0.86	0.67	0.93	0.81	0.82
F-score	0.72	0.74	0.88	0.83	0.88	0.78	0.75	0.92	0.80	0.87
Macro-Precision	0.85					0.84				
macro-Recall	0.79					0.82				
macro-F score	0.81					0.83				
Accuracy	**0.80**					**0.82**				

Table 5.10: Different data split properties

RQ1) Which possible algorithm to perform better performance intent analysis model? For this experiment (LSTM) used, a well–known deep learning model. widely and successfully used for sentence–level natural language processing tasks. To evaluate our model, Bidirectional model applied on different English intent analyses model(Luong et al., 2017). Tenfold, fivefold, and threefold cross-validations used and all experiments were repeated five times. Confusion matrix contains information on the current and expected classifications used by the classifier.



Training and validation accuracy BILSTM



Training and validation loss BILSTM

Figure 5.21 Figure: Training vs Validation accuracy and loss.

The difference model properties for this Bidirectional long short term memory change the drop outs defined 0.8 drop out and activation function for BILSTM "relu" used for the previous LSTM "soft-max" activation used. For this experiment most same training model properties to train model. LSTM network switch 64 layer with 5 dense defined layer for output. A complete 5*5 confusion matrix was created at the end of each trial, accuracy was calculated for the resulting network, and precision recall and f-measure were calculated for each target neuron detection rate. The goal production with the highest numerical value was used for the purpose classification. Educated LSTM networks have benefited at least partly from all five intention groups.

Performance measure	Domain of class				
	positive	**negative**	**suggestion**	**question**	**wish**
Precision	0.74	0.74	0.99	0.90	0.96
Recall	0.87	0.78	0.88	0.81	0.84
F Score	0.80	0.76	0.93	0.86	0.90
Macro-Precision	0.87				
Macro-Recall	0.84				
Macro-F score	0.83				
Accuracy	0.82				

Table 5.11 Comparison of BI LSTM performance with defined domains

Both the above experiment with nearest probability configuration BILSTM is less performance measurement when it compared with the first base LSTM model. Both contain their own feature when it observed from the previous papers. Wisely this research list out different point to compare these two model. So their difference in configuration and feature gives awareness. As the above result describes LSTM better on the numerical measurement. Also BILSTM on our experiment take a lot of time to train when it compares to the LSTM model.

Performance measure	Domain of class				
	positive	**negative**	**suggestion**	**question**	**wish**
Precision	0.76	0.60	0.93	0.85	0.89

Recall	0.66	0.83	0.84	0.77	0.89
F Score	0.71	0.70	0.88	0.81	0.89
Macro-Precision	0.80				
Macro-Recall	0.80				
Macro-F score	0.83				
Accuracy	0.78				

Table 5.12 Comparison of MNB performance with defined domains

Two of the classifiers have shown high performance with average F-score of 83%, 78% for BI LSTM and MNB respectively. BILSTM outperforms MLP and BI LSTM by 5% respectively. Compared to results from proposed approach LSTM it scores lower by 2%, for BILSTM and 5% for MNB. The lower F-measures could be due to three possible reasons, the first one is train test split used in these experiments discards 20% of training data for test.

This affects the accuracy of the model. The other reason is that, deep neural networks need very large amount of data to give better performance and the data set used is not large enough. Finally, the network parameters used in our experiments might not be in their optimized value.

5.8.1 K-fold cross validation

The data is split into k sub-sets in K Fold cross validation. Now the holdout process is repeated k times, so that each time one of the k subsets is used as a test set/validation set and the other k-1 subsets are grouped together to form a training set. tenfold, fivefold and three fold cross validations used and all experiments were repeated 10 times.

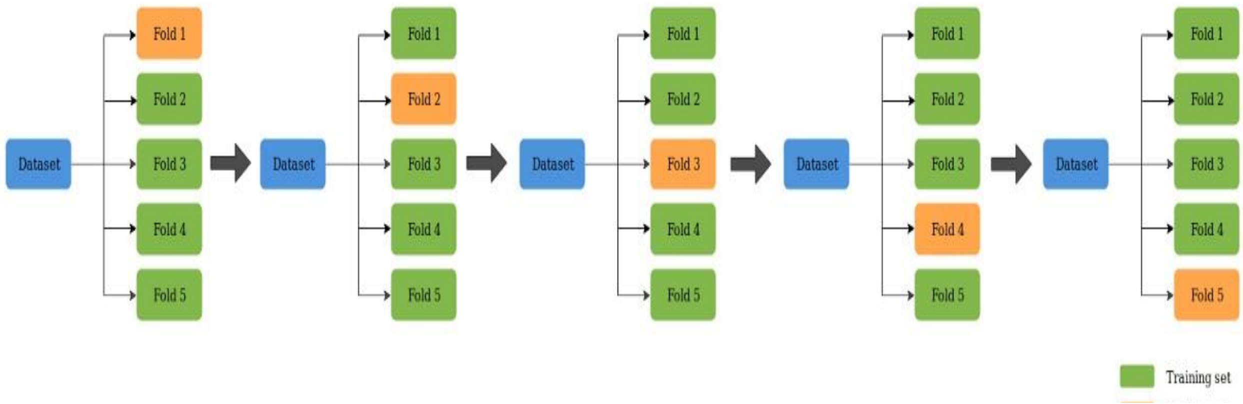


Figure 5. 22: Represent of fivefold cross validation

Algorithm	Average accuracy with 10-fold cross validation	Average accuracy with 5-fold cross validation	Average accuracy with 3-fold cross validation
LSTM	**82.75**	82.35	81.67
BI-LSTM	81.84	**85.19**	80.10
MNB	78.9	**79**	78

Table 5. 13: K-fold for three selected algorithms

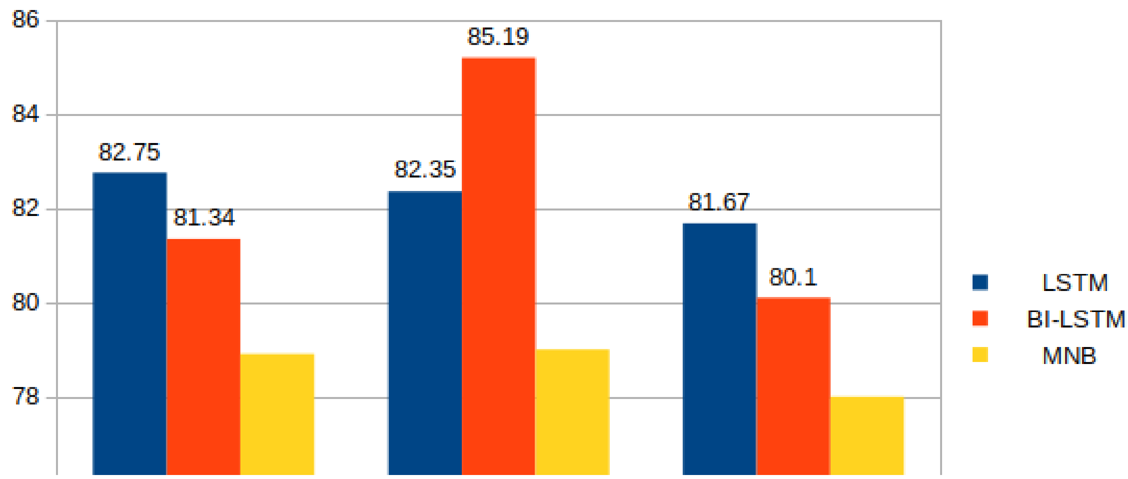


Figure 5. 23: k fold validation summary

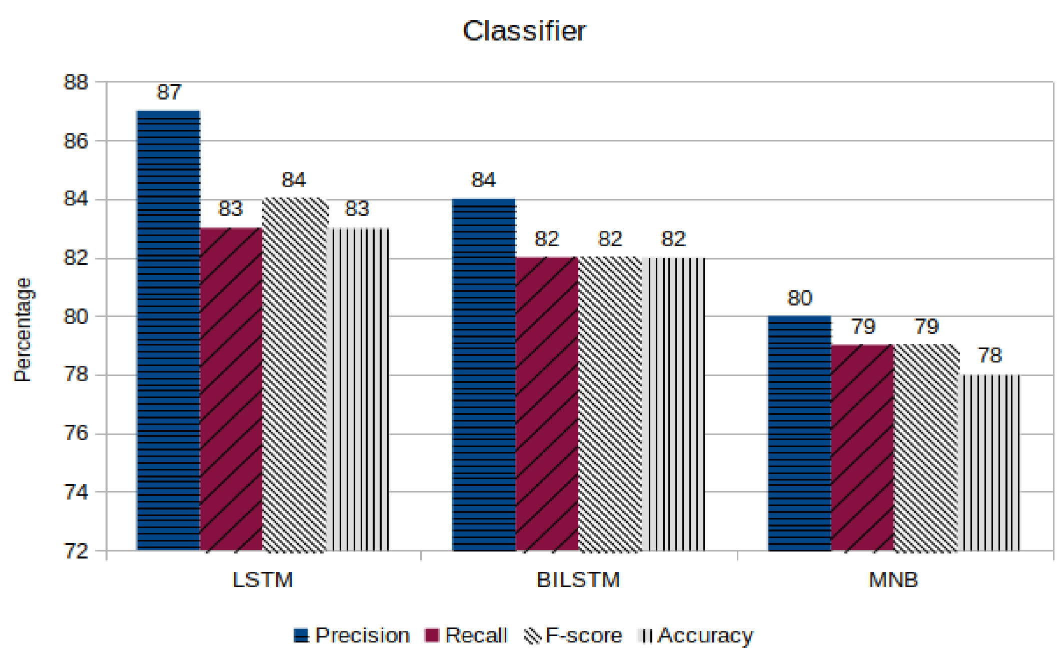


Figure 5.24 Research question (1) Summary

RQ2) **[Could the inclusion of extra white space and misspelled sentence affect Amharic intent analysis?]**

Because large, human annotated data sets suffer from labeling errors, it is crucial that deep neural networks can be formed in the presence of label noise. This section, investigate extra white space and input misspelled data could lead miss classification or variation between wrong domain. Making experiment by give noise input data to see the effect. (ትችላ አባቴ) example of noise data that must be represent as (ትችላለህ አባቴ), (እሰከመ ራብ ድረሰ ደርሰ ነበር) that might represent correctly (እስከመራብ ድረስ ደርሶ ነበር)

Type of data	**Amount of data**		
	Sentence	**predicted**	**Actual label**
Noise extra white space data	በመልም ቀ ን ደስ ይበልህ	Negative	Negative
Noise misspelled data	የሞ መሣርንም አዘጋጀ	Negative	Negative
Noise extra white space data	ፍት ህ ለሚገባቸው ወገኖች ፍት ህ ሥጥ ?	question	question
Noise misspelled data	የት ነም የተቀበው?	question	question
Noise (wish)	እድሜና ጤና ተመልሽ	wish	wish

Table 5.14 Test variation between domains white space and misspelled input

The result from Table 5.14, incorporating noise comment did not affect the model by interchange the actual label form. The above data express without normalized input data it performs well in accurate label prediction. The researcher proposes to study the problem of inferring intent categories for comment. the problem formulated as a miss classification problem. Test model with noise input data to see the probability of vary on final output label. Test all selected noise data listed on the above table, Result shows noise input not affect classification vary between predefined label in terms of target or expected output in LSTM based intent analysis model.

5.8.2 Overview of Framework Flask

Flask was developed as an extensible platform from the ground up; it offers a solid core of basic resources, while extensions provide the remainder.

5.8.3 System Design

The framework architecture to incorporate the web-based system on the structure flask will be explained in this section. This design would create a folder with a framework for the project and essential files will be visible. The phase diagram of the device is shown in figure 5.25.

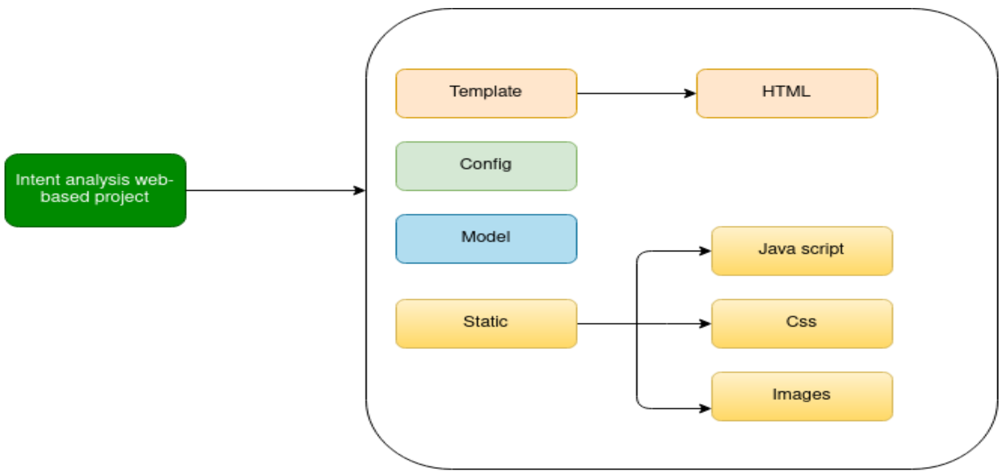


Figure 5. 25: system process diagram

5.8.4 Produce the project Structure

This section describes the system design to implement the flask structure website. In this system, the main architectural concept has basic components such as Model, View and server Figure 5.26 to the model, output from the model will be sent to the server after the server receives model outputs, data will be sent to view.

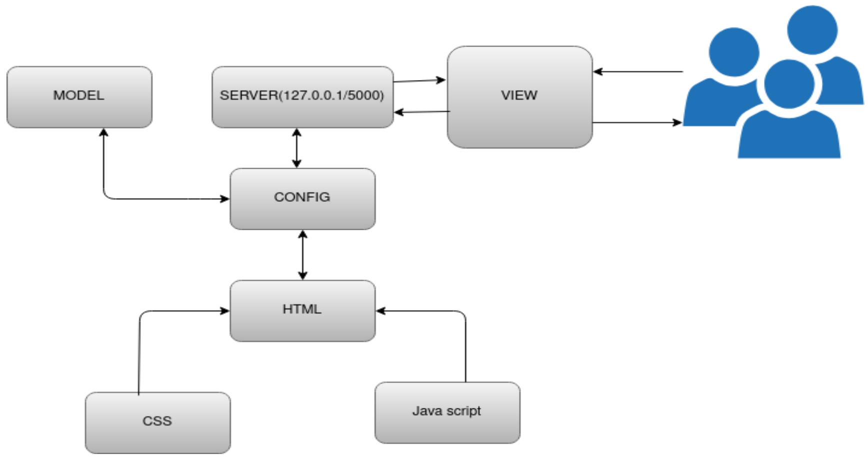


Figure 5. 26: Intent analyzer pattern in web based-system

5.8.5 Performance Evaluation

In this portion, a test process will be carried out to evaluate the web-based model design of the Flask system using the Python language. This test will be divided into two parts, namely the implementation of the website and the integration of the LSTM model using the flask framework.

a) Implementation of website creation

To build a new website, using Flask to make it from a project generator. Project folders and files are organized, a folder structure that is often fitted with an app.py web server. The view to run the server, and start page view of the home view that is linked to the. Index html.

b) Startup Server

The framework instance has a run method that launches the integrated web server of Flask:

```
if __name__ == '__main__':
    app.run (debug=True)
```

The __name__ == '__main__' here, the Python idiom is used to ensure that the development web server only runs when the script is directly executed. After running the server it generate local server port which is http://127.0.0.1:5000/ running this address opens the home page figure 5.27 shows home page which contain graphical and numerical data representation and display result in easy way.

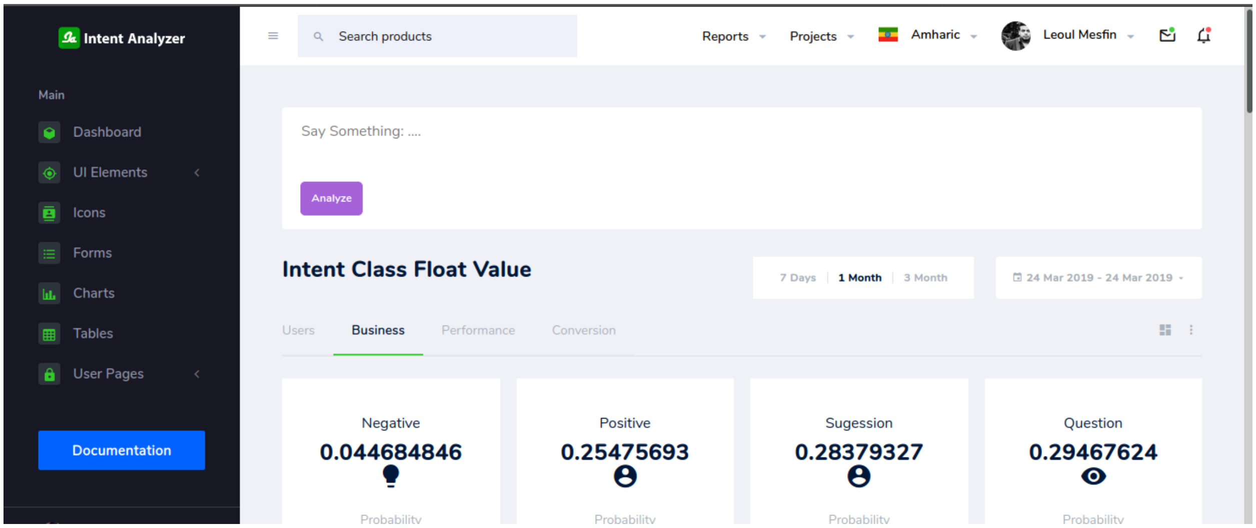


Figure 5. 27: Display home page

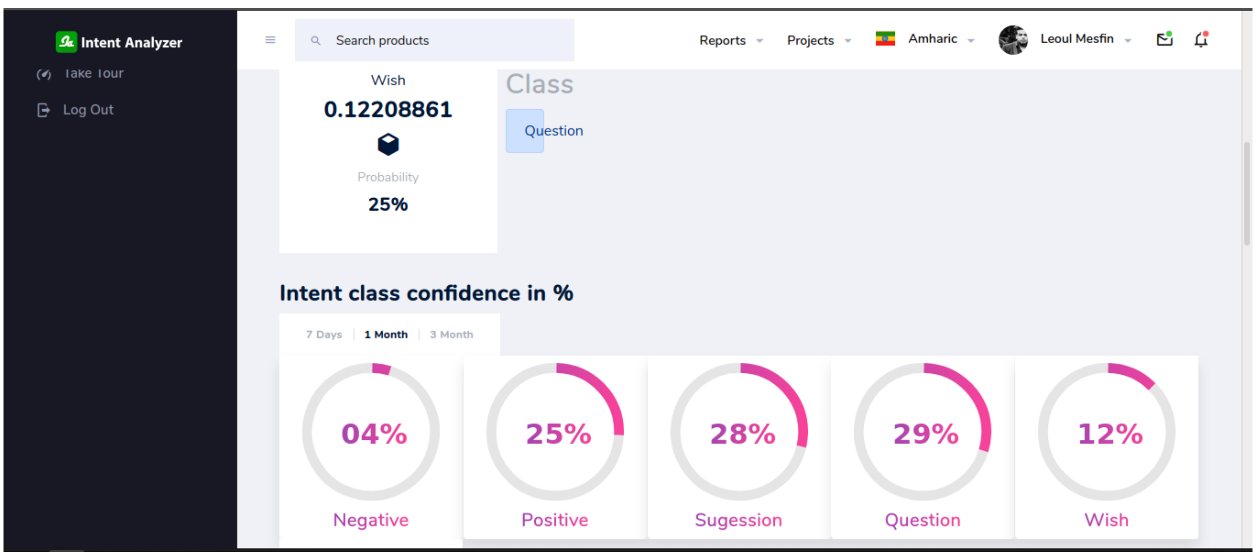


Figure 5. 28: Circular progress bar shows the result of input text

Figure 5. 29: Line chart and bar chart for training data set description on website

CHAPTER SIX

6 CONCLUSION AND FUTURE WORKS

This chapter is conclusion of observations from these research. It also contains future works to show further researches that can be done in the future.

6.1 Conclusion

The main objective of this research is to build intent analysis model and also investigate the effect of noise testing or input data affect the classification task by vary between expected label and other predefined class domains. Furthermore, its shown that is possible to achieve an accuracy of 82.75% both BI LSTM and Multinational Naive Bayes (MNB) are also perform well, MNB achieves 78%, BILSTM achieves 82%. Both architecture they have also their own drawbacks like running time BILSTM because of double cell on each gate it consumes time MNB have lack of memory management. Also they have their own features like MNB perform well if many feature selection task applied. This paper has concluded that noise test data not affect classification model.

This research collects and contribute Amharic based intent classification domain corpus for the next researches. And Develop a Long short term memory based Amharic intent analysis classification model. The data that input from the user to the model which is a proposed LSTM not only formal there might be noise data given by the user this research analyzing noise data could lead miss classification on proposed model.

6.2 Future works

Developing economical fully-functional intent analysis needs coordinated team efforts that comprise linguistic skilled, technology skilled and others that may collect additional comments from each public society and social media. Good coordination of these different professionals can result in a full functional intent analysis model and spell checker model. This research paper identifies the following direction as future work.

1. This model specifically on human intention scope, as a future work the researcher suggest that collaborate different NLP task like automatic feature selection opinion mining and other Ares integrate with automated web interface. To different extraction, classification, integration tasks perform on one model

2. It is believed that the deep neural network used can be further refined by changing the network type and architecture, to maintain the performance.

3. This research paper on intent analysis is limited to comments written in Amharic language only. However, most users write a comment by transliterating Amharic alphabets to English alphabets and others write a comment by combining Amharic language with other languages like English, Tigrigna, and Oromifa. For next works adding these languages will help in developing a full model of intent analysis.

4 When conducting our experiments, small amount of corpus for word intent analysis used. Thus large data to train classifier believed better results can be achieved in the future.

5 Research focused on only basic named entity classes (Positive, negative, suggestion, question and wish) further researches could consider additional categories of intention analyses to have complete system.

REFERENCE

Addis, D., Liu, C. M., & Ta, V. D. (2018). Printed Ethiopic Script Recognition by Using LSTM Networks. *2018 International Conference on System Science and Engineering, ICSSE 2018*, 1–6. https://doi.org/10.1109/ICSSE.2018.8519972

Agarwal, S., & Sureka, A. (2017). But i did not mean it! - Intent classification of racist posts on tumblr. *Proceedings - 2016 European Intelligence and Security Informatics Conference, EISIC 2016*, 124–127. https://doi.org/10.1109/EISIC.2016.032

Akulick, S., & Sayed Mahmoud, E. (2017). Intent Detection through Text Mining and Analysis. *Proceedings of the Future Technologies Conference (FTC), Vancouver, Canada*, *493*(November), 493–496.

Alemu, Y. (2018). *Deep Learning Approach for Amharic Sentiment Analysis University of Gondar*.

Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2018). Hate speech detection in the

Indonesian language: A dataset and preliminary study. *2017 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2017*, *2018-Janua*, 233–237. https://doi.org/10.1109/ICACSIS.2017.8355039

Alias, N., Razak, S. H. A., elHadad, G., Kunjambu, N. R. M. N. K., & Muniandy, P. (2013). A Content Analysis in the Studies of YouTube in Selected Journals. *Procedia - Social and Behavioral Sciences*, *103*, 10–18. https://doi.org/10.1016/j.sbspro.2013.10.301

Arega, Y. (2019). *Trend detection and sentiment analyses in Amharic social media*. *1*, 90.

Argaw, M. (2019). *Amharic Parts-of-Speech Tagger using Neural Word Embeddings as Features*. Addis Ababa University Addis Ababa.

Azzouni, A., & Pujolle, G. (2017). A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction. *ArXiv Preprint ArXiv:1705.05690*. http://arxiv.org/abs/1705.05690

Balodis, K., & Deksne, D. (2019). Fast text-based intent detection for inflected languages. *Information (Switzerland)*, *10*(5), 161. https://doi.org/10.3390/info10050161

Cohan-Sujay, C., & Madhulika, Y. (2012). Intention Analysis for Sales, Marketing and Customer Service. *COLING (Demos)*, *December 2012*, 33–40. https://doi.org/10.1088/0004-637X/718/1/177

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2018). Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. *ArXiv Preprint ArXiv:1801.02143*.

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2020). Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Forecasting Network-wide Traffic State with Missing Values. *ArXiv Preprint ArXiv:2005.11627*.

De, A., & Kopparapu, S. K. (2010). A rule-based short query intent identification system. *Proceedings of the 2010 International Conference on Signal and Image Processing, ICSIP 2010*, 212–216. https://doi.org/10.1109/ICSIP.2010.5697471

Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, & Mounika Bagadi, Mounika Belusonti. (2020). Text based Sentiment Analysis using LSTM. *International Journal of Engineering Research And*, *V9*(05). https://doi.org/10.17577/ijertv9is050290

Eyassu, S., & Gambäck, B. (2005). *Classifying Amharic news text using self-organizing maps*.

71. https://doi.org/10.3115/1621787.1621801

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Gebremeskel, S. (2010). Sentiment Mining Model for Opinionated Amharic Texts. *Unpublished Masters Thesis, Addis Ababa University, Addis Ababa*.

György, G. G. (2017). Using machine learning techniques for deanonymization. *Informacios Tarsadalom*, *7*(1), 72–86. https://doi.org/10.22503/inftars.XVII.2017.1.5

Hashemi, H. B., Asiaee, A., & Kraft, R. (2016). Query Intent Detection using Convolutional Neural Networks. *WSDM QRUMS 2016 Workshop*. https://doi.org/10.1145/1235

Hassan, A., & Mahmood, A. (2017). Deep Learning approach for sentiment analysis of short texts. *2017 3rd International Conference on Control, Automation and Robotics, ICCAR 2017*, 705–710. https://doi.org/10.1109/ICCAR.2017.7942788

Heikal, M., Torki, M., & El-Makky, N. (2018). Sentiment Analysis of Arabic Tweets using Deep Learning. *Procedia Computer Science*, *142*, 114–122. https://doi.org/10.1016/j.procs.2018.10.466

Jabbar, H., & Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 163–172.

Jurek, A., Mulvenna, M. D., & Bi, Y. (2015). Improved lexicon-based sentiment analysis for social media analytics. *Security Informatics*, *4*(1), 1–13. https://doi.org/10.1186/s13388-015-0024-x

Kelemework, W. (2013). Automatic Amharic text news classification: Aneural networks approach. *Ethiopian Journal of Science and Technology*, *6*(2), 127–137.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. https://doi.org/10.1007/s13748-016-0094-0

Kululo, H. W. (2020). *COLLEGE OF NATURAL SCIENCES Information filtering of social media Amharic texts Based on Sentiment Analysis COLLEGE OF NATURAL SCIENCES Information filtering of social media Amharic texts Based on Sentiment Analysis*.

Le, X.-H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. *Water*, *11*(7), 1387.

Lewis, G. (2016). Sentence Correction Using Recurrent Neural Networks. *Department of Computer Science, Stanford University*, 7.

Li, W., Liu, P., Zhang, Q., & Liu, W. (2019). An improved approach for text sentiment classification based on a deep neural network via a sentiment attention mechanism. *Future Internet*, *11*(4), 96. https://doi.org/10.3390/FI11040096

Li, Y., Sun, G., & Zhu, Y. (2010). Data imbalance problem in text classification. *Proceedings - 3rd International Symposium on Information Processing, ISIP 2010*, 301–305. https://doi.org/10.1109/ISIP.2010.47

Luong, T. Le, Cao, M. S., Le, D. T., & Phan, X. H. (2017). Intent extraction from social media texts using sequential segmentation and deep learning models. *Proceedings - 2017 9th International Conference on Knowledge and Systems Engineering, KSE 2017*, *2017-Janua*, 215–220. https://doi.org/10.1109/KSE.2017.8119461

Mihret, M., & Atinaf, M. (2019). Sentiment Analysis Model for Opinionated Awngi Text. *IEEE AFRICON Conference*, *2019-Septe*, 49–51. https://doi.org/10.1109/AFRICON46755.2019.9134016

Nigam, A., Sahare, P., & Pandya, K. (2019). Intent Detection and Slots Prompt in a Closed-Domain Chatbot. *Proceedings - 13th IEEE International Conference on Semantic Computing, ICSC 2019*, 340–343. https://doi.org/10.1109/ICOSC.2019.8665635

Nowak, J., Taspinar, A., & Scherer, R. (2017). LSTM recurrent neural networks for short text and sentiment classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10246 LNAI*, 553–562. https://doi.org/10.1007/978-3-319-59060-8_50

Olah, C. (2015). *Understanding LSTM Networks [Blog]*. Web Page. https://doi.org/10.1007/s13398-014-0173-7.2

Purohit, H., Dong, G., Shalin, V., Thirunarayan, K., & Sheth, A. (2015). Intent classification of short-text on social media. *Proceedings - 2015 IEEE International Conference on Smart City, SmartCity 2015, Held Jointly with 8th IEEE International Conference on Social Computing and Networking, SocialCom 2015, 5th IEEE International Conference* *on Sustainable Computing and Communic*, 222–228. https://doi.org/10.1109/SmartCity.2015.75

Rahi, S. (2017). Research Design and Methods: A Systematic Review of Research Paradigms, Sampling Issues and Instruments Development. *International Journal of Economics & Management Sciences*, *06*(02), 1–5. https://doi.org/10.4172/2162-6359.1000403

Rajput, N. K., Grover, B. A., & Rathi, V. K. (2020). Word Frequency and Sentiment Analysis of Twitter Messages During Coronavirus Pandemic. *ArXiv*.

Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information (Switzerland)*, *11*(4), 193. https://doi.org/10.3390/info11040193

Romanyshyn, M. (2013). Rule-Based Sentiment Analysis of Ukrainian Reviews. *International Journal of Artificial Intelligence & Applications*, *4*(4), 103–111. https://doi.org/10.5121/ijaia.2013.4410

Science, C. (2014). *Rule Based Sentiment Analysis System* (Issue July).

Shang, Y. (2017). Studies on User Intent Analysis and Mining. In *ProQuest Dissertations and Theses*. Drexel University. https://search.proquest.com/docview/2014424447?accountid=14169

Sikdar, U. K., & Gambäck, B. (2018). Named entity recognition for amharic using stack-based deep learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10761 LNCS*, 276–287. https://doi.org/10.1007/978-3-319-77113-7_22

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

TESHOME, Y. (2019). *SENTENCE LEVEL OPINION MINING FOR AMHARIC LANGUAGE*. *1*, 80.

Tilahun, T. (2014). Linguistic Localization of Opinion Mining From Amharic Blogs. *International Journal of Information Technology & …*, *3*(1), 890. http://pezzottaitejournals.net/index.php/IJITCSP/article/view/1250

Wang, J., Cong, G., Zhao, W. X., & Li, X. (2015). Mining user intents in Twitter: A semi-supervised approach to inferring intent categories for tweets. *Proceedings of the National Conference on Artificial Intelligence*, *1*, 318–324.

Wang, X., Chen, X., Tang, M., Yang, T., & Wang, Z. (2020). Aspect-Level Sentiment Analysis Based on Position Features Using Multilevel Interactive Bidirectional GRU and Attention Mechanism. *Discrete Dynamics in Nature and Society*, *2020*. https://doi.org/10.1155/2020/5824873

Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. *IEEE Access*, *6*, 13825–13835. https://doi.org/10.1109/ACCESS.2018.2806394

Wondwossen Mulugeta. (2014). A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts. *HiLCoE Journal of Computer Science and Technology*, *2*(2), 80–87.

APPENDIX

Appendix A

Normalized characters

ዐ=አ	ዖ=ኦ	ኀ=ሀ	ሓ=ሀ	ጻ=ፃ	ሤ=ሴ
ዓ=አ	ዉ=ው	ኸ=ሀ	ሔ=ሄ	ጼ: ፄ	ሥ=ስ
ዑ=ኡ	ኃ=ሀ	ሖ=ሆ	ኆ=ሆ	ጾ=ፆ	ሦ=ሶ
ዒ=ኢ	ኄ=ሄ	ኃ=ሀ	ጽ=ፅ	ሠ=ሰ	ሒ=ሂ
ዓ=አ	ኁ=ሁ	ሃ=ሀ	ጸ=ፀ	ሣ=ሳ	

ዔ=ኤ	ኂ=ሂ	ሐ=ሀ	ጹ=ፁ	ሡ=ሱ
ዕ=እ	ኄ=ሄ	ሑ=ሁ	ጺ=ፂ	ሢ=ሲ

Word Normalization

ኢቲዮፒያ=ኢትዮጵያ	ይመችክ=ይመችህ
1ኛ=አንደኛ	ኡፍፍፍፍፍፍ=ኡፍፍ
ኡፍፍፍፍፍፍፍ=ኡፍፍ	ፍልም=ፊልም
ኮመንት=አስተያየት	ፍቅርርር=ፍቅር
ሙቭ=ፊልም	ዋውውውው=ዋው
ላይክ=ዉደድ	ውድድድ:ውድድ
ውድድድድድ=ውድድ	ኢትዬ=ኢትዮጵያ

Appendix B

LSTM model

NB_WORDS = 20000 # Parameter indicating the number of words we'll put in the dictionary
VAL_SIZE =1000 # Size of the validation set
EPOCHS = 20 # Number of epochs we usually start to train with
BATCH_SIZE = 512 # Size of the batches used in the mini-batch gradient descent

df = pd.read_csv('ala.csv', encoding ='utf-16')
df.head()

```
df = df.reindex(np.random.permutation(df.index))
df = df[['comment', 'intent']]
df.head()
X_train, X_test, y_train, y_test
train_test_split(df.comment, df.intent, test_size=0.2, random_state=37)
print('Training Data:', X_train.shape[0])
print('Test Data:', X_test.shape[0])
tweet_df = df[['comment','intent']]
tk = Tokenizer (num_words=NB_WORDS,split=" ")
tk.fit_on_texts(X_train)
sentiment_label = tweet_df.intent
len(df.comment)
print(tk.word_counts)
word_index = tk.word_index
print('Found %s unique tokens.' % len(word_index))
model = models. Sequential()
model.add  (layers.Embedding(NB_WORDS, 8, input_length=MAX_LEN))
model.add(layers.LSTM(64))
model.add(layers.Dropout(0.2))
model.add(layers.Dense(5, activation='softmax'))
model.compile(loss='categorical_crossentropy',optimizer='adam', metrics=['accuracy'])
history = model.fit(X_train_emb, y_train_emb, epochs=EPOCHS, batch_size=BATCH_SIZE,
validation_data=(X_valid_emb, y_valid_emb))
def plot(history, metric_name):
    metric = history.history[metric_name]
    val_metric = history.history['val_' + metric_name]
    e = range(1, EPOCHS + 1)
    plt.plot(e, metric, 'ro', label='T raining ' + metric_name)
    plt.plot(e, val_metric, 'r', label='Validation ' + metric_name)
    plt.legend()
```

```
plt.show()
```

WEBSITE SAMPLE CODE

```
</head>
  <body>
    <div class="container-scroller">
      <!-- partial:partials/_navbar.html -->
      <nav class="navbar default-layout-navbar col-lg-12 col-12 p-0 fixed-top d-flex flex-row">
        <div class="text-center navbar-brand-wrapper d-flex align-items-center justify-content-center">
          <a class="navbar-brand brand-logo" href="index.html"><img src="{{ url_for('static', filename='assets/images/j/logos.svg')}}" alt="logo" /></a>
```

```
          <a class="navbar-brand brand-logo-mini" href="index.html"><img src="{{
url_for('static', filename='assets/images/j/logo-mini.svg')}}" alt="logo" /></a>
        </div>
        <div class="navbar-menu-wrapper d-flex align-items-stretch">
          <button class="navbar-toggler navbar-toggler align-self-center" type="button" data-
toggle="minimize">
            <span class="mdi mdi-menu"></span>
          </button>
          <div class="search-field d-none d-xl-block">
            <form class="d-flex align-items-center h-100" action="#">
              <div class="input-group">
                <div class="input-group-prepend bg-transparent">
                  <i class="input-group-text border-0 mdi mdi-magnify"></i>
                </div>
                <input type="text" class="form-control bg-transparent border-0"
placeholder="Search products">
              </div>
            </form>
          </div>
          <ul class="navbar-nav navbar-nav-right">
            <li class="nav-item  dropdown d-none d-md-block">
              <a class="nav-link dropdown-toggle" id="reportDropdown" href="#" data-
toggle="dropdown" aria-expanded="false"> Reports </a>
              <div class="dropdown-menu navbar-dropdown" aria-labelledby="reportDropdown">
                <a class="dropdown-item" href="#">
                  <i class="mdi mdi-file-pdf mr-2"></i>PDF </a>
                <div class="dropdown-divider"></div>
                <a class="dropdown-item" href="#">
                  <i class="mdi mdi-file-excel mr-2"></i>Excel </a>
                <div class="dropdown-divider"></div>
```

```
              <a class="dropdown-item" href="#">
                <i class="mdi mdi-file-word mr-2"></i>doc </a>
            </div>
          </li>
          <li class="nav-item  dropdown d-none d-md-block">
            <a class="nav-link dropdown-toggle" id="projectDropdown" href="#" data-
toggle="dropdown" aria-expanded="false"> Projects </a>
            <div class="dropdown-menu navbar-dropdown" aria-
labelledby="projectDropdown">
              <a class="dropdown-item" href="#">
                <i class="mdi mdi-eye-outline mr-2"></i>View Project </a>
              <div class="dropdown-divider"></div>
              <a class="dropdown-item" href="#">
                <i class="mdi mdi-pencil-outline mr-2"></i>Edit Project </a>
            </div>
          </li>
          <li class="nav-item nav-language dropdown d-none d-md-block">
            <a class="nav-link dropdown-toggle" id="languageDropdown" href="#" data-
toggle="dropdown" aria-expanded="false">
              <div class="nav-language-icon">
              <div class="p-3 text-center bg-primary">
                <img class="img-avatar img-avatar48 img-avatar-thumb" src="{{ url_for('static',
filename='assets/images/faces/face28.png')}}" alt="">
              </div>
              <div class="p-2">
                <h5 class="dropdown-header text-uppercase pl-2 text-dark">User Options</h5>
                <a class="dropdown-item py-1 d-flex align-items-center justify-content-between"
href="#">
                  <span>Inbox</span>
                  <span class="p-0">
```

```
          <span class="badge badge-primary">3</span>
          <i class="mdi mdi-email-open-outline ml-1"></i>
        </span>
      </a>
      <a class="dropdown-item py-1 d-flex align-items-center justify-content-between"
href="#">
        <span>Profile</span>
        <span class="p-0">
          <span class="badge badge-success">1</span>
          <i class="mdi mdi-account-outline ml-1"></i>
        </span>
      </a>
      <a class="dropdown-item py-1 d-flex align-items-center justify-content-between"
href="javascript:void(0)">
        <span>Settings</span>
        <i class="mdi mdi-settings"></i>
      </a>
      <div role="separator" class="dropdown-divider"></div>
      <h5 class="dropdown-header text-uppercase  pl-2 text-dark mt-2">Actions</h5>
      <a class="dropdown-item py-1 d-flex align-items-center justify-content-between"
href="#">
        <span>Lock Account</span>
        <i class="mdi mdi-lock ml-1"></i>
      </a>
      <a class="dropdown-item py-1 d-flex align-items-center justify-content-between"
href="#">
        <span>Log Out</span>
        <i class="mdi mdi-logout ml-1"></i>
      </a>
    </div>
```

```
    </div>
   </li>
   <li class="nav-item dropdown">
      <div class="collapse" id="ui-basic">
      <ul class="nav flex-column sub-menu">
        <li class="nav-item"> <a class="nav-link" href="{{ url_for('static',
filename='pages/ui-features/buttons.html')}}">Buttons</a></li>
        <li class="nav-item"> <a class="nav-link" href="{{ url_for('static',
filename='pages/ui-features/dropdowns.html')}}">Dropdowns</a></li>
        <li class="nav-item"> <a class="nav-link" href="{{ url_for('static',
filename='pages/ui-features/typography.html')}}">Typography</a></li>
      </ul>
    </div>
   </li>
   <li class="nav-item">
    <a class="nav-link" href="{{ url_for('static', filename='pages/icons/mdi.html')}}">
      <span class="icon-bg"><i class="mdi mdi-contacts menu-icon"></i></span>
      <span class="menu-title">Products</span>
    </a>
   </li>
      <span class="icon-bg"><i class="mdi mdi-format-list-bulleted menu-
icon"></i></span>
      <span class="menu-title">Register</span>
    </a>
    <div class="dashboard-progress dashboard-progress-2 d-flex align-items-center
justify-content-center item-parent"><i class="mdi mdi-account-circle icon-md absolute-center
text-dark"></i></div>
            <p class="mt-4 mb-0">Probability</p>
            <h3 class="mb-0 font-weight-bold mt-2 text-dark"></h3>
           </div>
```

```
          </div>
        </div>
        <div class="col-xl-3  col-lg-6 col-sm-6 grid-margin stretch-card">
          <div class="card">
            <div class="card-body text-center">
              <h5 class="mb-2 text-dark font-weight-normal">Question</h5>
              <h2 class="mb-4 text-dark font-weight-bold">{{dom4}}</h2>
              <div class="dashboard-progress dashboard-progress-3 d-flex align-items-
center justify-content-center item-parent"><i class="mdi mdi-eye icon-md absolute-center
text-dark"></i></div>
            justify-content-center item-parent"><i class="mdi mdi-cube icon-md absolute-
center text-dark"></i></div>
              <p class="mt-4 mb-0">Probability</p>
              <h3 class="mb-0 font-weight-bold mt-2 text-dark">25%</h3>
            </div>
          </div>
          <div class="col-md-3">
            <h1>Intent Class</h1>
            <div class="alert alert-primary" role="alert">
              {% if probability %}
                {{ sentiment }}
                {% endif %}
            </div>
          </div>
        </div>
```

Appendix C

Data set sample for intent classification

	A	B
1	comment	intent
2	አበያችን ኦርልን ምርጡ መሪያችን	1
3	ንገስ ደጋግመህ ደግሞ ኮራሁ በገሬ	1
4	እያችሁ ጃንሆይ ምን ያህል ሃብታም እንደ ነበሩ?	3
5	አብዛኛው ህዝቡ ግን በከፉ ድህነት ውስጥ ነበረ	0
6	የአገር ሰላም ባልተረጋገጠበት ይህን እንዲህ ግልፅ ማድረጉ ለዝርፍያ ሊያጋልጠው ይችላል	2
7	ገና ስልጣኑን በደንብ ሳይረጋጋ ይህን ሁሉ የቤተመንግስት ምስጢር ፍተሻ የገባው ምን አስቦ ነው?	3
8	እነዚህ ቅርሶች ግን ከዚህ በሃላ ለረዥም ዘመን የመቆየታቸው ነገር እየያሰሰበኝ ነው	2
9	አረ የት ሄጀ ልሣቅ ጉድ ነው	0
10	በዘመናችን እንዲንተ አይነት መልካም መሪ ስላየን በጣም እድለኞች ነን	1
11	እግዚአብሔር አብዝቶ አብዝቶ ይባርክህ	4
12	ወይኔ እምዬ ኢትዮጵያን አላውቃትም ነበር ማለት ነው?	3
13	አበቹዬ ሺ አመት ምራን	1
14	እኔ በግሌ አብይን እወደዋለሁ የሚሠራቸው ነገሮች ሁሉ የመኖር ተሥፋዎቻችን ናቸው	1
15	ግን አደዚጋ ቅሬታ አለብኝ	0
16	ፍትህ ለሚገባቸው ወገኖች ፍትህ ሥጥ ?	3
17	እደተደበቀው መኪና ነፍሠጡር ያረዱትን ወንጀለኞች ከተደበቁበት አውጣ?	3
18	ለናቶች ፍትህ ሥጥ እንባቸውን ጥረግ ያኔ ያሠብከው ይሣካል ይቀናሀል	2
19	ይህ መልካም የአዛብሄር ሰው ስሩ ሳይል ስርቶ በተግባር ያሳየ ታላቅ መሪ	1
20	ትንሽ ግዜ ካገኘ እና ሚያግዘው ሚያዳምጠው ሰው ህብረተሰብ ካገኘ ሚያቆመው ሰው አይኖርም	1
21	አይደለም ለኢትዮጵያ ለአፍሪካም አይገባትም በዱላ ለሚያምን እድሜልኩን ሲገርፍ የኖረ ዶማ ህብረተሰብ ሚያስረው ሚረግጠው ሚያደማው እንደቀድሞ አመራሮች ጨፍልቆ ሚገዛው ነው	0
22	ቀበሮዎቹ ግን ቂሮ ነገር ናቸው ከልል ይወዳሉ	0
23	ዶክተር አብይ አንተ እግዚአብሔር ለኢትዮጵያ የሰጣት ስጦታ ነህ	1
24	ቸሩ እግዚአብሔር በምታደርገው ነገር ሁሉ ይቅናህ	4
25	የተደበቀውን ታሪክ ሁሉ እንዲወጣ አደረከው ንጹህ ኢትዮጲያዊ ወንድማችን	1
26	በኦሮሚያ ክልል የተፈናቀሉት የሚያስታውሳቸው የሃገር መሪ አጥተው ስለ ዛፍ ያወራሉ	0
27	የአንተ አይነት ጥላቻን የሚዘሩ ሰዎች ባሉበት ሀገር ሁልጊዜም አገራችን ለቅሶ ቤት ናት ቂመኛ በሽተኛ ያለፈውን ረስተሽ ወደፊት ተጓዥ እኛ ለቅሶ እና ስደት ነው የሰለቸን	0
28	ዕድሜ ይስጥህ	4
29	ወይ አብቹ ስንት የተደበቀ ታሪክ አወጣክ ይገርማል ኢትዮጵያ ያኔ እና ዘንድሮ	1
30	እግዛብሄር ብርክ ያድርግህ ጌታ ይጠብቅህ	4
31	ኢግዚያብሔር ባንተ የጀመረውን ይጨርሳል በዘመኔ ያላየሁትን ያገሬን ሁበት ስላየሁ ደስ ብሎኛል	1
32	[illegible]	1