

JIMMA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
JIMMA INSTITUTE OF TECHNOLOGY (JiT)
SCHOOL OF COMPUTING
DEPARTMENT OF INFORMATION TECHNOLOGY

AMHARIC ASPECT BASED OPINION SUMMARIZATION USING BOOTSTRAP ON HOTEL DOMAIN

BY
SEID HUSSEIN

A THESIS SUBMITTED TO
THE SCHOOL OF GRADUATE STUDIES OF JIMMA UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

Jimma Ethiopia

June 15, 2019




JIMMA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
JIMMA INSTITUTE OF TECHNOLOGY (JiT)
SCHOOL OF COMPUTING
DEPARTMENT OF INFORMATION TECHNOLOGY

**AMHARIC ASPECT BASED OPINION SUMMARIZATION
USING BOOTSTRAP ON HOTEL DOMAIN**

**BY
SEID HUSSEIN**

A THESIS SUBMITTED TO
THE SCHOOL OF GRADUATE STUDIES OF JIMMA UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

Name and Signature of Members of the Examining Board

Title	Name	Signature	Date
1. Advisor	Dr. Debela Tesfaye		June 15, 2019
2. External Examiner	Dr. _____	_____	June 15, 2019
3. Internal Examiner	_____	_____	June 15, 2019
4. Chairperson	_____	_____	June 15, 2019

DEDICATION

Dedicated to my Mather Fatuma Adem and my Mather country Ethiopia!
መታሰቢያነቱ ለወድ እናቴ ፋጡማ አደም እና ለእናት ሀገሪ ኢትዮጵያ
ይሁን!

Declaration

I declare that the work described in this thesis is entirely my own. No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or institute. Any help or source information, which has been availed in the thesis, has been duly acknowledged.

Signature

SEID HUSSEIN

Department of Information Technology,

Jimma Institute Technology (JiT),

Jimma University,

Jimma, Ethiopia

Acknowledgment

First of all, I would like to almighty thank for Allah for giving me the strength to achieve whatever i have achieved so far and for guiding me all the way through. Next to Allah, there are people without whose help this thesis would not have come into being fruitable. I would like to take this opportunity to thank them.

My foremost gratitude goes to my advisor Dr. **Debela Tesfaye** for his consistent follow-up and his willingness to offer me his time and knowledge from the inception to the completion of this thesis. Next I also want to thank Mr. **Teferi K.** for his supportive ideas when we had time to discuss together. I would like to extend my special thanks to many of my friends who have helped me in this thesis work. Especially, many thanks to my classmates: **Smegnew A., Gashaw Z., Abdulkadir A., Mohammed N., Tesfu M., Abaynew G., Mathews B., Nigusu Y., Workneh** and my colleague: **Tefera A., Wegderes T.,** and his friend **Damana D.** Really Damana help me more to do codes of the system.

In addition, I would like to thank all staff members of Jimma University, School of Computing, for your kind help during my stay especially for Dr. Getachew M., Mr. Ymam N., Salahadin S., Seid y., Dr. Melita and Dr. Dipack.

Finally, my thank goes to everyone who has contributed negative and positive impacts to the successful realization of this thesis work, who are not mentioned in name but whose support helped me complete the study successfully. Thanks for all.

(SEID HUSSEIN)

Table of contents

TABLE OF CONTENTS	V
LIST OF FIGURES	VIII
LIST OF TABLES	VIII
LIST OF ACRONYMS	IX
ABSTRACT	X
CHAPTER ONE.....	- 1 -
1.1 Introduction.....	- 1 -
1.2. Major Contributions	- 2 -
1.3. Statement of the Problem	- 2 -
1.4. Objective of the research.....	- 3 -
1.4.1. General objective.....	- 3 -
1.4.2. Specific objectives.....	- 3 -
1.5. Methodology and Design of the Study	- 4 -
1.5.1 Literature Reviews.....	- 4 -
1.5.2 Design of the Study	- 4 -
1.5.3 Collect Opinions	- 5 -
1.5.4 Amharic Opinionated Text Analysis	- 5 -
1.5.5. Building Bootstrap with Seed opinion word lexicon.....	- 5 -
1.5.6 Implementation Tool	- 5 -
1.5.7 Evaluation Mechanism	- 6 -
1.6. Scope and Limitation of the research	- 6 -
1.6.1. Scope	- 6 -
1.6.2. Limitation	- 7 -
1.7. Significance of the Research	- 7 -
1.8. Thesis Organization.....	- 8 -
CHAPTER TWO.....	- 9 -
LITERATURE REVIEW.....	- 9 -
2.1. Introduction.....	- 9 -
2.2. Opinion summarization	- 10 -
2.2.1. Non Aspect based opinion summarization	- 11 -

2.2.2. Aspect based opinion summarization	- 11 -
2.2.2. 1. Aspect/Feature Identification	- 12 -
2.2.2.2. Sentiment Prediction	- 14 -
2.2.2.3. Summarization Methods	- 16 -
2.3. Amharic language study	- 17 -
2.3.1. Amharic Word Structure	- 18 -
2.3.2. Amharic opinion Lexical category	- 19 -
CHAPTER THREE	- 22 -
RELATED WORK TO OPINION SUMMARIZATION	- 22 -
CHAPTER FOUR	- 27 -
RESEARCH METHODS AND DESIGN	- 27 -
4.1 The proposed Architecture	- 27 -
4.1.1 Amharic Opinions (Reviews)	- 27 -
4.1.2 Preprocessing Phase	- 28 -
Tokenization	- 29 -
Normalization	- 29 -
4.1.3 Aspect Identification and Opinion Learner	- 30 -
4.1.3.1 Identifying Aspects	- 30 -
4.1.3.2. Extracting /Learn opinions	- 31 -
4.1.3.3 Polarity Defining	- 33 -
4.1.4 Opinion word Seed Lexicon	- 34 -
4.1.5 Prediction the Aspect and Opinion Pair	- 34 -
4.1.6 Aspect Based Amharic Opinion Summarization by Graph	- 35 -
4.2 Algorithm	- 36 -
4.2.1. Bootstrapping	- 36 -
4.2.2 Naive Bayes Classifier	- 40 -
4.3. Review data collections	- 41 -
CHAPTER FIVE	- 44 -
EXPERIMENTAL RESULTS AND EVALUATION	- 44 -
5.1 Introduction	- 44 -
5.2 The Research Implementation	- 44 -
5.2.1 Tools	- 44 -
5.2.2 Graphical User Interface	- 45 -

5.3. Testing Environment	- 47 -
5.4. Evaluation Metrics	- 47 -
5.4.1. User-Centered Evaluation.....	- 48 -
5.4.2. System- Centered Evaluation	- 50 -
5.5. Experimental Results and Discussions	- 50 -
5.5.1 Experiments and Their Result	- 50 -
5.5.2. Experiments.....	- 51 -
5.5.3 Discussions about the Results	- 53 -
5.5.4. Training Corpus preparation for Naïve Bayes Classification.....	- 54 -
CHAPTER SIX	- 57 -
6. CONCLUSION AND RECOMMENDATION	- 58 -
6.1 Conclusion.....	- 58 -
6.2. Recommendation as Future Work.....	- 60 -
REFERENCES	- 62 -
APPENDIX	- 67 -
A. Sample of Amharic Review on Hotel Domain Collected From Customer.....	- 67 -
B. Sample of result visualization with graph.....	- 68 -
C. Sample of source code Amharic Aspect Opinion Summarization Using Bootstrap on Hotel Domain	- 69 -
D. List of positive and negative seed opinion word lexicons	- 73 -

List of Figures

FIGURE 1: <i>GENERAL ARCHITECTURE OF AMHARIC ASPECT OPINION SUMMARIZATION</i>	- 28 -
FIGURE 2: BOOTSTRAPPING ALGORITHM FOR AMHARIC ASPECT OPINION SUMMARIZATION ON HOTEL DOMAIN	- 37 -
FIGURE 3: DRAW BAR CHART CODES FOR ASPECT BASED SUMMARIZING	- 39 -
FIGURE 4: SUPERVISED NAIVE BAYES BASED ASPECT BASED SUMMARIZATION	- 40 -
FIGURE 5: SAMPLE OF AMHARIC OPINIONS	- 42 -
FIGURE 6: ASPECT BASED AMHARIC OPINION SUMMARIZATION GUI BEFORE GIVING INPUT OPINIONS	- 46 -
FIGURE 8: ASPECT BASED AMHARIC OPINION SUMMARIZATION GUI AFTER INPUT & EXECUTE OPINIONS	- 47 -
FIGURE 9: SUMMARIZE RESULTS BY BAR GRAPH BOOTSTRAP	- 51 -
FIGURE 10: SYSTEM CENTERED EXPERIMENT RESULT OF BOOTSTRAP APPROACH	- 53 -
FIGURE 11: UI AND ENTERING REVIEWS FOR NAIVE BAYES	- 54 -
FIGURE 12: GRAPHICAL REPRESENTATIONS OF NAIVE BAYES RESULT	- 55 -

List of Tables

TABLE 1: NORMALIZATION OF AMHARIC	- 30 -
TABLE 2: ADJECTIVES FORMED BY ADDING PREFIXES	- 33 -
<i>TABLE 3: SELECTED SEEDS WORD</i>	- 38 -
TABLE 4: ASPECTS AND THEIR PROPORTION NUMBER OF COLLECTED OPINIONS	- 42 -
TABLE 5: QUESTIONS AND ANSWER FOR USER CENTERED EVALUATION	- 49 -
TABLE 6: BOOTSTRAP SYSTEM CENTERED EXPERIMENT RESULTS	- 52 -
TABLE 7: CONFUSION MATRIC FOR GENERAL SYSTEM PERFORMANCE TESTING	- 52 -
TABLE 8: COMPARISON OF OUR WORK	- 56 -

List of Acronyms

ABAOS	Aspect Based Amharic Opinion Summarization
GUI	Graphical User Interface
UTF	Unicode Transformation Format
JSON	JavaScript object notation
IDE	NetBeans is an Integrated Development Environment
IR	Information Retrieval
IE	Information Extraction
MAP	Mean average precision
NLP	Natural language processing
MLP	Multilayer Perception
POS	Part of speech tagging.
TF-IDF	Term Frequency and inverted document frequency
WWW	World Wide Web
ወ/ሮ	ወይዘሮ
ደ/ብርሃን	ደብረ ብረሀን
ኢ/ፌ/ደ/ሪ	የኢትዮጵያ ፌዴራላዊ ዲሞክራሲያዊ ሪፐብሊክ

Abstract

Over the last few years, this special task of summarizing opinions has stirred tremendous interest amongst the NLP and Text Mining communities. The simplest form of an opinion summary is the result of sentiment prediction. Aspect-based summarization divides input texts into aspects, which are also called as features and subtopics, and generates summaries of each aspect. Today millions of web-users express their opinions about many topics through blogs, wikis, web, chats and social networks. Especially sectors such as e-commerce and e-tourism, it is very useful to automatically analyze the huge amount of social information available on the Web, but the extremely unstructured nature of these contents makes it a difficult task. One of the main reasons for the lack of study on opinions is the fact that there were little opinionated texts available before Web. Today millions web-users express their opinions about many topics through blogs, wikis, web, chats and social networks. The objective of this research is to design and develop aspect opinion summarization for opinionated Amharic documents. The major components of this work are Amharic opinions/reviews, preprocessing phase, aspect and opinion learner, opinion word Seed Lexicon, polarity defining and Aspect based Amharic opinion summarization by graph. The Average performance of user center evaluation for our aspect based Amharic opinion summarization system for bootstrapping is 91.38% or 4.569 out of 5 weigh. In system centered evaluation the semi supervised bootstrap method achieved the averages of positive and negative effectiveness 92% precision, 72.8% recall and 81.40% F-measure. Also in the naïve Bayes hotel review classification approach average performance test result is that 75.68% precision, 87.05% recall and 77.86% F-measure less than bootstrapping. Up to now there is no systems that digesting those huge amount of customer opinions given in Ethiopic (Amharic language) for understanding opinion holder (customers) need on a given domain particular entity. Therefore in this work Ethiopic (Amharic language) customer opinions on a hotel domain was summarize with respect to their aspects /features by graph visualization with the performance of above. This work will help for organization (such as hotels), individual (such as hotel users), government intelligence, and business intelligence.

Keywords: Aspect opinion Summarization, Hotel aspect, Seed opinion word lexicon, bootstrap, naïve Bayes.

Chapter One

1.1 Introduction

Over the last few years, this special task of summarizing opinions has stimulated tremendous interest amongst the Natural Language Processing (NLP) and Text Mining communities. ‘Opinions’ mainly include opinionated text data such as blog/review articles, and associated numerical data like aspect rating is also included. While different groups have different notions of what an opinion summary should be, we consider any study that attempts to generate a concise and digestible summary of a large number of opinions as the study of Opinion Summarization. The simplest form of an opinion summary is the result of sentiment prediction (by aggregating the sentiment scores). The task of sentiment prediction or classification itself has been studied for many years. Beyond such summaries, the newer generation of opinion summaries includes structured summaries that provide a well-organized break down by aspects/topics, various formats of textual summaries and temporal visualization. There are many ways to use the mining results. One simple way is to produce a feature-based summary of opinions on the object [1]. The different formats of summaries complement one another by providing a different level of understanding. For example, sentiment prediction on reviews of a product can give a very general notion of what the users feel about the product. If the user needs more specifics, then the topic-based summaries or textual summaries may be more useful. Regardless of the summary formats, the goal of opinion summarization is to help users digest the vast availability of opinions in an easy manner. The approaches utilized to address this summarization task vary greatly and touch different areas of research including text clustering, sentiment prediction, text mining, NLP analysis, and so on. Some of these approaches rely on simple heuristics, while others use robust statistical models [2].

Aspect-based summarization divides input texts into aspects, which are also called as features and subtopics, and generates summaries of each aspect. For example, for the summary of ‘iPod’, there can be aspects such as ‘battery life’, ‘design’, ‘price’, etc. By further segmenting the input texts into smaller units, aspect-based summarization can show more details in a structured way. Aspect segmentation can be even more useful when overall opinions are different from opinions of each aspect because aspect-based summary can present opinion distribution of each aspect separately. The aspect-based approaches are very popular and have been heavily explored over

the last few years [1]. A standard setting for sentiment summarization assumes a set of documents $D = \{d_1, d_2, d_3 \dots d_m\}$ that contain opinions about some entity of interest. The goal of the system is to generate a summary S of that entity that is representative of the average opinion and speaks to its important aspects [3].

1.2. Major Contributions

This thesis proposes a domain specific of Amharic Aspect-Opinion summarization Using Bootstrap as well as naive Bayes classification in Hotel Domain. This work aims to automatically learn extraction hotel customer opinion from linguistic analysis of opinionated text, taking hotel domain reviews as input. The mapping between the linguistic arguments and the target pattern arguments is specified automatically. The learning method and its setup are general enough to enable the adaptation to new domains and new tasks.

Our system is highly scalable and adaptable with respect to new domains and relations of different complexity. The scalability and adaptability starts with the decision of taking the relation instances as seed for the bootstrapping-based learning and training data for naïve Bayes. The relation instances are sample of the target association pattern relations defined by the user. Thus, the learning process is driven by the target pattern structures. The seed opinion word's helps us identify the explicit Amharic Opinionated text. An interesting learning including an empirical investigation analyzes the influence of the seed opinion words on the learning performance, considering under specification and over specification of the seed opinion words and size of opinions training data set.

1.3. Statement of the Problem

Today millions of web-users express their opinions about many topics through blogs, wikis, web, chats and social networks. For sectors such as e-commerce and e-tourism, it is very useful to automatically analyze the huge amount of social information available on the Web, but the extremely unstructured nature of these contents makes it a difficult task. One of the main reasons for the lack of study on opinions is the fact that there were little opinionated texts available before the World Wide Web. Before the Web, when an individual needed to make a decision, he/she typically asked opinions from friends and families. When an organization wanted to find the opinions of humans about its products and services, it conducted opinion

polls, surveys, and focus groups. However, with the Web, especially with the explosive growth of the user generated content on the Web in the past few years, the world has transformed [33] [34] [37]. This above entire situation is also parallel true to Amharic language (Ethiopic) users.

As online business is becoming more and more popular, the quantity of reviews toward products given by customers is growing rapidly as well. Hence it is difficult for a customer, seller or the producer to read all of customer reviews and then make a reasonable decision when she/he is facing the problem whether to purchase a certain product, use certain service or not [4]. Due to the availability of opinion rich documents on review sites, forums, discussion groups, blogs, social networks such as Facebook, etc., there are too many opinions and reviews to be read which is very difficult and hence traditional opinion polls, survey, and focus group techniques are inadequate, time consuming and hard. So there is a need for good sampling and classification techniques for these reviews and opinions. For this reason many researches on opinion or sentiment analysis have been done and are being under taken for English and other Latin languages such as Germany, French [6] and Chinese language [40]. Therefore, this study investigates and aims to develop an Aspect based Amharic opinion summarization for opinionated Amharic texts on the hotel domain with answering and addressed the following research questions.

- How we can prepare Amharic language (Ethiopic) hotel customer reviews?
- How to learn opinion words from hotel reviews?
- To what extent our Aspect based Amharic opinion summarization prototype is performs?

1.4. Objective of the research

The general and specific objectives of this study are given below:

1.4.1. General objective

The general objective of this research work is to design and develop aspect based Amharic opinion summarization using bootstrapping for opinionated Amharic hotel reviews.

1.4.2. Specific objectives

To achieve the above objective, the following specific objectives will be done:

- Review literatures in the area of opinion mining, opinion summarization, aspect based summarization and different approaches needed to solve it.
- Study the nature and characteristics of Amharic language such as word structure and lexicon category
- Build bootstrapping with selected seed opinion Amharic words Lexicon.
- Collect hotel reviews for the performance evaluation.
- Adopt Bootstrap algorithm as well as Naive Bayes classification for extraction of opinion in hotel domain.
- Develop a model for Amharic Aspect opinion summarization Using Bootstrap
- Evaluate performance of the model

1.5. Methodology and Design of the Study

The methodology and design of Amharic Aspect Opinion Summarization Using Bootstrap on Hotel Domain is described as following to achieve the objectives of this work.

1.5.1 Literature Reviews

Sentiment summarization and Opinion mining related literatures from different sources such as published papers, journal articles and other materials are reviewed in detail to get better understanding of the area and to have detail knowledge on the various techniques of sentiment summarization specially bootstrapping. Subsequently this research work is mainly concerned with Amharic opinionated summarization on hotel domain, it was essential to examine the nature of Amharic documents that contain opinions especially in hotel domain.

1.5.2 Design of the Study

This Amharic Aspect Opinion Summarization Using Bootstrap on Hotel Domain work is done based on the idea of empirical experimental ways of research. As we describe in above section; it concentrated on Amharic opinion summarization by examine nature of opinion word, we design architecture, rule and prototype. Therefore rules and methods were proposed to identify or categorize Amharic opinion terms and summary them. To identify Amharic opinionated words and check them; it was mandatory to analyze the nature of Amharic documents that contain opinions. Therefore rules and methods were proposed to identify or categorize Amharic opinion terms given to hotel domain. The research design is depending on seed lexicon of Amharic opinion terms. The seed lexicon that we were built contains Amharic opinion terms has polarity

positive and negative. The purpose of Amharic opinionated seed word lexicon contains opinion terms of Amharic language terms which are domain specific only on hotel domain which also within specific nine hotel aspects.

1.5.3 Collect Opinions

Amharic opinions or dataset or also called reviews are collected manually from opinion holders in hotel domain. The dataset is collected from five hotels on nine aspects. The domain is selected by the reason that there is better availability of a number of opinion holders in this environment which could be volunteer to give/write their opinions about a hotel that they were use it.

1.5.4 Amharic Opinionated Text Analysis

In Amharic language a number of opinions are posted on different electronic media nowadays. To digest this huge number of user generated opinion text data and take the reasonable discussion making, we must be analysis Amharic opinion text. Therefore rules and methods will be proposed to identify and categorize opinions on the features of the objects. We will consult linguistic professional for better understanding on the relationship between features and opinions in Amharic text and take out patterns getting from our bootstrap method.

1.5.5. Building Bootstrap with Seed opinion word lexicon

Amharic opinion words Polarities are important for our Aspect-opinion summarization in hotel domain. Thus, the newly extracted opinion words should be assigned with polarities. We now propose a polarity assignment method based on the related evidence. We will prepare the seed opinion word lexicon having two polarity ‘positive’ and ‘negative’. The seed opinion word lexicon contains specific number up to 10 seed opinion word lexicons. See at appendix D. Therefore given the hotel reviews, the task is to find the hotel features and opinion words. As observed, the opinion words mostly appear around the features in the review sentences. They are highly dependent on each other. So we adopt bootstrapping method to find opinion words. We use selected seed opinion word lexicon for this bootstrap approach of Amharic Aspect opinion summarization in hotel domain.

1.5.6 Implementation Tool

In order to achieve our objective, we used different environments and tools. We use NetBeans IDE 8.0.1 for implementing the algorithms and design the graphical User Interface

(GUI), Text Editor to process our Amharic opinionated text document, MS Word for document writing and jfreechart-1.0.19 jar files for the purpose of drawing opinions in graph.

1.5.7 Evaluation Mechanism

To evaluate our Aspect based Amharic opinion summarization; we use user centered and system centered performance evaluation mechanism. We evaluate by humans after giving the system for usage with enough distributions to them. In the system centered evaluation tool, we evaluate with Precision, Recall, and F-measure which are the evaluation parameters of information retrieval (**IR**), which are also used in text classifications. Precision measures the exactness of a classifier. Precision is the ratio of the number of reviews classified correctly to the total number of reviews in a given category. A high precision means less false positive, while a lower precision means more false positives. Recall measures the completeness or sensitivity of a classifier. A high recall means less false negative, while lower recall means more false negatives. There is trade-off between precision and recall. Greater precision decreases recall and greater recall leads to decreased precision. The F-measure is the harmonic mean of P and R and takes account of both the measures.

1.6. Scope and Limitation of the research

1.6.1. Scope

Opinion summarization is a complex and recent growing [4] research discipline that requires the effective analysis and processing of documents. Since there are no publicly available Natural Language Processing (NLP) tools and other resources for Amharic language that can be integrated with our Aspect based Amharic opinion summarization model, the scope of our research work is:

- Limited to Amharic opinion sentences.
- We use domain specific hotel review texts that are grammatically checked and organized.
- Develop Amharic language opinion lexicon with having polarity positive and negative words.
- Extract hotel aspects that opinion is given to it and classify opinions as positive and negative
- Design summary in graphical form for up to nine hotel aspects.

1.6.2. Limitation

Although certain of Amharic NLP tools have been done by some researchers, they are not publicly available for use. Because of this reason and time constraint, the following are limitations of our research work:

- Opinion spam detection, opinion holder identification, and fake reviews detecting, aren't covered in our aspect based Amharic opinion summarization research.
- Because of complicated nature of Ethiopic (Amharic language) expressions such as ግጥም, ተረት እና ምሳሌ, ቅኔያዊ አነጋገር (ሰም እና ወርቅ), ቃለ አጋኖ, and other type of Amharic expressions are not handling with our work.
- Word sense disambiguate of Ethiopic (Amharic language) is not handled by this system.
- The system can't be used to answer opinion summarization questions.

1.7. Significance of the Research

As described in [89] Due to the rapid increase of Internet, web Amharic opinion documents dynamically emerge which is useful for both potential customers and product manufacturers for prediction and decision purposes. Therefore knowing these Amharic language opinions plays an important role in decision making processes involving regular customers to executive managers. Aspect-based Opinion Summarization is one of summarization techniques which provide brief yet most relevant information about different features related to the target product. Hence, the Amharic opinion summarization model can be used for different purposes. In general the study has a vast value for hotel domains by taking customer reviews on nine hotel aspects to make efficient and effective decision. Some of significances are:

- Improve hotels by understanding customers like and dislike from reviews.
- Make hotels competitive in hotel business world by taking customer opinions.
- Aspect based Amharic opinion summary can be saving efforts and time by helping the hotel to find which hotel aspect will be improved.
- Customers can get better summarized information about hotel in general or about each aspect of hotels

1.8. Thesis Organization

The report of this thesis is organized as follow. The report is consisting six chapters with appropriate descriptions. Chapter one introduce statement of the problems, objective, scope, limitation, and significance of the study.

Chapter two introduces an overview of opinion mining (sentiment mining) and different techniques used in sentiment mining researches. Moreover, the general steps in sentiment mining are also discussed in this chapter. Chapter three presents' reviews of related researches conducted on Aspect based opinion summarization. In this chapter, in depth reviews of researches done on Aspect based opinion summarization using different techniques for different languages is presented. Chapter four describes the general architecture of the proposed model for Aspect based Amharic opinion summarization model and the construction of Amharic language (Ethiopic) lexicon.

Also implementation related issues to our Aspect based Amharic opinion summarization system are explained in this section. Chapter five presents the experimental results of the proposed model and evaluates the Aspect based Amharic opinion summarization model. Finally, recommendations and conclusions are given in the last sixth chapter.

Chapter Two

Literature Review

2.1. Introduction

An opinion is the personal state of individuals, and as such, it represents the individual's ideas, beliefs, assessments, judgments and evaluations about a specific item/subject/topic and those opinions collected from individuals have a great impact on and provide guidance for individuals, governments, social communities and organizations in the decision-making process [5]. An opinion is defined as a quintuple, $(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$ where e_i is the name of an entity, a_{ij} is an aspect of e_i , oo_{ijkl} is the orientation of the opinion about aspect a_{ij} of entity e_i , h_k is the opinion holder, and t_l is the time when the opinion is expressed by h_k [6]. The opinion orientation oo_{ijkl} can be positive, negative or neutral, or can be expressed with different strength/intensity levels. When opinion is about entity itself as a whole, Bing Liu use the special aspect GENERAL to denote entity as he point to in his survey. Opinion mining has been an emerging research field in Computational Linguistics, Text Analysis and Natural Language Processing (NLP) in recent years. It is the computational study of people's opinions towards entities and their aspects [7], [6]. Entities usually refer to individuals, events, topics, products and organizations. Traditional market research methods such as interviews, surveys, and observations usually investigated customers' opinions from sample groups instead of a target consumer population. However, following the emergence of the internet, online communication has become a common channel between customer and company, and has changed markets and organizations [5], [11]. Aspects are attributes or components of entities. In the last few years, social media has become an excellent source to express and share people's opinion on entities and their aspects. With the availability of vast opinionated web contents in the form of comments, reviews, blogs, tweets, status updates, etc. it is harder for people to analyze all opinions at a time to make good decisions. So, there is a need for effective automated systems to evaluate opinions and generate accurate results [8]. Two major tasks in aspect based opinion mining are aspect extraction and aspect sentiment classification. Process of identifying the opinion words from the given sentence is called aspect extraction and categorizing the extracted opinion words into one of the polarity scales is called aspect sentiment classification. People express opinions either

implicitly (Indirect) or explicitly (Direct). Extracting implicit aspect expression is a difficult task since opinion words differs from people to people [8].

2.2. Opinion summarization

Because of extra people post their opinions on the web; the Internet is becoming a fashionable and active source for people opinions these days [9]. Those opinions posted by peoples on the web are increasing within day to day. Therefore opinion mining, sentiment analysis and summarization become a serious necessity. Due to the large volume and wide range of opinionated text data, also there is a growing need to summarize those opinionated documents for decision making. In general Summarization is a way of presenting large amount of information using limited words still maintaining its meaning and relevancy. In the same fashion opinion summarization show a summary for large number of opinionated sentences [12]. Kim and his colleague stats that Opinion summarization can be performed at three levels of granularity at document level, sentence level or at aspect level [2]. For document level mining, a document is considered as a single entity to be observed. Similarly for sentence level mining, a single sentence and for aspect level mining, different aspects/features of an entity are taken into consideration.

Opinion summarization, also known as sentiment summarization[10], and it can take many forms. The main aim of opinion summarization is to generate a sentimental summary on opinions in a text and it has been illustration more attention recently in NLP due to its significant contribution to various applications[8] [9]. From those previous studies on extracting opinion summaries, most of them focus on text reviews, such as product reviews[12] [13], movie reviews [14] [15], hotel and restaurant review[16] [17] and aspect-based opinion mining in tourism products reviews[18] and few of them also on conversation [7] as well as on speech[91]. Nataliaia discusses the most comprehensive review of available opinion summarization options by dividing in to two big groups: aspect-based and non-aspect based in her thesis work [19]. According to Nataliaia; Non-aspect based opinion summarization have several types like text summarization, entity-based summary, basic sentiment classification and all types of visualization. Basic sentiment classification is about summarizing only sentiment and presenting it in aggregated form with polarity, percentage or polarity plus intensity. It doesn't suit our need to improve textual summarization. In contrastive summarization, which is other popular way

where summary consists of two parts with positive and negative opinions respectively. It suits well for contradictory topics with two possible points of view both having meaning to the summary consumer [8]. Another big opinion summarization group is Aspect-based summarization that is the most common type of opinion summarization and our intention is here. Aspect based opinion summarization is also known as subtopics or features. The goal is to extract sentiment for set of aspects. Bellow this we try to describe more about these two opinion summarizations in detail.

2.2.1. Non Aspect based opinion summarization

Warih Maharani et al.[20] Called traditional-based summarization which is a technique creates a generalized summary over any target without considering its aspects or features [21]. In Kim et al. [2] survey, non- aspect based opinion summarization methods are generalized into four main ways which are Basic Sentiment Summarization, Text Summarization, Visualization and Entity-based summary. Basic sentiment summarization shows the overall opinion of input data set without identification aspects using sentiment classification prediction results. Also Kim et al.[2]; show that text summarization would be handled in different methods. Opinion integration, Contrastive Opinion Summarization, Abstractive Text Summarization, and Multi-lingual Opinion Summarization are text summarization methods. Visualization is useful and need to obtain better intuition to the summarization results. In the survey [2] finally entity based summary shows the entities in text and their relationships with opinion polarity annotations. In work [20] non aspect based opinion summarization methods are similar to text summarization methods by extracting the most important sentences in the document unlike aspect based opinion summarization below ([section 2.2.2.](#)).

2.2.2. Aspect based opinion summarization

One opinion from a single person is usually not sufficient for action. Then most opinion mining applications need to study opinions from a large number of opinion holders. Due to this reason some form of summary of opinions is needed. The quintuples definitions of Opinion in section [2.1](#) above provide an excellent source of information for generating both qualitative and quantitative summaries. A common form of summary is based on aspects[22], and is called aspect-based opinion summary. Aspect-based opinion summarization which is the most common type of opinion summarization technique [2] [8][23][24] generates summaries of opinions for the

main aspects of an object or entity. Objects could be services, products, organizations (e.g., a Smartphone), and aspects are attributes or components of them. Kim et al. [2], elaborates that aspects are usually arbitrary topics that are considered important in the text being summarized such as the battery or the screen for a Smartphone. An automatic system of aspect-based opinion summarization receives as input a set of opinions about an object and produces a summary that expresses the sentiment for some relevant aspects. Aspect based summarization is pass through three different steps as Kim et al. [2] discusses in their survey, which are aspect identification, sentiment prediction, and summary generation. Liu et al. [25] also performs these three tasks in customer reviews summarization. Different researchers use different approaches for aspect or feature summarization according to their work. As evidence Kurian N and Asokan S. [9] Opinion Summarization consist of five steps which are identifying the important aspects, identifying the opinion words, classifying the sentences containing aspects according to their polarity and generating the actual summary. As we point in the above the most common type of opinion summarization technique is aspect-based opinion summarization. It involves generating opinion summaries around a set of aspects or topics (also known as features) and those aspects are usually arbitrary topics that are considered important in the text being summarized [23]. Generally, aspect-based summarization is made up of three separate steps which are aspect/feature identification, sentiment prediction, and summary generation [2], [19], [20], [26]. Some approaches, however, integrate some of the three steps into a single model. So now her let's see about the three common steps bellow.

2.2.2. 1. Aspect/Feature Identification

Aspects identification was study for the first time by M. Hu and B. Liu [25] as stated [70]. Aspects or features are an opinion targets that are related to the topic under discussion. That means aspect are the subject of review on which user comments such as for a product like “mobile” its aspects would be “battery”, “display”, “camera”, etc. Extract and find important topics in text which will be used to summarization are the main target of this step. Feature extraction is to automatically identify aspects or features of products mentioned in the users' opinions. To developing automatic feature extraction hot motivations offered in past works. Due to two main reasons some may argue that the product features can be obtained from manufactures[27]. Those two causes are customers may comment on unexpected features that manufactures have never thought about and the terminology used by manufactures may be

different to the terms used by customers. Aspects are usually named entities or noun phrases frequently mentioned. Her taking to account that all nouns are not well be aspects [28]. Implicit and explicit aspects are two main types of aspects [29], [30], [31]. Explicit aspects are concepts that clearly denote targets in the opinionated sentence posted by users. Explicit aspects are occurs in the sentence. For example “The camera of this phone is good quality” here the sentence clearly defines about the good camera quality of phone. So “camera” is aspect of product phone. Implicit aspects are an aspect that can be expressed indirectly. As evidence, in the sentence “It’s too large to handle” does not give a clear view that it is describing about the large size of the camera.

To identify aspects Kim et al. [2] discusses two main approaches’ which are NLP technique and data/text mining based approach. Minghui et al. [32] use the OpenNLP toolkit to perform chunking and obtain noun phrases and the Stanford NER tagger to identify named entities from the posts. For feature discovery in the opinion text many approaches [25], [28], [33] attempt with the help of NLP based techniques. POS (Part-of-speech) tagging and syntax tree parsing are very common starting points for feature identify. As an example, as aspects are usually noun phrases, still basic POS tagging allow people to find candidate aspects. Aspects also identify with parsing in the case of short comments [64]. Short phrases are used to express the most opinions in the case of short comments like an example ‘well studied’ and ‘excellent seller’[2]. If parsing and POS tagging are well studied and many state-of-the-art parsers and taggers are known to have high accuracies those NLP approaches are quite effective for aspect extraction. The weaknesses of these NLP techniques are the practicality and may not be sufficient in discovering all the features. Practically the speed of parsing or tagging is still not ‘efficient’ enough for large scale processing and because of features are not always nouns, and often times they are not explicitly specified in the text, NLP techniques are may not sufficient for discovering aspects or features.

The other method that used to identify feature or aspects is data mining technique. Minqing Hu and Bing Liu[25] uses association mining algorithm to identify frequent aspects (which are itemset) in addition to POS tagging. In this case itemsets are a set of words or a phrase that occurs together within in some given sentences. The weaknesses of pureNLP-based techniques can be compensated by frequent itemset mining as Hyun Duk Kim and his colleagues’ indicated by their survey in 2011[2]. Frequent itemset mining technique is an approach that do not restricts

only certain types of words or phrases which can become candidate features. This approach used support information to determine a particular word or phrase is an aspect/feature or not. This approach to aspect/feature discovery shows reasonable performance especially with product reviews [2]. According to Minqing Hu and Bing Liu [25] customer review contain many things that are not directly related to product features and different customers have different stories. But when customers comment on product features, the words that they use meet. As a result Minqing Hu and Bing Liurun the association miner based on Apriori algorithm to find frequent itemsets is appropriate. Because those frequent itemsets are likely to be product features except some candidate frequent features are not generated with association mining and those unlikely features are removed by compactness pruning and redundancy pruning. Infrequent features are not hot like frequent features. These features are talked by only a small number of people having interest potential to some customers as well as manufacturer of the product.

2.2.2.2. Sentiment Prediction

After feature identified sentiment prediction is the next stage to determine the sentiment orientation on each aspect[2], [26] and [34]which allow for the discovery of sentiment orientation (positive, negative and neutral) on the aspect/feature. According Suganya et al.[74], this phase depends on a sentiment word dictionary that contains a list of positive and negative words (called opinion words) which are used to match terms in the opinionated text. In addition special linguistic rules proposed, because of special words change the orientation like consider negation words “not” or “no” with handling care, since not all occurrences of such rules or words apparitions will always have the same meaning. Suganya and colleagues indicate that opinion words are basis to determine the sentence orientation, so they try resolve by developing set of rules. They put algorithm by applying set of linguist rules to determine the orientation of each word in a sentence. Edison Marrese-taylor et al.[18] and [26] also considering opinion words as a basis, taking into account the work of Ding et al. (2008) as inspiration, developed a set of rules to determine the sentence orientation. In the work of [18] and [26] they settle on aspect orientation rule and word orientation rules. In word rules Positive opinion words, negative opinion words, and neutral words will basically have a score of 1, denoting a normalized positive orientation, -1 denoting a normalized negative orientation, and 0 denoting neutral (not opinion words) respectively. Negations can affect the opinion or neutral words. Then special action needed and Edison Marrese-taylor et al.[18] and [26] apply three rules: Negation Negative → Positive,

Negation Positive → Negative and Negation Neutral → Negative. Following the above rules now determine the final orientation of a sentence on aspect. Edison Marrese-taylor et al.[18] Determiner the aspect orientation rules by aggregation function. Therefore if scores for each opinion word and neutral word in sentences are known, score for each aspect word (aw_{ij}) in sentence s can be find by the following equation as Edison Marrese-taylor and his colleagues set.

$$score(AW_{ij}, S) = \sum_{ow_j \in S} \frac{score(ow_j)}{WD(ow_j, AW_{ij})} \dots \dots \dots \text{equation 2.1}$$

Where ow_j is an opinion word or neutral word in sentence s ; $WD(ow_j, aw_{ij})$; is the word distance between the aspect word aw_{ij} and the opinion word ow_j in s . But this way is not handle compound aspect word. Therefore they also Edison Marrese-taylor et al.[18], set other aspect aggregation rule taking into consideration the score so fall the words that compose it. See the following.

$$Score(ai, S) = \sum_{aw_{ij} \in AW_i} Score(aw_{ij}, S) \dots \dots \dots \text{equation 2.2}$$

On the other hand Kansal and to shniwal predict the sentiment orientation by using opinion lexicon[35], an online dictionary, containing the collection of positive and negative opinion words. In this research [35] first they identify subjective words that are mostly expressed by adjectives and verbs in sentences. Due to this reason they extract all the adjectives and verbs from the tagged sentences and following form the feature opinion pair. They also map each opinion word to the nearest feature noun and form the pair. After forming pair (feature-opinion) Kansal and to shniwal find the polarity of each opinion word by lexicon. Unfortunately they face to other issue related to context dependent opinion words whose polarities depend on the context. One opinion word can have different meaning in the different domain. To predict such context dependent opinion word polarity Kansal and his colleague use linguistic rules like “and”, “but” and “however”. Using these rules they predict the polarity of opinion words before or after the conjunctions. In the case of "but" and "however", the polarity reverses after conjunction but in the case of "and", the polarity before and after the conjunction remains same. For example, "Picture Quality of this iPhone is excellent, but battery life is bad" and "Picture Quality of this iPhone is excellent and battery life is amazing" are valid sentences. They also handle phrases which are "but-like" structure and contain the "but" word, but do not change the orientation after

conjunction. As an example, "but also" in the following sentence indicate this; "I not only like camera of this iPhone but also its keypad".

Kim et al. [2] in their survey on comprehensive Review of Opinion Summarization provide different ways of sentiment prediction methods. Rule/lexicon based, Learning-based and other techniques are explained. In the context of opinion summarization lexicon based sentiment prediction is popular [25]. Lexicon based method is relies on a sentiment word dictionary. The lexicon typically contains a list of positive and negative words that are used to match words in the opinion text. To predict opinion orientation Hu and Liu [25] proposed a simple until now effective method based on WordNet. They create manually a small list of seed adjectives tagged with positive or negative labels and this seed adjective list is domain independent where the lexicon is domain dependent (e.g., movies) and must be rebuilt for each new domain as they stated. The other sentiment prediction method is learning based in the survey [2] which formulate a problem as sentiment classification and incorporate may features. The above lexicon based method can be one of the important features for learning-based predictions.

2.2.2.3. Summarization Methods

To present processed results in a simple manner, summary generation is the last step. Opinion summary generation is handle in different methods as discuss in [2] and [21]; summary with timeline, aggregate rating, text selection, and statistical summary methods are most used in various work. Statistical Summary is common and it uses a list of aspects and results of sentiment prediction. In [25] Bing Liu and Minqing Hu and in [14] Nirali Makadia et al., to generate feature based product review summary first discover feature and assign the polarity of the opinionated sentence, and secondly according to the frequency appearance in the review features are ranked. Also statistical summary shows in graphic representation format and enables to users to compare opinion of several products [2] [36] [37]. It uses to understand overall people's opinion on something. But sometimes it is needed to see (read) the actual opinion sentences to understand specifics. To showing a complete list of sentences is not very useful due to large volume of opinions on even in a single topic. Text selection solve this problem by showing smaller pieces of text as the summary using word, phrase and sentences level granularities [2] [38] [20]. By combines statistical summary and text selection together aggregate rating opinion summary technique also proposed as well by Bing Liu as stated in [2]. Opinions

come from people change from time to time within a target and help to figure out what changes people's opinions. For this purpose also use timeline methods as summary.

In [20] Warih Maharani et al. try to put two ways for summary generation in aspect based opinion summarization namely Text-based Opinion Summarization and Visual-based Opinion Summarization. Visual-based Opinion Summarization make the summaries easier and quicker to read people's opinion than text representation via rating, value, stars, bar and other visualization form [37] [20] [39]. In work [40] Opinion summary generation done with first semantic relationship mining that is selection of representative features and secondly opinion sentence extraction to extract the representative opinion sentences to generate an opinion summary. B. Cristian et al. [41] Has also different assumptions to summarize the people's opinion. They propose as best solution; find short sentences between three and eight words that can summarize the set of reviews as good as possible like in text summarization [42] but it is not common because it is clearly different from traditional text summarization by only interested in feature of the product that people give opinion on [12].

Visual representation ways of summary generation is more interesting and affects users' perceptions. Commercial websites like TripAdvisor, Rotten Tomatoes, Amazon, Yelp, IMDb, and eBay present reviews product summarization within various visualization such as thumbs up/ thumbs down, positive and negative signs, star rating and rating meter [20]. Different visualization categories are available. From those visual-based summarization can categorize as content based and quantitative. Works [36] and [20] are focus on quantitative based rating visualization. With the graph representation, we can obtain people's overall opinions about the target more automatically. This format of visualization summary has been widely adopted in the business website world [2]. Therefore we are interesting to do Aspect based Amharic opinion summarization by graph to contribute for our Ethiopic (Amharic) opinions within business environment.

2.3. Amharic language study

The official language of Ethiopia is Amharic that belongs from Geez (ልዩነ-ግእዝ as William says [82]). It is the second vastest semantic language next to Arabic in the world [75] [76] [77]. It is the official working language of the Federal Democratic Republic of Ethiopia and thus has official status nationwide. It is also the official or working language of several of the

states/regions within the federal system, including Amhara and the multi-ethnic Southern Nations, Nationalities and Peoples region. Outside Ethiopia, Amharic is the language of millions of emigrants [77]. Amharic has an estimated of over 40 million speakers [78] in (Egypt, Israel Sweden and Eritrea [77]) and increased through time due to it is the working language and Amharic is a written language with its own alphabet and written materials. So Amharic subjective documents on the web increases gradually as many newspaper agencies, review sites, forums, discussion groups, blogs and so on, provide availability of opinion rich documents electronically. For different purpose as stated in ([section 1.8](#)) summarizing these opinionated document is needed. To do this the Amharic language different knowledge's are compulsory like word structure. Let's see some of Amharic language knowledge's in the following sections.

2.3.1. Amharic Word Structure

As indicated in Muluaem work [78] Amharic shows a root-pattern morphological phenomenon similar to Arabic. It is morphological rich language like Arabic because of the reason of the nature of the language that is highly inflectional and derivational [88]. Amharic has 8 word classes [78] [79] which are interjection, verb, adverb, adjective, pronoun, preposition, noun, and conjunction in early studies like [82] [80]. But in [81] Amharic word class or POS (part of speech) is grouped in to five (Noun, Verb, Adverb, Adjective and Preposition) where pronouns and conjunctions are included under noun and preposition categories, respectively and remove interjection. In addition to the difficulty morphologically complex and difficult in phoneme's Amharic also challenging in writing where misspell words in different ways like use interchangeably the third and the sixth phonemes. Amharic writing system also (alphabet called "Fidel"), has 33*7 much number characters [105, 108]. It is also has 34*7 number of characters according to work [63]. From these characters some of the mare pronounced the same but symbolically different homophones. For instance William says [82] in the present Amharic *ሀ*, *ሐ*, and *ኸ* are pronounced as 'h' in house and are often switched in objective Amharic sentence. Tulu [88] also says no clear rule to use *ሀ* or *ሐ* for which word. So peoples use interchangeable for the same word in meaning. This is also true for subjective/opinionated Amharic sentence in the blog, forum, social network (Face book, tweeter, and etc.) and others user reviews. Therefore this situation increases the number of features going to be extracted without any benefit. Furthermore many ways of writing for semantically same word (e.g. the word head can be written as አራስ or ራስ) mainly for loan words (e.g. the word computer can be written as ኮምፒዩተር or ኮምፒውተር.)


2.3.2. Amharic opinion Lexical category

Adjectives classes of a language are identified as opinionated words by a number of researchers such as Bo Pang et al. [83]. Noun also the most important word class for aspect based opinion tasks as Tulu indicates [88]. So it is very important to deal about deferent part of speech of Amharic. In different work [90] usually nouns are aspects. Therefore to identify and extract aspect, we must analysis word POS. For example, “face recognition”, “zoom”, and “screen” are aspects of the product “camera” as Amani et al. indicates [90].

A. On the Noun:

According early work like William [82], four considerations are known in speaking of Amharic noun according to their formation (termination, species), gender, number, and declension. Noun formation can Simple, Compound, and Augmented or Primitive or Derived [82]. Based on their termination Amharic noun can end in any of the six orders except first order. Tulu [88] also try to classify nouns in two; basic nouns (which are giving meaning by their own) and derived nouns (derived from verbal roots, adjectives, stems, compound words and nouns) as he mention as the recent work. Getahu Amare [55] also indicates that this two classes of nouns as a modern Amharic language word classes and we uses this word classification for our work. William [82] also says derived noun can be drive from particles. Noun can be name of a place, person, thing, or idea.

On the gender of the noun: Amharic nouns have two genders masculine and feminine. The female is distinguished by the termination ት, ታ, and ቱ [82]. Tulu also say by affixation of the morpheme ኡ or -ኡ.ት [88].

On the number of the noun: The number is twofold either singular or plural. When Amharic singular noun establish no particular forms found. But when plural nouns form have one form, which is the termination ኦች och; in which that recognize in the Arabic  as William indicate in his Amharic Grammar [82]. Tulu [88] also indicate adding by affixation “-ኦች” can make plural noun form similar to William [82]. For example the plural form of noun word ቤት (house) is ቤቶች (houses) by adding affixation “-ኦች” at the end of word. In work [82] other form of plural is also present; termination with “an” or ኦን (e.g. መት → መታን), “at” or ኦት (e.g. ቅድስ → ቅድሳት), and irregular form (e.g. አንበሳ → አናበሳት).

On the declension of the noun: Noun declension is very simple and uniform as William expressed. Nouns are inflected through four (nominative, dative, genitive, and accusative) cases in singular and plural. William also takes as example the singular word ቤት as nominative, የቤት as genitive, ለቤት as dative, and ቤትን as accusative.

In modern Ethiopic (Amharic) language word class are Noun, Verb, Adverb, Adjective and Preposition [55]. We are more interesting on both noun and adjective class of words due to different researches are indicating those two word class are very important for the work of opinion mining as well as opinion summarization. As we discuss in the above in several work we observed that Nouns are taking አች to indicating plural form of words. But all words which take አች are not always noun. For example the word ጎበዝ, ቀጭን, and ሰነፍ are takes አች but they are differing in their position. Nouns are classifying into two major types by Getahun [55] in 2003. These are primitive/basic and derived nouns. Primitive nouns are noun which are not come from other word. Derived nouns also come or built from noun or other word class. Examples for primitive and derived nouns are ቤት and ቀለም from (“ቅጽልኝም”) respectively. Baye [62] and Getahun [55] discussed this idea in detail.

B. On the adverb:

The most essential part of speech next to noun is adverb in Amharic language. We can identify from other class of words in two ways by their position in the sentence at end of it and by structure. They found at the end of Amharic sentences. For example in the sentence “አናት አገር ኢትዮጵያ ነጻነቱዋና አንደነቱዋ እስከ መጨረሻው ተጠብቆ ይኑር”, the word “ይኑር” found at the end of sentence. Hence this word is taken as verb. In the works Getahun [55] Baye [62] and also William [82] are support and describe this core part of word class in detail. Due to the reason it has less rolls for our work opinion related, we try to focus on the next critical needed Amharic part of speech word class called adjective.

C. On the Adjectives:

Adjectives are clarifying nouns coming before to it in the Amharic sentence structure. But all word coming before noun is not necessary adjective. As Getahun [55] articulated that before adjective words in the sentence we can add the word “በጣም” (very) and then the sentence is validated. For example on the phrase “ምርጥ ሰው” we can add the word “በጣም” and forms new

phrase “በጣም ምርጥ ሰጧ” which is valid. Then the word “ምርጥ” is adjective according to Getahun says. Also words under this class grouped into primitive/basic and derived adjectives. Basic adjectives are adjectives which are not constructed from other class of words such as “ደቆ”, “የዋህ”, “ሞኝ”, and “ደግ”. They are less in their number. But derived adjectives are many in their number and they derived from verb, noun, and adding adjective former affixes at the end of word [55]. This class of word is very useful for us to extract the subjective Amharic sentences. Many researches indicate that adjective class of words of a language is an indication of opinionated sentences [88] [53].

In general Ethiopic (Amharic) language is very challenging language due to many reasons as tulu [88] says. The major one is it has very complex morphology. Amharic language is highly inflectional and derivational that makes morphological analysis very complex [88] [82]. Secondly it has many characters with having similar pronounce and some of writers wrote in different symbols. As tulu say one word may has different symbols not only browning words [88] but also one original amharic words also can represented in several character symbols. As evidence in Amharic language can be written as “አገር” and “ህገር” which have the same meaning in both way of representation which mean country.

Chapter Three

Related Work to Opinion Summarization

The growth of the Internet as a commerce medium, and particularly the Web 2.0 phenomenon of user-generated content, have resulted in the proliferation of massive numbers of product, service and merchant reviews. While this means that users have plenty of information on which to base their purchasing decisions, in practice this is often too much information for a user to absorb. To alleviate this information overload, research on systems that automatically aggregate and summarize opinions have been gaining interest [3]. With the growth of the web over the last decade, opinions can now be found almost everywhere blogs, social networking sites like Facebook and Twitter News portals, E-commerce sites, etc. While these opinions are meant to be helpful, the vast availability of such opinions becomes overwhelming to users when there is just too much to digest [2]. Over the last few years, this special task of summarizing opinions has stirred tremendous interest amongst the Natural Language Processing (NLP) and Text Mining communities. ‘Opinions’ mainly include opinionated text data such as blog/review articles, and associated numerical data like aspect rating is also included. While different groups have different notions of what an opinion summary should be, we consider any study that attempts to generate a concise and digestible summary of a large number of opinions as the study of Opinion Summarization [5], [12].

The simplest form of an opinion summary is the result of sentiment prediction (by aggregating the sentiment scores). The task of sentiment prediction or classification itself has been studied for many years. Beyond such summaries, the newer generation of opinion summaries includes structured summaries that provide a well-organized breakdown by aspects/topics, various formats of textual summaries and temporal visualization. The different formats of summaries complement one another by providing a different level of understanding. For example, sentiment prediction on reviews of a product can give a very general notion of what the users feel about the product. If the user needs more specifics, then the topic-based summaries or textual summaries may be more useful. Regardless of the summary formats, the goal of opinion summarization is to help users digest the vast availability of opinions in an easy manner. The approaches utilized to address this summarization task vary greatly and touch different areas of research including text clustering, sentiment prediction, text mining, NLP analysis, and so on [5].

Research in the area of summarizing documents focused on proposing paradigms for extracting salient sentences from text and coherently organizing them to build a summary of the entire text [5]. The relevant works in this regard includes Paice and Kupiec [12] the earlier works focused on summarizing a single document, later, researchers started to focus on summarizing multiple documents. Due to the characteristics of data itself, opinion summarization has different aspects from the classic text summarization problem. In an opinion summary, usually the polarities of input opinions are crucial. Sometimes, those opinions are provided with additional information such as rating scores. Also, the summary formats proposed by the majority of the opinion summarization literature are more structured in nature with the segmentation by topics and polarities. However, text summarization techniques [43], [5] still can be useful in opinion summarization when text selection and generation step. After separating input data by polarities and topics, classic text summarization can be used to find/generate the most representative text snippet from each category [5]. One of the work's on Aspect Based opinion Summarization is the work of Dim and Thin [89].

In most case humans are rely on the customer comments and experiences of other persons before they make a discussion and make purchasing as discussed in [65]. This is fact in different day to day human activities. Sonal [65] says the first hand experiences are generally more sought after rather than the seller's description of the product. Therefore if a system could be developed that gives the detailed summary of the various opinions existing online then this can greatly help the potential buyer and companies selling their products can also use this to track how well their products are being received by the users. For the benefit and efficiency use of this summarization it is needed to develop mechanism to generate it by considering several aspects. To opinion mining and summarization related works are mostly in movie reviews and product [5, 14, 25, 66, 67, 68, and 98].

Ahmad works is much related to us [66] on the title Review Mining for Feature Based Opinion Summarization and Visualization. In this work major components of the work include subjectivity/objectivity analyzer, feature and opinion learner that includes rule based approach for feature-opinion pair extraction and anaphora resolution for feature-opinion binding, feasibility analyzer, sentiment analyzer, and the final task is feature based review summarization and visualization. The classification of objectivity/subjectivity is used to eliminate unwanted and

unnecessary objective sentences from further preprocessing and each sentence are tokenized into unigrams. Various classifiers are practiced for identify efficiency of feature determination of subjectivity like naïve Bayes, J48, multilayer perceptron. The next architectural component of this Ahmad [66] work is feature and opinion learner. This model is responsible for comprising feature, modifier, and opinion to extract from subjective review. Mining processing is initiated with statically parser and facilitated by rule base system to identify candidates. Feature-opinion pairs are tracking with technique inferring anaphora pronoun. The third component is feasibility analyzer to eliminate noisy feature-opinion pairs extracted before with different algorithm. Other critical task in Ahmad [66] research is sentiment analyzer that classifies sentiment or polarity (positive, negative, or neutral) of opinion bearing words present as a part of information components. To determine the sentiments of opinionated words supervised approach based on statistical and linguistic features for word-level sentiment classification is applied. Point wise Mutual Information, Mutual Information, Log Likelihood Ratio and some linguistic features (including negation, tf-idf and modifier for classification purpose) are considered in this task. The sentiment analyzer implemented in two phase training and testing. Decision Tree and Bagging Algorithms are implemented in WEKA due to having best performance after considering four (Naive Bays, Decision Tree (J48), Multilayer Perceptron (MLP), and Bagging) classifier considered as above. Finally Ahmad employed Opinion Summarization and Visualization System with providing graphical representation by using bar and pie charts for every product feature using JSON. The system Opinion Summarization Visualization Sentiment interface screen shows two main rows, upper (having three panels' upper-left, upper-middle, and upper-right) and lower (having two panels lower-left and lower right). Finally in this work for a selected review document, the lower-right panel presents the list of extracted results that includes feature, modifier (if any), opinion, orientation (positive, negative, and neutral) and opinion indicator of the feature-opinion pair.

Another related work is done by Chinsha TC and S. Joseph [72] with having seven main tasks in their proposed model. In this research first they collect restaurant reviews from the web using crawler that extracting review from web pages. Then after collecting restaurant reviews, preprocessing of reviews is continue, to improve the accuracy of opinion mining process and to avoid the unnecessary processing overhead. Thirdly extracting aspect is going on. Aspect may be a single word or phrase. Since aspects are noun and noun phrase in this work they use Stanford

POS tagger to identify word classes. All sentence come from webpage as reviews are not express opinion. So it is needed to remove objective sentences for further preprocessing. This classification is achieved by sentiwordnet (which contains opinion words) and aspect dictionary (contains important aspects created earlier). After classification they try to identifying aspect related opinion words. To do this they [72] use POS information of a word like adjective, adverb, noun and verb are used for extracting the opinion words in a sentence after checking any aspect is present in a sentence. Next to extracting opinion words then the orientation of opinion i.e., polarity scores of opinion on aspect using SentiWordNet is done. Here in [72] does not consider the word sense disambiguate. The final seventh work is aspect based summary. Positive and negative scores of aspects are separately aggregated; hence they get an aggregate positive and negative score of aspect to generate restaurant review summary using visualizing tool. The scores of opinions on each aspect in all reviews can be aggregated using the formula $Aggregate_Positive_polarity_{[j]} = \sum_i Positive_Pol_{i,j}$ and $Aggregate_Negative_polarity_{[j]} = \sum_i Negative_Pol_{i,j}$ for each aspect j of the restaurant. They also evaluate their system at the end. The evaluation is done by measuring accuracy, precision and recall.

Minqing Hu and Bing Liu also work ‘Mining and Summarizing Customer Reviews’ [73] which is very connected to our work. The feature based opinion summarization system of M. Hu and B. Liu performs three main tasks (1) mining product features that have been commented on by customers; (2) identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; and (3) summarizing the results having many sub-steps. The first sub-step is POS tagging which helps us to find opinion features. Features are usually noun and noun phrase [25, 66, 71, and 72] from review sentences. Therefore POS is necessary and M. Hu & B. Liu used the NLProcessor linguistic parser to parse each review to split text into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc.). The process also identifies simple noun and verb groups. NLProcessor generates XML output. Also some pre-processing of words is also performed, such as removal of stop words, stemming and fuzzy matching that is used to deal with word variants and misspellings. In the next step M. Hu and B. Liu do identify frequent features that number of people expresses their opinions. To do this they use association mining Apriori algorithm to find all frequent itemsets. An itemset is simply a set of words or a phrase that occurs together in some sentences. But all candidate frequent features generated by association mining are not genuine features.

Some unlikely features should be removed by using Compactness pruning and Redundancy pruning.

The third sub step in M. Hu and B. Liu work is opinion word extraction. In this work adjectives are taken as opinion words. After opinion word is extracted the orientation of opinion word is identified which predict semantic orientation of each opinion sentence which indicates the direction that the word deviates from the norm for its semantic group. Unluckily, dictionaries and similar sources, i.e., WordNet do not include semantic orientation information for each word. To predict the semantic orientations of simple and yet effective method by utilizing the adjective synonym set and antonym set in WordNet. Due to this absence of semantic orientation, B. Liu and M. Hu uses new strategy a set of seed adjectives, which they know their orientations and then grow this set by searching in the WordNet. Infrequent feature identification is the next sub step employed by B. Liu and his colleague. Infrequent feature identification using opinion words is that it could also find nouns/noun phrases that are irrelevant to the given product. Then the sixth sub task is predicting the orientation of opinion words. In the case where there is the same number of positive and negative opinion words in the sentence, B.Liu and M. Hu predict the orientation using the average orientation of effective opinions or the orientation of the previous opinion sentence. The final task is straightforward summary generation in two steps. The first step is discovered feature, related opinion sentences are put into positive and negative categories according to the opinion sentences' orientations. The second step is all features are ranked according to the frequency of their appearances in the reviews. Feature phrases appear before single word features as phrases normally are more interesting to users. Other types of rankings are also possible.

M. Hu and B. Liu [73] also evaluate their aspect based short summary system based on three perspectives effectiveness of feature extraction, opinion sentence extraction, and accuracy of orientation prediction of opinion sentences. They do experiments using the customer reviews of five electronics products: 2 digital cameras, 1 DVD player, 1 mp3 player, and 1 cellular phone. The reviews were collected from Amazon.com and C|net.com. The average recall and precision of frequent features using association mining are 0.68 and 0.56 respectively for 69 features identified manually. For infrequent feature identification they achieved 0.80 Recall 0.72 Precision.

Chapter Four

Research Methods and Design

This chapter gives a detailed description of system architecture and functional details of the proposed system to identify aspect-opinion pairs for Amharic Aspect Opinion Summarizing. The main goal of our Summarization task is to generate an aspect/feature based summarization on a hotel domain customer reviews. The proposed system consists of two different major functional components namely bootstrap and Naïve Bayes. These major functional components are Amharic opinions/reviews, preprocessing phase, aspect identification and opinion learner, Opinion word seed lexicon, Aspect-Opinion pair prediction and finally Amharic Aspect opinion summarization by graph visualization as illustrated bellow in [figure 1](#).

4.1 The proposed Architecture

As shown in the following figure 4.1, the general architecture of the proposed system (Amharic Aspect Opinion Summarization) is established with different components depend on the required processes. The General architecture of Amharic Aspect Opinion Summarization Using Bootstrap and naïve Bayes on Hotel Domain system consists of different things. Let's discuss each components of the proposed system (architecture) of aspect based Amharic opinion summarization.

4.1.1 Amharic Opinions (Reviews)

Amharic Opinion (Review) component prepares the raw text for the next step that recognizes and extracts opinions. This step collects opinions (sentiments) or customer reviews written in Amharic language which had been given by a number of hotel users on different hotel aspects in different hotels. Our primary assumption was to collect opinionated Amharic reviews posted by different opinion holders, on different social networks such as Face book, blogs, tweeter, and on different websites about our country hotels. But finally there is no more Amharic language opinionated text on such media. So due to the less availability of Amharic opinion text on social media, we should collect manually from different opinion holders. This Amharic opinions or comments are following the Amharic language normal sentence structure. It does not contain Amharic poem (ግጥም), proverb (ምሳሌያዊ አነጋገር), folklore (ተረትና ምሳሌ), poetry (ቅኔያዊ አነጋገር) objective sentences and other Amharic language category.

4.1.2 Preprocessing Phase

The preprocessing phase is the second step in the aspect based Amharic opinion summarization. In the preprocessing, only subjective (opinionated) sentences are submitted to a pipeline as input and accept it as responsibility. For our work, the pre-processing components needed are decided. The decided components are sentence tokenization and normalization. Here described below.

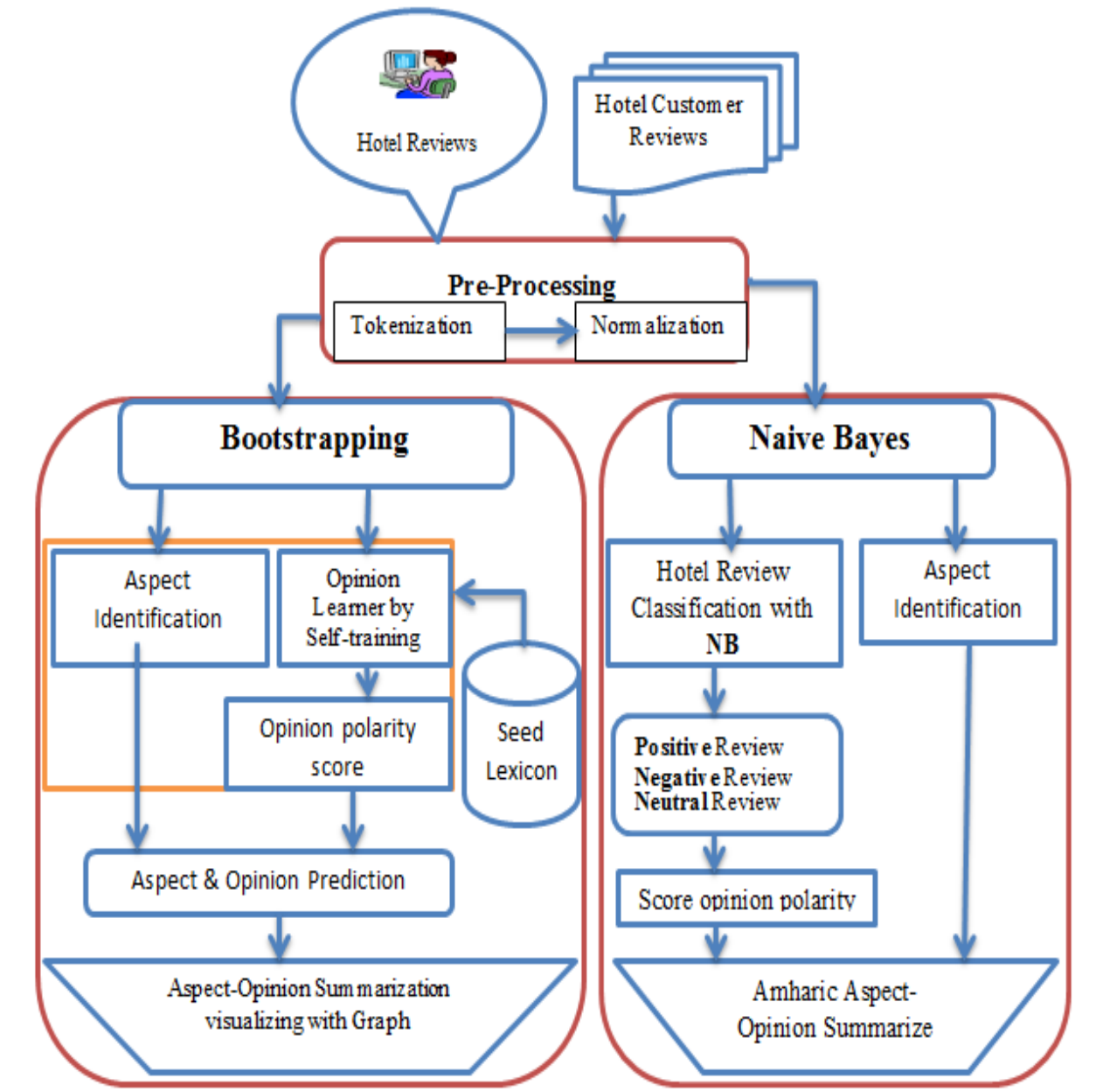


Figure 1: General Architecture of Amharic Aspect Opinion Summarization

Tokenization

The input reviews or opinions are tokenizing after opinionated sentences are splitting. Inputs for the tokenization preprocessing are opinions/reviews sentences that are splitting for processing to the next activity. This tokenization action reads a sequence of characters as a string and tokenizes them using predefined list of delimiters such as special characters like (#, @, %, ^, &, *), new lines, space and punctuation marks. This tokenization is used to break down a stream of text into phrases, words, and symbols which will be used as meaningful elements called tokens.

In this tokenization preprocessing sub component we are not concerning on Amharic punctuation marks called netela serez (ነጠላ ሰረዝ (፣)) which is used as comma even if it play a vital role for extraction of text segment. Because nowadays many Amharic modern writers and posters do not use such punctuation mark and we do not get it more on the posted and written given Amharic opinions or comments. In contrast Amharic full stop :: (አራት ነጥብ) used as independent and identifying the sentence separation. Mostly if the Amharic opinionated sentences contain more than one opinion then it is separated with full stop :: (አራት ነጥብ) in most Amharic opinion or sentiment rich text documents.

Normalization

The sub component of normalization Amharic opinionated text document will be normalized to show similar standard. In fact in Amharic language alphabet there are redundant symbols to represents a single word with having different latter orthography. In Amharic language it is common that various Amharic characters with the same pronunciation but different symbols. For instance, it is common that the character ‘አ’ and ‘ዐ’ are used interchangeably as ‘አዲስ’ and ‘ዐዲስ’ to mean ‘new’ and also the word ጸሀይ, ፀሀይ, ጸሃይ, ፀሃይ to mean sun. Such types of inconsistencies in writing Amharic words are handled by replacing character variants by a common symbol. Due to this reason such alphabets (“ፊደል”) in Amharic ሀ, ኸ, and ሐ and ኀ should be replaced with common ሀ alphabets (“ፊደል”). Normalization also handles short form of Amharic language abbreviation words which are written using period (’.’) or forward slash (’/’). For example the word ወይዘሮ is written as “ወ.ሮ” or “ወ/ሮ”, ደብረ ብርሃን written as “ደ.ብርሃን” or “ደ/ብርሃን”, ዶክተር written as “ዶ.ር” or “ዶ/ር”, የኢ.የጳ.ያ ፌደራላዊ ዲሞክራሲያዊ ሪፑብሊክ written as “ኢ.ፌ.ዲ.ሪ” or “ኢ/ፌ/ዲ/ሪ” and so on. In the following table we indicate that the Amharic character normalization as it is listed.

Group	Equivalent alphabets with different orthography	Possible normalized alphabet
a.	1. 'ሀ', 'ሁ', 'ሂ', 'ሃ', 'ሄ', 'ሀ', 'ሀ' 2. 'ሐ', 'ሐ', 'ሐ', 'ሐ', 'ሐ', 'ሐ', 'ሐ' 3. 'ኀ', 'ኀ', 'ኀ', 'ኀ', 'ኀ', 'ኀ', 'ኀ'	'ሀ', 'ሁ', 'ሂ', 'ሃ', 'ሄ', 'ሀ', 'ሀ'
b.	1. 'አ', 'አ', 'አ', 'አ', 'አ', 'አ', 'አ' 2. 'ዐ', 'ዐ', 'ዐ', 'ዐ', 'ዐ', 'ዐ', 'ዐ'	'አ', 'አ', 'አ', 'አ', 'አ', 'አ', 'አ'
c.	1. 'ሰ', 'ሰ', 'ሰ', 'ሰ', 'ሰ', 'ሰ', 'ሰ' 2. 'ሠ', 'ሠ', 'ሠ', 'ሠ', 'ሠ', 'ሠ', 'ሠ'	'ሰ', 'ሰ', 'ሰ', 'ሰ', 'ሰ', 'ሰ', 'ሰ'
d.	1. 'ጸ', 'ጸ', 'ጸ', 'ጸ', 'ጸ', 'ጸ', 'ጸ' 2. 'ፀ', 'ፀ', 'ፀ', 'ፀ', 'ፀ', 'ፀ', 'ፀ'	'ጸ', 'ጸ', 'ጸ', 'ጸ', 'ጸ', 'ጸ', 'ጸ'

Table 1: Normalization of Amharic

An alphabet listed in above table 4.1 which are classifies in similar group has similar sound in our Amharic opinionated text or corpus. If the opinion/sentiment/comment/reviews contains such those characters then our system “Amharic aspect opinion summarization using bootstrap on hotel domain” replace into one normalized character as shown in above. Aspect based hotel review summarization with semi supervised approach is follow the bootstrap mechanism. The bootstrap uses some seed opinions words as lexicon. The possible process that will happen in this approach is discussed in the following section.

4.1.3 Aspect Identification and Opinion Learner

In fact these two information extraction (IE) tasks are more challenged in aspect based Amharic opinion summarization system. After preprocess the Amharic opinionated text document aspect and opinion learner component of our system is applied. To learn the opinion and the aspect pair, the previously preprocessed Amharic opinionated text document is analyzed and generates all possible information components from them. After detecting aspect and opinion the prediction for aspect-opinion pair is handling as illustrated in the architecture [figure 1](#). We will employ the Bootstrapping unsupervised method for opinion detection by using Seed word lexicon.

4.1.3.1 Identifying Aspects

One of the critical things in our work is extracting aspects or features or same times called review to summarize the user generated data. Most aspect indicating words are nouns and noun phrases as presenting in different work that discussed in chapter two above in section 2.2.2. 1. Therefore, to recognize all of the aspects in simple and complex Amharic opinionated sentence, defining the pattern is the effective way. As a result, the system can extract the aspects almost entirely even though Amharic opinionated sentences are not written in grammatical structure simply via observing the word matrix. See the linguistic filtering patterns how to identify a noun and noun phrase in implementation part ([section 4.2](#)).

Amharic is under- resourced language. In Ethiopian Amharic language there is no freely availability of ‘part of speech tagging’ tool which are necessary to detect word class of the language. Therefore there is no way to identify and extract the aspect word class ‘noun’ and ‘noun phrase’ in easy way by using unsupervised method even if unsupervised method is employed in any domain and useful for all domain area. To extract aspects in our hotel domain in case of our system we use seed opinion word pair with which aspect. In our opinionated Amharic dataset we are classified with nine specific aspects. When we collect Amharic opinionated text from hotel customers, we cluster opinions with most known hotel aspects manually. The aspect is the title of opinion. Nine aspects cluster as a title in our hotel domain within the collected Amharic opinion text. The categorized aspects are parking (መኪና ማቆሚያ), food/drink (ምግብ/መጠጥ), cleaner (ፅዳት ሠራተኛ), waiters (አስተናጋጅ), swimming pool (መዋኛ ገንዳ), hotel (ሆቴል), security guard (ጥበቃ), service (አገልግሎት), and bedroom (የመኝታ ክፍል). Such Amharic words (like ሆቴል, ጥበቃ, አገልግሎት, አስተናጋጅ, ፅዳት ሠራተኛ, መዋኛ ገንዳ, ምግብ/መጠጥ, መኪና ማቆሚያ, and የመኝታ ክፍል) are noun and noun phrases.

4.1.3.2. Learn opinions By Self-training

In the extracting/learn opinions sub component extracts by learning from the hotel aspect related seed opinion words lexicon. Opinion words are mostly expressed by adjectives and verbs in the sentences from the linguistic structure to identify the opinion word. In different previous work trust on a manually constructed lexicon of terms which are strongly positive or negative regardless of context. Sometime researchers’ neutral polarity also included. In this work we ignored the polarity ‘neutral’ because of our system requires the summary indication of one of the two polarities (positive either negative) of customer reviews which are given on hotel domain to build seed Opwords lexicon. And also a number of researches are focusing on domain specific opinion mining and summarizing. Because same of customer review opinions polarity are different in different domains. Like ways in our work Amharic language opinion has opinionated words that have different polarity in different domain. See the following example.

Examples of opinions word having different polarity in different domain

A. ባጠቃላይ ውድ ምግብ አላቸዉ::

As we can read from the above hotel user comment, here the word ‘ውድ’ (means ‘expensive’ in English) is negative polarity in hotel domain specifically for food aspect.

B. አሚና ለቤተሰቦቻቸው ወደ ልጅ ናት፡፡

Again when we see example ‘B’ in the given opinionated Amharic sentence the word ‘ወደ’ is also indicating positive polarity of such aspects lets we say ‘family’ aspect. Therefore we can say that opinion words are domain dependent as we can observe in the above example. Even if the given opinion words have different polarities in the above two examples, the word class of ‘ወደ’ is similar to both opinionated sentence ‘A’ and opinionated sentence ‘B’. For both above two opinionated sentences the word ‘ወደ’ is an adjective.

Extracting or detecting opinions (also called sentiments/comments/reviews) is achieved with different techniques. Lexicon-based methods are the most common opinion extract methods in different work. Lexicon based method is popular. For employed opinion lexicon there are three approaches namely corpus based, dictionary based and manual approaches. For our work we use unsupervised Association based Bootstrapping approach. For this purpose we developed Amharic **opinion word seed lexicon** for Amharic aspect opinion summarization using bootstrap on hotel domain as bootstrap. See in [section 4.1.4](#).

These opinion words are normally for domain specific dependent for our work on hotel domain. Domain specific lexicon increases our system effectiveness in the domain hotel. The reason is that some Amharic opinion words have different polarities in different domains. For example the amharic opinion word “ቀዉጢ” in the opinionated sentence “ከትፎኦቻዉ ቀዉጢ ነዉ” , have positive polarity in the hotel domain with respect to food/drink (“ምግብ/ምጠጥ”) aspect. It means “ከትፎኦቻዉ አሪፍ ነዉ” , or “ከትፎኦቻዉ ጥሩ ነዉ” in the other Amharic language expression. Also it has negative polarity in other domain such as city (“ከተማ”). The possible example for this domain can be taking the Amharic opinion sentence “በነበረዉ የእሳተ አደጋ የጅማ ከተማ ቀዉጢ ሆነች” .In this opinionated Amharic sentence the word “ቀዉጢ” has negative polarity. It means that “በነበረዉ የእሳተ አደጋ የጅማ ከተማ ተረበሸ” or “በነበረዉ የእሳተ አደጋ የጅማ ከተማ ሰላሙ ታወከ”. So domain specific Amharic opinion lexicon is more suitable for our Amharic aspect opinion summarization Using Bootstrap system.

In many work the word class adjectives are the main indicator of opinion words in any language. In Amharic language adjective can be grouping into existing “ነባር ቅፅል” and derived “ዉልድ ቅፅል” as we can observe. The existing adjectives are limited in number. Example of these existing

adjective are such as የዋህ, ደግ, ሞኝ, በጎ, ዲዳ, ከንቱ, and so on. But in contrast derived adjectives “ወልድ ቅፅል” are very much in number. Derived adjective is constructed from root word (አምድ), noun (ስም), and root verb (ከግስ ስር). From root verb (ከግስ ስር) can be made by inserting different vowels “አናባቢ” between root verbs (ከግስ ስር). For such type of derived adjective the possible example can be the opinion word adjective “ሰነፍ” constructed from root “ሰ-ን - ፍ” by adding vowel ኧ at the beginning & pre-ending and at the ending & pre-ending of the opinion lexicon word like that “ሰኧንኧፍ”. Another type of adjective is constructed by inserting two type vowels in the root verb. For example from root verb word “አ-ጥ-ር” with the vowel of “ኧ”, and “ኢ” (አጭጭኢር) opinion word “አጭር” is formed. In Amharic language to build adjective from root verb can be inserting the Amharic vowels “ኧ” at the beginning and “ኡ” at the pre-ending of the root verb. For instance from root verb “ግ-ዝ-ፍ” the adjective word ግዝፍ can be formed by adding vowels ኧ and ኡ like ግኧዝኡፍ.

Adjective also formed from noun by adding different prefixes in Amharic language. Those known prefixes are ኧ, ኡ, ኢታ, ኧኛ, አማ, አም, and አዊ. Table 4.2 shows examples of adjectives formed by adding prefixes into noun word class in Amharic.

አምድ/ስም	ቅጥያ	ወልድ-ቅፅል	ቅፅል
ቀዳድ	ኧ	ቀዳድ -ኧ	ቀዳዳ
ንቅ	ኡ	ንቅ -ኡ	ንቁ
ሳቀ	ኢታ	ሳቀ -ኢታ	ሳቂታ
ሃይል	ኧኛ	ሃይል -ኧኛ	ሃይለኛ
ፍሬ	አማ	ፍሬ -አማ	ፍሬያማ
ወሽት	አም	ወሽት -አም	ወሽታም
ባህል	አዊ	ባህል -አዊ	ባህላዊ

Table 2: Adjectives formed by adding prefixes

4.1.3.3 Polarity Defining

Here also the main activity is that identifying the polarity of seed opinion words lexicon with bootstrap approach by classifying opinions using seed opinion lexicon. Seed Opinions word extracted from hotel customer reviews have word class adjectives as we discuss in the above. These adjectives word class of amharic language words are classified into two major polarity ‘positive’ and ‘negative’ in our system. In some research the opinion polarity also can be classify into negative, positive, and neutral. For our system that not so much useful to group opinions

into neutral. Because customers want a short summary of hotel aspects or product features in general. By this reason we prepare the seed opinion word lexicon having these two polarity ‘positive’ and ‘negative’. The seed opinion word lexicon contains 10 seed opinion word lexicons. The polarity of those 10 seed opinion word lexicon is handled by Amharic linguistic professional. Mr. Tesfaye Shitto is an Amharic language teacher in Amhara national region north shewa zone at Mekoy preparatory school. The task of identifying polarity of opinion words is working by him.

4.1.4 Opinion word Seed Lexicon

As the most important part of aspect-based opinion summary system using bootstrap in hotel domain, this work focuses on identifying hotel aspects and learning opinion words from Amharic hotel customer reviews. So as we say in above [section 4.1.3](#) with identifying opinion polarity we construct the opinion word seed lexicon containing positive and negative seed opinion word lexicon. We prepare the seed opinion word lexicon having two polarity ‘positive’ and ‘negative’. The seed opinion word lexicon contains 10 seed opinion word lexicons. See at appendix D. Therefore these manually picked 10 seed words by linguistic are used for bootstrapping. Bootstrapping is a general agreement for improving a learner using unlabeled data. Typically, bootstrapping is an iterative process where labels for the unlabeled data are estimated at each round in the process, and the labels are then incorporated as training data into the learner.

4.1.5 Prediction the Aspect and Opinion Pair

After extracting the aspects/features and extracting opinion word from lexicon of Amharic opinion words, the relation of aspect/feature and opinion word pairing task is important to emphasize each of them. The pair of aspect and opinion words is realized by the co-occurrence of each word. In this research the hotel domain aspect and opinions given towards each aspect could make a pair to give the summary information useful for hotel users or any customers which want to get basic information about hotel. First the Amharic opinionated customer reviews sentences are split line by line. Then the aspect or category of the hotel will be found by title (‘ገፅታ’) of the opinion. After the aspects are categorized, the Amharic opinionated words are extracted from Amharic opinions lexicon that we are constructed. In other words we can find the polarity of the actual given opinion words from built opinion lexicon of Amharic as we indicate

the above [section 4.1.4](#). Finally count the number of categories and opinions found in the given customer reviews.

Therefore it differ from lexicon based approach (dictionary based and corpus based approaches) which use (a dictionary that contains synonyms and antonyms of a word for dictionary based) and (large human annotated corpus for corpus based). Instead we can use bootstrapping or semi-supervised learning, which needs only a very small hand-labeled training set, “*10 seed words*”.

4.1.6 Aspect Based Amharic Opinion Summarization by Graph

The summarization task is different from traditional text summarization because we only mine the features of the product (hotel aspects) on which the customers have expressed their opinions. We do not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization. After all the previous steps, we are ready to generate the final feature-based opinion summary by chart. Complete summary generation consists of the following steps: For each subjective sentence, feature and related opinion phrase are put into positive and negative categories according to the opinion sentences’ orientations. Our system aims to identify aspects of a given entity (hotel aspects), and summarize the overall Amharic opinion orientation or polarity (from our seed opinion word lexicon) towards each hotel aspect. This kind of summarization is useful for hotel users or consumers as well as for hotels owner or business class. However, it may lose some detailed information, which is also important for consumers to make decisions. For example, hotel users may wish to get more information on suggested bedroom in detail instead of only summarizing which hotel bedroom classes are good or bad.

After extracted aspects and opinion polarity from seed opinion word lexicon, then count the total polarity as positive and negative of opinions and aspects then summarize the polarity of opinions on a given hotel aspects. The aggregated positive score and negative score of opinions (user-generated content or collected from hotel user Amharic opinion text documents) has been used for generating opinion summary using visualization tool graph, which is easy to understand users opinion in structured form. In this Amharic Aspect Opinion Summarization Using Bootstrap system, we use bar chart for generating summarization of hotel users’ opinions about different aspects of five hotels.

Appendix C describes the sample source code of Amharic Aspect Opinion Summarization Using Bootstrap on Hotel Domain.

4.2 Algorithm

4.2.1. Bootstrapping

A. Aspect Identification

Machine learning algorithms ‘learn’ to predict outputs based on previous examples of relationships between input data and outputs (called training data). Also machine learning at a high level, that of letting large amounts of data dictate algorithms and solutions for complex inferential problems. Supervised machine learning requires a training data set with labeled data, or data with a known output value. For this work we were categorized by manual. So it is vital to categorize the hotel aspects for our domain to do our system aspect based Amharic opinion summarization by graph on the hotel domain. After preprocessing our Amharic opinionated text document then assign category by titles. Then titles are taken as aspects.

Aspect Identification source code

```
1      if (!(txtDisplay.getText().equals("")) {
2
3          String[] titles = txtDisplay.getText().split("ገጽታ-: ");
4
5          category =new String[titles.length];
6          pos =new int [titles.length];
7          neg =new int [titles.length];
8          for(int t=1;t<titles.length;t++){ //titles
9              negative=0;
10             positive=0;
11             String[] opinion = titles[t].split("\n");
12
13             title = opinion[0];//subtitle
14             category[t]=opinion[0];
```

B. Bootstrapping with seed

It is obvious that statistical approaches to information extraction and natural language processing tasks require vast amounts of information in order to perform acceptably and produce reliable results. Therefore training experimental algorithms requires huge corpora. A corpus of unannotated, unmarked natural language text is not often suitable for training purposes on most tasks; algorithms typically need their training data to be annotated in some fashion in order to be able to extract and learn the relevant textual features. Unfortunately, annotated corpora are in

short amount. Manually annotating even a small corpus is extremely time-consuming and is infeasible for larger corpora. Hence Bootstrapping provides an alternative to painstaking manual annotation. The techniques reviewed herein are all trained on unannotated corpora. The notion is to start with a small number of carefully chosen **seeds** positive (e.g., excellent) and negative words (e.g., bad), and then to use these words to induce sentiment polarity orientation for new words in a large unannotated set of texts (in our case, hotel reviews). The goal of the bootstrapper is to perform semi-supervised information extraction.

The opinion word learner component learns the opinions from constructed Opwords Seed lexicon as bootstrapping. Given the hotel reviews, the task is to find the hotel aspect and opinion words. As observed, the opinion words mostly appear around the Aspects in the review sentences. They are highly dependent on each other. So we adopt bootstrapping method using seed words. The bootstrapping consists of two steps: the Initialization and Iterate Bootstrapping. Initialization is a small number of hand-chosen seed information (in our case positive and negative keywords) for each class are applied to the entire set of unlabeled examples, resulting in labels for those examples matching the keywords. Iterate Bootstrapping is also Improve the Model and Relabeling.

1. Start with an empty list of things
2. Initialize this list with carefully chosen seeds
3. Leverage the things in the list to find more things from a training lexicon (as corpus)
4. Score those newly found things; add the best ones to the output.
5. Repeat from step 3
6. Stop after a set number of iterations

Figure 2: Bootstrapping algorithm for Amharic Aspect Opinion Summarization on Hotel Domain

The algorithm is given a small seed set (10 selected Amharic positive and negative words) of labeled instances of each word and a much larger unlabeled corpus (hotel review). The system first trains an initial classifier on the seed set. Then it uses this classifier to label the unlabeled corpus hotel reviews. The algorithm then trains a new classifier on seed set, and iterates by applying the classifier to the now-smaller unlabeled seed set, extracting a new training set “*seed set*”, and so on. The process is repeated until some sufficiently low error-rate on the training set is reached.

Selecting Candidate Seeds

In bootstrapping classifier the first thing is to choice a set S of seeds to try. For bootstrapping to work, it is crucial that this set contain a fertile seed. Seeds are chosen either at random by picking the top *N* most frequent terms of the desired class or by asking experts. None of these methods is quite satisfactory. The impact of seed set noise is reflects on the final performance. No general agreement regarding exactly how many seeds are necessary for a given task. Some researchers say 10 to 20 seeds are a sufficient starting set in a distributional similarity model to discover as many new correct instances as may ever be found. Others also say 1 or 2 instances are sufficient to discover thousands of instance attributes. In this work 10 seeds are selected by asking language expert (5 positive words and 5 negative words). See the following table.

Positive seeds word lexicon (ጥሩ, አሪፍ, ምርጥ, ጎበዝ, ሰፊ)

No	Hotel aspects	Amharic positive opinion words
1	ምግብ መጠጥ	ውሃው ጥሩ አይደለም
2	አገልግሎት	ስጋቸው አሪፍ ነው
3	መዋኛ ገንዳ	ገንዳው በወነቱ ምርጥ ነው
4	ፅዳት ስራተኛ	ፅዳት ስራተኛ ጎበዝ ናቸው
5	መኪና ማቆሚያ	መኪና ማቆሚያ ሰፊ በመሆኑ ደስ በሎኛል

Negative seeds word lexicon (መጥፎ, ጠባብ, አያኩብሩም, ስልቹ, ወድ)

No	Hotel aspects	Amharic negative opinion words
1	ጥበቃ	መጥፎ ጥበቃ ነው ያላቸው ቢያስቡበት
2	መኝታ ክፍል	ጠባብ ነው
3	ስራተኛ	ሀኒላንድ ሆቴል የሚሰሩ ፅዳት ስራተኞች ሰዎችን አያኩብሩም
4	አስተናጋጅ	አስተናጋጆቹ በጣም ስልቹ ናቸው
5	ምግብ	የምግብ ዋጋ በጣም ወድ ነው

Table 3: Selected seeds word

C. Classifications of Amharic hotel review polarity

This opinion classification identifies Orientation of an opinion on each aspect, i.e., polarity scores of opinion on aspect. Proposed method uses hotel review corpus for assigning importance scores to opinion words, which is a seed of opinion words which contains positive and the negative words. Suppose if a word is found in the seed lexicon and if its corresponding value is positive, then this opinion term is positive. Similarly, if a term is found in the seed lexicon and if its corresponding value is negative, then this opinion term is negative. Therefore the seed lexicon contains bootstrapper negative and positive words.

D. Aspect-Based Opinion Summary for visualization

Aspect based opinion summary algorithm take the polarity of positive and negative and aspects as input and then count the total polarity as positive and negative of opinions and aspects then summarize the polarity of opinions based on aspects. The aggregated positive score and negative score of user-generated content has been used for generating summary using visualization tools, which is easy to understand user’s opinion in structured form. In this study, bar chart is drawn for generating summary of hotel users’ Amharic review.

Code for bar chart draw

```

1  public DrawBarChart(String title) {
2      super(title);
3      CategoryDataset dataset = createDataset();
4      JFreeChart chart = createChart(dataset);
5      ChartPanel chartPanel = new ChartPanel(chart);
6      chartPanel.setFillZoomRectangle(true);
7      chartPanel.setFont(new java.awt.Font("Power Geez Unicode1", 1, 14));
8      chartPanel.setMouseWheelEnabled(true);
9      chartPanel.setPreferredSize(new Dimension(580, 270));
10     setContentPane(chartPanel);
11 }
12 private static CategoryDataset createDataset() {
13     DefaultCategoryDataset dataset = new DefaultCategoryDataset();
14     String [] category=OpinoInSummarize.category;
15     int [] pos=OpinoInSummarize.pos;
16     int [] neg=OpinoInSummarize.neg;
17     String pun="";
18     for(int i=1;i<category.length;i++){
19         if(i+2==category.length){
20             pun = " & ";
21             topic=topic+category[i]+pun;
22         }
23         else if(i+2<category.length){
24             pun = ", ";
25             topic=topic+category[i]+pun;
26         }
27         eLse if(i+1==category.length){
28             topic=topic+category[i];
29         }
30     }
31     dataset.addValue(pos[i], category[i], "አዎንታዊ ወይም ፖዘቲቭ");
32     dataset.addValue(neg[i], category[i], "አሉታዊ ወይም ነገጽኛ");
33 }
34 return dataset;
35 }
36 private static JFreeChart createChart(CategoryDataset dataset) {
37     Font font= new java.awt.Font("Power Geez Unicode1", 1, 14);
38     JFreeChart chart = ChartFactory.createBarChart(
39         "የተሰጡ አስተያየቶች በጣራ ስቃሙ", null /* x-axis Label*/,
40         "ጥቅም" /* y-axis Label */, dataset);
41     chart.addSubtitle(new TextTitle(topic, font));
42     chart.setBackgroundPaint(Color.white);
43     chart.getTitle().setFont(font);
44     CategoryPlot plot = chart.getCategoryPlot();
45     LegendTitle legend = chart.getLegend();
46     legend.setItemFont(font);
47     ValueAxis axis2 = plot.getRangeAxis();
48     CategoryAxis axis = plot.getDomainAxis();

```

Figure 3: Draw Bar chart Codes for aspect based summarizing

Aspect based hotel review summarization with supervised approach is uses Naïve Bayes classification for hotel reviews and Aspects/features also selected by categorizing reviews. The possible process that will happen in this approach is discussed in the following section.

4.2.2 Naive Bayes Classifier

The Bayesian Classification represents a supervised machine learning method as well as statistical methods for classification. As a classifier it learns from training data from the conditional probability of each attribute given the class label. Using Bayes rule to compute the probability of the classes given the particular instance of the attributes, prediction of the class is done by identifying the class with the highest posterior probability. Computation is made possible by making the assumption that all attributes are conditionally independent given the value of the class. Naive Bayes as a standard classification method in machine learning stems partly because it is easy to program, its intuitive, it is fast to train and can easily deal with missing attributes. The Naive Bayes (NB) classifier is derived from Bayes rule.

$P(c/d) = \frac{P(d/c)*P(c)}{P(d)}$, that is equivalent with **Posterior** = $\frac{Likelihood*Prior}{Evidence}$, Where $P(d)$ has no play role in selecting c . So, conditional probability of a hotel reviews is given as: -

$$P(polarity/review) = \frac{P(review/polarity) P(polarity)}{P(review)}$$

As described in our architecture the supervised hotel review summarization classifies reviews with Naive Bayes classifier and selects the possible hotel aspects/features using with chi squared. The possible algorithm for our supervised learning approach is show as following.

Algorithm: Supervised Hotel Review Aspect based Summarization

Input: a document d

A fixed set of classes $C=\{Positive, Negative, Neutral\}$

Output: a predicted class $c \in C$

Steps:

1. Pre-processing:
 - a. *Hotel Reviews collect from different hotels*
 - b. *Positive, negative & neutral hotel reviews were kept in pos.txt, neg.txt & neu.txt respectively*
 - c. *Tokenize and Normalize hotel reviews*
2. Hotel Aspects/Features will be manually identified by category
3. The classifier was trained using the dataset just prepared.
4. Aspect Based Hotel review Summary with Bar Graph and Text

Figure 4: Supervised Naive Bayes based Aspect Based Summarization

Feature Selection

It is energetic to categorize the hotel aspects for hotel domain. Feature selection was identified by manually. For hotel feature selection we use category. The hotel reviews are grouped into title called (ገጽታ). Then titles are taken as aspects as stated earlier.

4.3. Review data collections

As we stated in chapter one [section 1.6](#) our domain is hotel. The main reason that we are choosing hotel domain is that, in hotels there is a chance to get a number of opinion holders. Then we can get Amharic opinions from those customers in easy way rather than other domains. We choose five hotels having different standard. The opinions are collected focusing on nine the most common hotel aspects.

The Amharic opinion reviews (also called comments or sentiments in this work as well as in other work) are collected from three geographically far apart places. We were collected from five hotels which are allocated in three regions in Ethiopia namely Oromia national region (Boni, central Jimma, and honey land hotels from Jimma), Amhara national region (Chefa valley from kemissie) and southern nation's nationalities and peoples regional state (Tepi & Friwa from tepi). Unfortunately I was move around these three regions due to my work and family settlement. So I was interested to collect opinions from those different places. The chance that I was seen those three places (Jimma, Tepi and Kemissie) is motivate to me to collect data from these different environment. We were collect more than 440 sentences from a number of hotel users in 9 aspects which are containing around 799 reviews. Here is below figure 4.3 presents Amharic collected sample of opinions given by different individuals from different hotels on a particular aspect so called food/beverage in Amharic (ምግብ/መጠጥ).

አስተያየት የተሰጠበት ገጽታ:- ምግብ/መጠጥ
 ተራ ቁጥር የተሰጠው ዐስተያየት
 1 ጨው አታብዙ ደስ ይላል።
 2 ከሌሎች ሆቴል ምግባቸው አሪፍ ነው ።
 3 ስጋቸው አሪፍ ነው።
 4 ጅም ሴንተራል ሆቴል ምግባቸው በጣም ጣፋጭ እና አቀራረቡም ማራኪ ነው ።
 5 ምግባቸው ጨው ይበዛበታል።
 6 ፒዛችሁን ውድጃላችዋለሁ።
 7 ባጠቃላይ ምግባቸው ውድ ነው ።
 8 የበግ አሩስቸው ዐሪፍ ነው ።
 9 የምግብ ሜኑ በስረዓት መዘጋጀት አለበት።
 10 የቴረና ፍሬዋ አስተናጋጆች በስረዓት አያስተናግዱም ።
 11 ሙሉ የሚፈልጉ ምግቦች አይዘጋጁም!
 12 የመጠጥ አይነቶች በአይነት አይገኙም!
 13 ሽንኩርት መቁረጫ በጣም ይደብራል ።
 14 የምግብ ሜኑ በስረዓት መዘጋጀት አለበት።
 15 በፍስክ ቀን የስጋ ዘር ብቻ ነው የሚዘጋጀው! ስለዝህ ማስተካከል አለባቸው።
 16 ስጋ መቁረጫው ግን እንጨት በመሆኑ ተቆራርጦ ይቀላቀላል።
 17 ምግባቸው ጨው ይበዛበታል።

Figure 5: Sample of Amharic opinions

Amharic opinionated text on hotel domain is collected from five hotels as we discussed above. This collected data is focused on nine hotel aspects. We are selected hotel aspects by considering the most useful and usable aspects/features in the hotel domain by our perspective in our country Ethiopia. The Selected aspects are beverage/food (ምግብ/መጠጥ), cleaner (ፅዳት ሠራተኛ), waiters (አስተናጋጅ), swimming pool (መዋኛ ገንዳ), hotel (ሆቴል), garden (ጥበቃ), parking (መኪና ማቆሚያ), service (አገልግሎት), and bedroom (የመኝታ ክፍል). The table blow describes the aspects and their proportional opinions collected from different opinion holders after discarded 100% similar comments to eliminate opinion sentences replica and trying to catch up different opinion sentences.

S. No.	Aspect	Numbers of amharic opinion collected		
		Positive	Negative	Total
1	ምግብ/መጠጥ	65	75	140
2	ፅዳት ሠራተኛ	42	80	122
3	አስተናጋጅ	63	59	122
4	መዋኛ ገንዳ	42	35	77
5	ሆቴል	57	41	98
6	ጥበቃ	45	35	80
7	መኪና ማቆሚያ	26	23	49
8	አገልግሎት	20	28	48
9	የመኝታ ክፍል	33	30	63
			Total reviews	799

Table 4: Aspects and their proportion number of collected opinions

In fact in real world situation helpful customer reviews and ratings promise many benefits in e-commerce systems such as Amazon, Yelp, flicker, and Google Play. Standing from this point of view, it is also has critical usage in different standard website of any organizations and business environment such as our domain hotels. In early traditional fashion; the customer opinions manually collected from the actual required environment in tedious and boring ways for their discussion making commitments. Therefore this type of works can solve such avoiding of additional tasks and helps for both hotel customers and business owners.

Chapter Five

Experimental Results and Evaluation

5.1 Introduction

In this chapter we present the experimental results of the developed aspect based Amharic opinion summarization system after implementation. The testing environment, evaluation, result and discussions are presented below her. In real world any system is developed to achieve certain basic functionality. These functionalities are evaluated to make sure that the systems are performing effectively [60]. Effectiveness refers to the extent a system fulfills its objective. So in our case the system ‘Aspect Based Amharic Opinion Summarization using bootstrap on hotel domain’ by Graph visualization, the exactness of extracting relevant hotel Aspects and the exactness of determining polarity (positive and negative) of opinion words are evaluated. Testing environment, evaluation metrics such as: precision, recall and F-measure, experimental results and discussions are all sub topics that will be discussed in the following sub sections.

5.2 The Research Implementation

In this section, the ‘Amharic Aspect Opinion Summarization Using Bootstrap on Hotel Domain’ for opinionated Amharic text document, the tools used for implementing the aspect based Amharic opinion summarization, the graphical user interface, the procedures to integrate the different components, the proposed algorithm, the input collected opinionated amharic text document or hotel users review and other related issues are described as shown in the next sub sections.

5.2.1 Tools

For development of our Amharic Aspect Opinion Summarization Using Bootstrap we use some tools. We used different environments and tools to achieve our objectives. These are NetBeans IDE 8.0.1 for design the graphical User Interface (GUI) in Java class, Text Editor to process our Amharic opinionated text document, and jfreechart-1.0.19 library for the purpose of drawing opinions in graph.

NetBeans is an Integrated Development Environment (IDE) used to easily develop desktop, web and mobile applications mainly in Java. It also provides tools for PHP, C/C++ and HTML5 languages. It is an open source cross-platform IDE. NetBeans IDE can run on any operating

system platforms that support a compatible Java Virtual Machine (JVM) since it is written in Java. NetBeans IDE 8.0.1 is the version and is used to implement this project. It also highlights source code semantically and syntactically. In the next sections, the details starting from the user interface for the data entry to the implementation of the aspect/Feature based Amharic Opinion summarization is presented.

5.2.2 Graphical User Interface

Our Graphical user interface is needed for help users to input Amharic opinionated reviews through it in the Amharic Aspect Opinion Summarization Using Bootstrap system. Users can post or view Amharic opinions through our User Interface that are collected from different opinion holders regarding about different hotels on different aspects (from five hotels namely Tepi & Friwa from tepi, Chefa valley from kemissie, Boni, central Jima and honey land hotels from Jimma in our case). We have developed graphical user interface for input the collected opinions or those posted Amharic opinions to our system. Graphical user interface is developed using java environment in order to help users send hotel user Amharic reviews via it.

Our GUI contains four parts with developed by JPanel container. The first one is the upper left part of Interface that displays customer Amharic opinionated text reviews (አስታየት) with respecting aspects that the opinion is given on it. The second is the upper right part which presents browse the opinions (አስተያየቱን ይክፈቱ) from user machines or computer. Third part of our GIU panel3 that is free space before running our system called Amharic aspect based opinion summarization. After executed our system, this part of Interface is became the place that the result is displayed. Then the final fourth of our JPanel is operations part. In this part we are going to do our system main actions with three Buttons namely open file (ፋይሉን ይክፈቱ), summarize in graph (በግራፍ አሳይ), and Clear the browsed opinion (ያጥፉ). The snapshot of our system user interface is indicated below before executed the main action of our system. See the following figure 6.

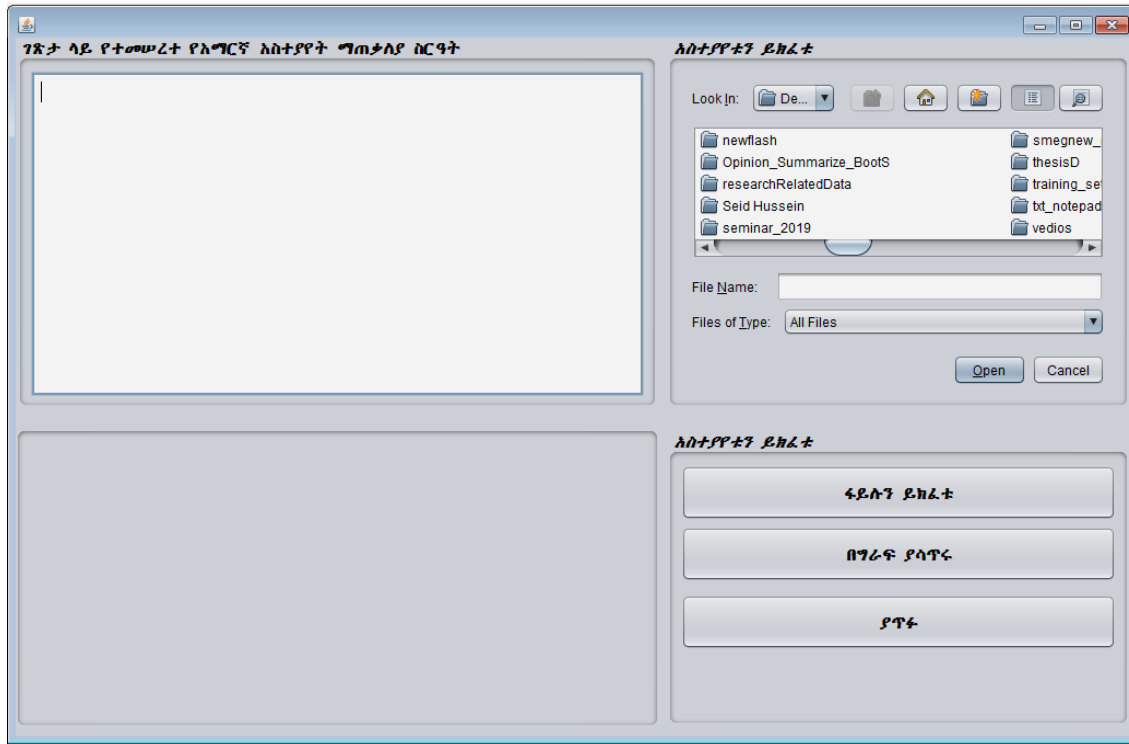


Figure 6: Aspect Based Amharic Opinion Summarization GUI before giving input opinions

After browsing the Amharic opinionated text reviews or after writing opinion sentences with congruent to particular aspect into our Interface part one ‘jPanel1’ from GUI part two ‘jPanel2’, we can perform action taken by our system. The main action in our Aspect Based Amharic opinion Summarization system is that to summarize the given customer opinionated sentences in graphical way to view easily and useful for higher hotel manager’s discussion making. We also can view the executed opinion texts and the executed result on GUI part one “jPanel1’ and GUI part three ‘jPanel3’ with new form as shown in the following figure 7.

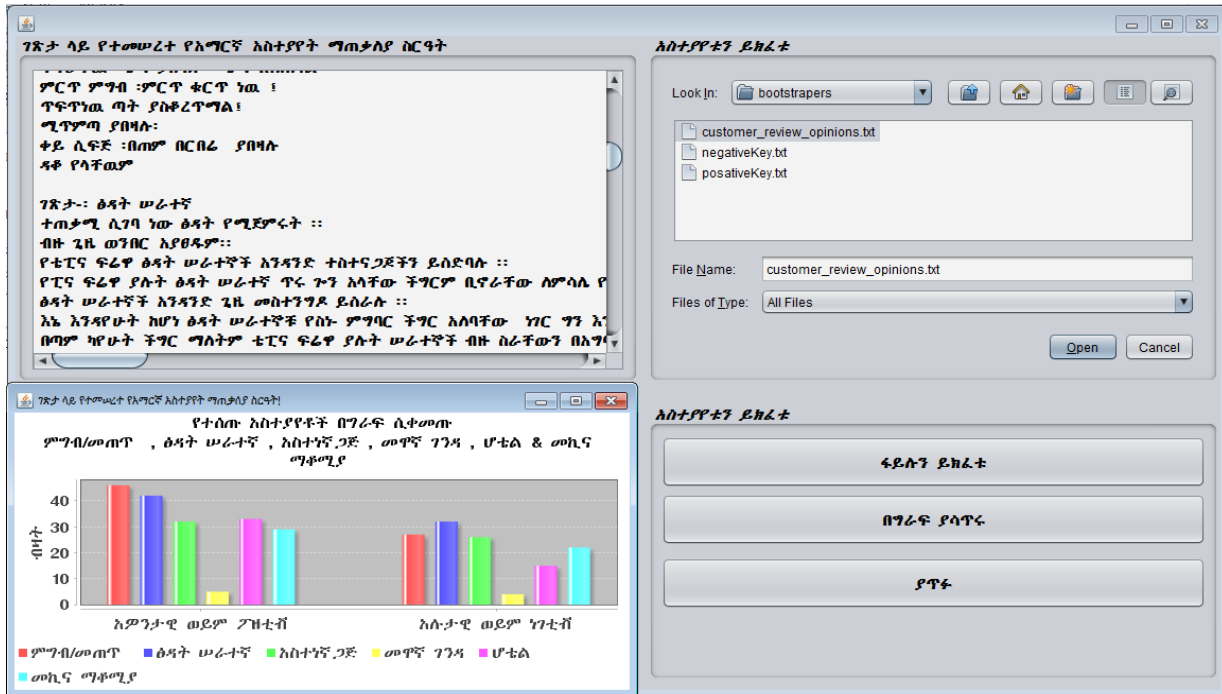


Figure 7: Aspect Based Amharic Opinion Summarization GUI after input & execute opinions

5.3. Testing Environment

For testing our aspect based Amharic opinion summarization system, we use our dell laptop with the following specifications. The testing has been done on Windows 7 ultimate operating system (OS), 2.2 GHz Intel(R) Core(TM) i5-5200U CPU, 32-bit operating system, 4GB RAM and having the storage capacity 931 GB hard disk size. On our dell laptop with the above description, we install the up-to-date JDK version ‘jdk1.8.0_45’ and NetBeans 8.0.1 was installed and prepare for usability. We also organize the Amharic opinion lexicons for the purpose of detected opinions words by classification their polarity in ‘positive’ and ‘negative’ taking as tool. The collected user generated data (in our case Amharic opinionated text document which are collected from different five hotels namely Chefa valley, tepi & Friwa, honey land, Boni, and Jimma central) has been saved with UTF-8 encoding system for processing unicode Amharic characters by text editor.

5.4. Evaluation Metrics

After we implement our proposed Aspect Based Amharic opinion Summarization system, we also must validate and verification to check whether our system is right or not and to Check whether our system is the right system or not. The system evaluation criteria mainly to check the

performance of ‘Aspect Based Amharic Opinion Summarization by Graph’ system is efficiency and effectiveness. The efficiency of our system is focus on Time and space complexity. And effectiveness of the system also for measures that the quality of classification decisions, including precision, recall, *F1*, and accuracy. Therefore to measure the effectiveness of our system, we use the most common evaluation techniques, which select a small sample of words and compare the results of the system with a human evaluator. Evaluation of our proposed system is made with the evaluation parameter that matches the number of collected Amharic opinion reviews, which are categorized in to accurately and inaccurately. We use the effectiveness metrics such as accuracy, precision P, recall R, F-measure F1 as we introduce in the above. In other word we use two evaluation strategies for our which are user-centered evaluation and system- centered evaluation to assess the usability of our proposed system. Let’s elaborate in the following sections below.

5.4.1. User-Centered Evaluation

In fact for the purpose development successful and effective of our system user centered testing is evaluation strategy that is very useful. To evaluate our aspect based Amharic opinion summarization system acceptance by users, we use the interaction questionnaires. These questionnaires help to domain experts to make personal opinions by interacting with our proposed system. So this is used to evaluate the performance of our system from the user point of view. It can also evaluate and measure the applicability and acceptability of our aspect based Amharic opinion summarization system in hotels domain with the collected Amharic opinionated text documents (from five hotel customer reviews as we discuss in data collection section). We can also consider the attitude of the hotel users about our ABAOS system.

Seven different hotel users or customers and one hotel manager or owner are conducted to the questionnaires that we prepare for the purpose of system user acceptance testing process. We were gave introduction of our system in detail how the ‘Aspect Based Amharic Opinion Summarization’ system works, for those eight selected persons to create similar awareness among them. This is help to avoid the variation of knowledge between all eight evaluators about our system. Those eight selected persons have giving their feedback based on questionnaires that we prepared for purpose of measuring user satisfaction after they (eight selected evaluators) interacts our system having similar parameters. We prepare graphical user interface as we

described in chapter four. Then questionnaires are prepared based on user interface for satisfying hotel customers or business class. The questions also measure the user interface of our ‘aspect based amharic opinion summarization by graph’ system in different characteristics. The major measure’s by questions are to be easy to use by customers, the system accuracy, the system time efficiency, attractiveness for users, adequacy, ability of problem solving, and contribution of the system ‘aspect based amharic opinion Summarization’. We prepare nine questionnaires’ for our selected evaluators. We also determine the answer words having weighed from 1 up to 5 like the work of Wegderes [61]. Answer words are outstanding, Very Good, Good, Fair, and Poor with having weigh 5, 4, 3, 2, and 1 respectively. Table 5.1 is shown the questionnaires and their answer by evaluators.

No	Questions for 8 selected evaluators	Outstanding (5)	Very good (4)	Good (3)	Fair-minded (2)	poor (1)	Average
1	How do you get the user interaction with interface and easiness use of the system?	6	2				4.75
2	Is our system aspect based amharic opinion summarization attractive or not?	6	1	1			4.625
3	How is our system efficiency in case of time?	8					5
4	Is the system giving the right opinion polarity classification based on aspects?	2	3	2	1		3.75
5	Is the system indicating right chart for aspect based amharic opinion summarization?	6	2				4.75
6	How do you percentage contribution of our system?	4	4				4.5
7	Do you get satisfaction with our new aspect based opinion summarization system?	6	2				4.75
8	Do you think the system is important for hotel domain?	8					5
9	Is the hotel aspects are much enough or less?	3	4	1			4
		Average					4.569

Table 5: Questions and Answer for User Centered Evaluation

Eight selected evaluators answered the above nine questions with weighed as shown in table 5.1. For example for question three (3), all eight evaluators agree on that the time efficiency of our aspect based Amharic opinion summarization system is 100% outstanding. Also for question four (4) one evaluator expressed as fair (12.5%), two evaluators expressed as good (25%), three evaluators expressed as the system is very good (37.5%), and the remaining two evaluators expressed as the system is as outstanding or excellent (25%) in case of give right polarity of opinion classification based on aspects. In general the average of user centered evaluation

performance for our aspect based Amharic opinion summarization system is 91.38% or 4.569 out of 5 weigh as we shown in table 5.1. This result is very interesting and great achievement for area of opinion summarization based on product aspects in general.

5.4.2. System- Centered Evaluation

System centered evaluation works based on reference judgment. Therefore as mulualem [78] says that based on the concept of relevance (i.e. given query or information need), there are several techniques of measures of IR performance available, such as, precision and recall, F-measure, E-measure, MAP (Mean average precision), R-measure and so on [60]. Relevance judgment is usually subjective (depends upon a specific user's judgment), situational (relates to user's current needs), cognitive (depends on human perception and behavior), and dynamic (changes over time). In this our work, the three widely used techniques precision, recall, and F-measure are used to measure the effectiveness of the IR system designed. We shall describe these measurements of our system effectiveness in the following section after meet the experiment within our system.

5.5. Experimental Results and Discussions

In this section, we present the results of our experiments doing for the system. We have used Precision, Recall, and F-measure metrics to evaluate the effectiveness of our system aspect based Amharic opinion summarization by graph as we wrote in the above [section 5.3.2](#). Also these three metrics recall, precision, and F-measure are the major IR effectiveness evaluation metrics [60]. The result of experiment is also describes in this section.

5.5.1 Experiments and Their Result

To achieve our system correctness, we were developed system to checks functionalities of it. For purpose of checking functionalities of our research model, we use opinion seed lexicon in judging the polarity of Amharic subjective sentences. Therefore as we say in above section 5.4, we used Precision, Recall, and F-measure metrics to evaluate the effectiveness of our system. Both positive and negative ABAOS-opinion detections are evaluated by the micro-average precision (P), recall (R), and F-measure (F), where $F = \frac{2 * P * R}{P + R}$ [52]. Therefore;

Precision (P) is the fraction of retrieved documents that are relevant. For our domain hotel reviews (HR) opinion detection uses the formulas:-

$$Precision = \frac{\#(relevant\ items\ retrived\ (classified\ correctly))}{\#(retrived\ items\ (classified\ (correctly+Incorectly))} \dots\dots\dots equation\ 5.1$$

This is indicating the Amharic opinionated document classification which is (relevant|retrieved).

$$Recall = \frac{\#(relevant\ items\ retrived\ (classified\ correctly))}{\#(relevant\ items\ (classified\ (correctly+Missed))} \dots\dots\dots equation\ 5.2$$

This formulation is also indicating the Amharic opinionated document classification that was (retrieved|relevant). Also **F-measure** is a harmonic mean evaluation measurement, which combines both recall and precision for both opinion polarity and identifying hotel Aspects.

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \dots\dots\dots Equation\ 5.5$$

5.5.2. Experiments

This experiment used our built Amharic opinionated lexicon of opinion terms. This is the one of the task of Aspect Based Amharic opinion summarization. The experiment uses more than 440 sentences which contain 799 opinions collected in the hotel domain. Sentences are settled with respective to nine hotel aspects. The following figure depicts the result of experiment.

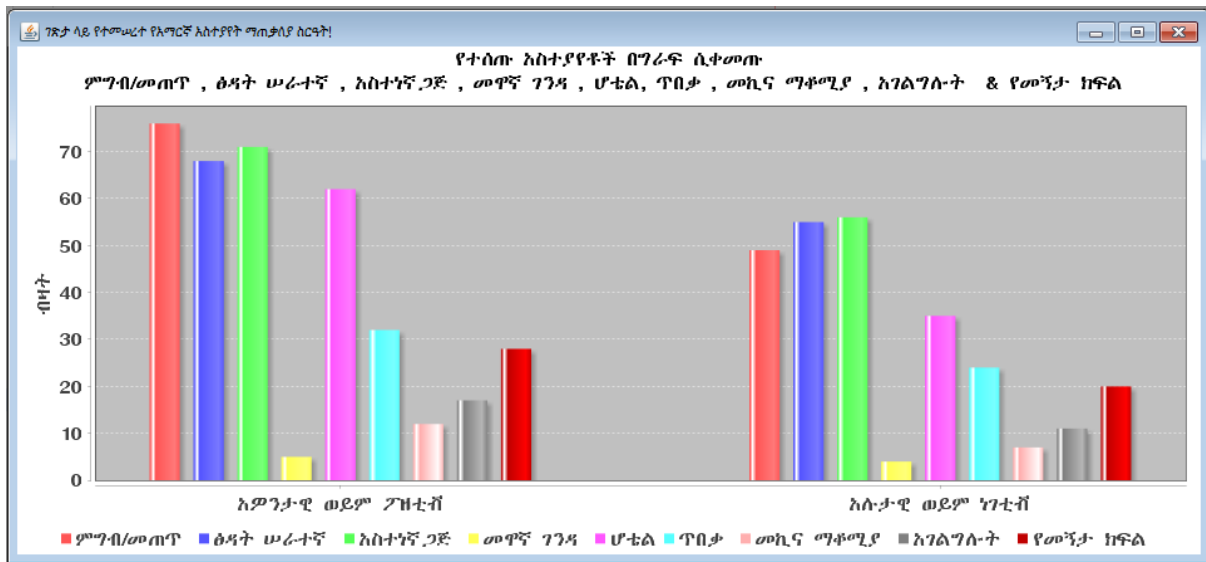


Figure 8: Summarize results by bar graph bootstrap

In the experiment we gave 799 hotel customer opinions for our developed system. After testing the system the result of experiment for opinions detection and hotel aspects identification is shown as the following table 6.

For opinion detections:					
Our system	Domain reviews	Class of opinion	Precision	Recall	F-measure
Amharic opinion classification	Hotel user reviews	Positive	0.8652	0.8167	0.8403
		Negative	1	0.6428	0.7826
For aspect identification:					
For the purpose of hotel Aspect identification we use categorized by title (ገጽታ) and then the categorized title identifies each aspects.					

Table 6: Bootstrap system centered experiment results

Generally in the following table we also examine the Confusion Matrix of System Performance Testing to shows the effectiveness of the proposed model in determining polarity of opinion words. The result of the experiment shows all the positive and negative class of opinions performance of the method.

	Relevant	Non Relevant
Retrieved	True positive (TP) #582	False positive (FP) #50
Not Retrieved	False negative (FN) #217	True negative (TN) # 0

Table 7: Confusion matrix for general system performance testing

Here the precision ($\frac{TP}{TP+FP}$) of the proposed system is 0.920 is achieved for Amharic aspect opinion summarization. Recall ($\frac{TP}{TP+FN}$) of the whole system is also 0.728 and F-measure ($\frac{2 * precision * recall}{precision+recall}$) is 0.8140. In the experiment 217 opinions (72 positive +145 negative) are not retrieved out of 799 given opinions. The system also retrieved 50 incorrect opinions especially from positive opinions.

Comparisons between manpower and automatic aspect based opinion summarization system shows saving time and manpower costs. Within 30 seconds 799 opinionated reviews are classified by the proposed system while it takes 268 minutes. We classify the polarity and aspect based opinion summarization using four person. This means that the proposed system saves more time wasted by person to clarify manually.

5.5.3 Discussions about the Results

As we seen in the above experiments; the effectiveness of our proposed system called ‘Aspect Based Amharic Opinion Summarization using bootstrap’ on hotel domain is measured by visualizing the system output. The experiments demonstrations performance the system focusing on two main perspectives. Hotel review (opinion) classification perspective and the other is hotel Aspects. The result of experiment is shown as follow by visualizing in the prototype that we were built.

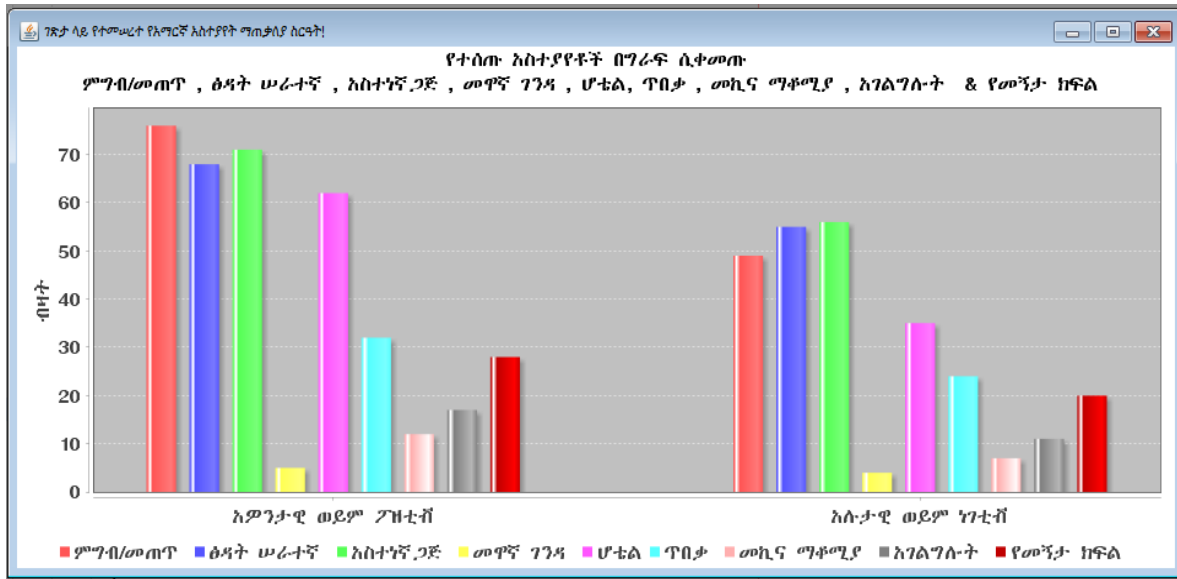


Figure 9: system centered experiment result of bootstrap approach

As shown at table 5 in the above the experiment indicates slightly different performances evaluation results for positive and negative hotel reviews. In this experiment opinion classification for *positive* polarity the performance of the system is achieved **86.52%** precision, **81.67%** recall, and **84.03%** F-measure for determination of positive opinion word. Similarly for the *negative* opinion words the system performance is achieved **100%**, **64.28%**, and **78.26%** for precision, recall, and F-measure respectively. The general effectiveness of our system for both positive and negative opinions (hotel reviews) together also recognized at confusing matrix at table 6 shown above. Here the system performance achieved **92%** precision, **72.8%** recall and **81.40%** F-measure which are the averages of positive and negative effectiveness. Here also we can find the Accuracy of our system “Amharic Aspect opinion Summarization” using confusion matrix. Accuracy ($\frac{TP+TN}{TP+FN+TN+FP}$) is 0.6855 or **68.55%** and it is very good result. Also the case

of hotel aspects/features identification is realized with categorized the title by (ገጽታ-). We did not bootstrap aspects. They identified by category. Until we differentiate with title and give the title ገጽታ-: all aspects are retrieved. The precision for both negative (100%) and positive (86.52%) polarity indicates that (relevant|retrieved) opinions. Bootstrapping method has problems on seed selection and semantic crawl. These are weakness of bootstrap.

5.5.4. Training Corpus preparation for Naïve Bayes Classification

For experimental we have used hotel reviews training dataset or corpus. This set contains 440 reviews which were tagged with polarity values positive, negative and neutral. From more than 440 collected hotel review sentences we were prepare training set. The training corpus contains 309 negative, 290 positive and 91 neutral opinion training dataset. Those hotel reviews were annotated by human and covers 80% of collected opinionated reviews. For testing purpose we also use 107 opinions which are almost 20% of collected reviews. After giving the test data to our system using the UI as shown below the result is display with bar graph and text in table.

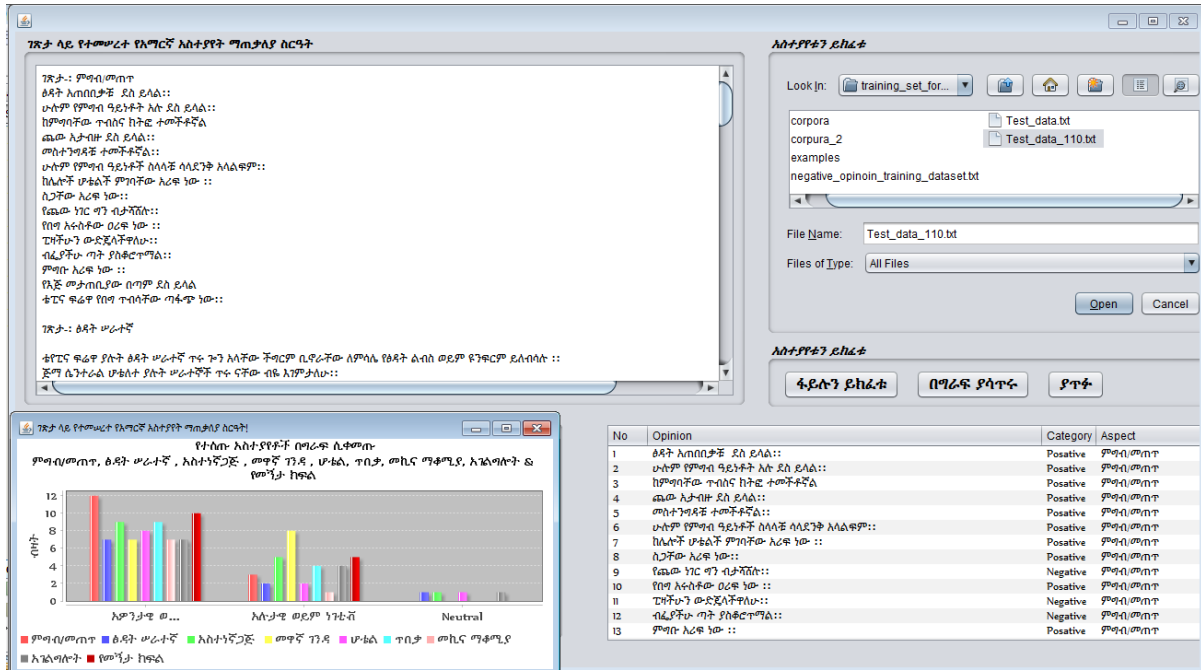


Figure 10: UI and entering reviews for Naive Bayes

Performance evaluation for Naive Bayes

In order to evaluate the effectiveness of the proposed supervised (Naïve Bayes) approach, we find the precision, recall and F-measure for experiment. The experiment result with supervised learning approach is depicted below.

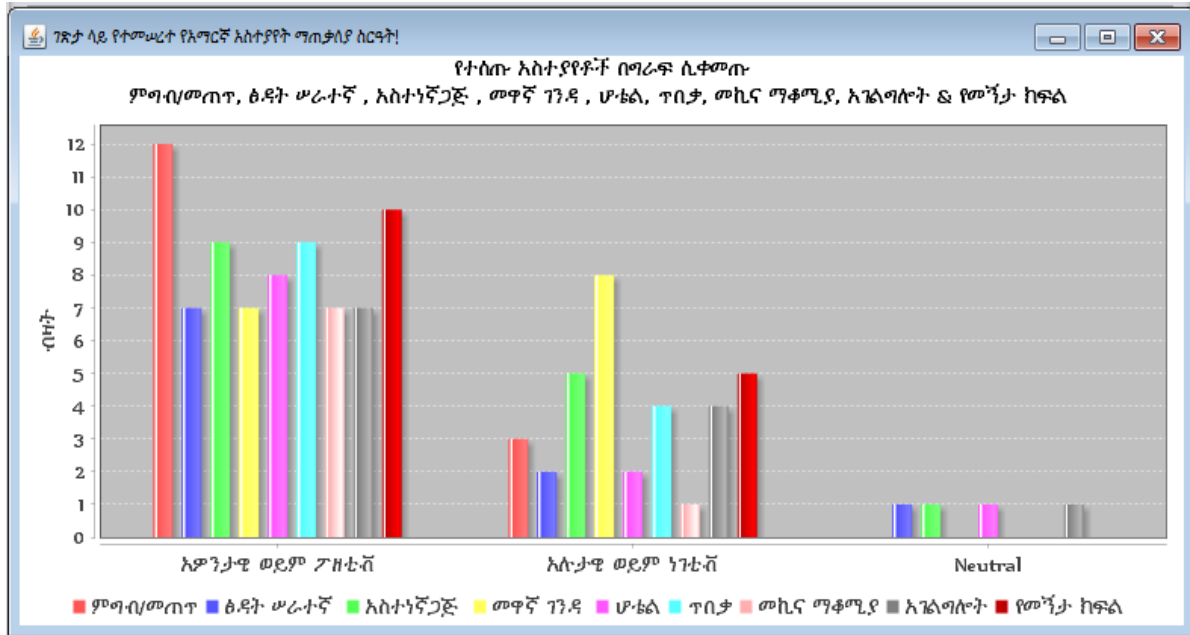


Figure 11: Graphical representations of Naive Bayes result

We use the following metrics to evaluate the classification performance: *precision*, *recall*, and *F-measure*. The precision is the percentage of documents that are correctly classified as positive out of all the documents that are classified as positive, and the recall is the percentage of documents that are correctly classified as positive out of all the documents that are actually positive. The metrics of precision and recall are defined as $(\frac{TP}{TP+FP})$ and $(\frac{TP}{TP+FN})$ respectively indicate at section 5.4.1. Among those measures that attempt to combine precision and recall as one single measure, the F1 measure is one of the most popular, which is defined by $(\frac{2 * precision * recall}{precision+recall})$.

We compute the efficiency of hotel review classification with naïve Bayes approach in the table below.

Class of opinion	Precision	Recall	F-measure
Positive	1	0.8117	0.8960
Negative	0.4705	1	0.6399
Neutral	0.8	0.8	0.8

Figure 8: Naive Bayes Classification system centered experiment results

The performance evaluation for positive hotel review is precision **100%**, recall **81.17%** and F-measure **89.60a%**. The negative opinion performance is also **47.05%**, **100%**, and **64%** for precision, recall and F-measure respectively. Neutral opinion performance achieved **80%** for all precision, recall and F-measure. Therefore generally average performance of test result for this naïve Bayes hotel review classification is that **75.68%** precision, **87.05%** recall and **77.86%** F-measure as we can observe from table above. Here we see less performance than bootstrapping approach opinion classification which score **92%** precision, **72.8%** recall and **81.40%** F-measure. This is due to the reason that naïve Bayes algorithm large amount of training data set to train. The problems in applying supervised learning are data annotation effort for training examples and difficult to scale up to a large number of application domain. Therefore, some research using unsupervised learning to overcome this problem. In the following table we try to compare the previous tulu work [53] for aspect opinion summary for Amharic opinionated text with morphological analyzer and lexicon and our work.

Done by	Method used	Perspective	Precision	Recall	F-measure	
Tulu T.	Morphological analyzer	Aspect/Feature	0.875	0.3	0.43	
		Opinion word	0.79	0.798	0.799	
Our	Bootstrap with seed words	Aspect/Feature	Identify all Aspect when title is categorized by (ገጽታ-:).			
		Opinion word	Positive	0.8652	0.8167	0.8403
			Negative	1	0.6428	0.7826
			Average	0.9326	0.7297	0.8114
	Naïve Bayes	Opinion word	Positive	1	0.8117	0.8960
			Negative	0.4705	1	0.6399
			Neutral	0.8	0.8	0.8
			Average	0.92	0.728	0.8140
Aspect/Feature		Identify all Aspect when title is categorized by (ገጽታ-:).				

Table 8: comparison of our work

Bootstrapping method achieve slightly better result than Naïve Bayes classifier when we compare these two methods as shown in the above. This is why naïve Bayes algorithm always needs huge amount of training data to increase the performance than bootstrapping. Due to time consuming, tedious task and expense language expert constraints we did not prepare large training data. So this less training data is influence our system performance. In fact bootstrapping algorithm has high precision and low recall. The weakness of bootstrapping such as semantic creep, difficulty in knowing when to stop the iterative cycles and choosing of seeds is arguably.

Chapter Six

6. Conclusion and Recommendation

Now in this section we try to set the conclusion and recommendations of the aspect based Amharic opinion summarization work. The research work presented in this work is summarized briefly as conclusion, and the future works which would enhance performance of this aspect based Amharic opinion summarization system are presented as recommendations.

6.1 Conclusion

Nowadays a huge number of peoples give their personal opinions on different issue using several site, blogs, social media, forums, and others by Amharic language over of Ethiopia and around the world. Due to the rapid growth of Amharic language based user generated content after web 2.0 created, free online opinions posted by different opinion holders on different domains. Up to now there is no systems that digesting those huge amount of customer opinions given in Ethiopic (Amharic language) for understanding opinion holder (customers) need on a given domain particular entity. Therefore in this work Ethiopic (Amharic language) customer opinions on a hotel domain was summarize with respect to their aspects /features by graph visualization. This work will help for organization (such as hotels), individual (such as hotel users), government intelligence, and business intelligence. Because of these reasons ‘Aspect Based Amharic Opinion Summarization’ system is needed.

To do this ‘Aspect Based Amharic Opinion Summarization Using Bootstrap’ system we were extract Amharic opinion words given on hotel domain and hotel aspects. After preprocessing collected Ethiopic (Amharic language) hotel customer opinions from different hotels (Tepi & Friwa, Chefa valley, central Jimma, Boni, and honey land hotels) we were use semi supervised (bootstrap approach) and supervised (Naïve Bayes classification) approaches’ for hotel reviews summary with respect to their Aspects/Features that are categorized by title called *gestita* (ገጽታ). In our research, the Amharic Opinion word extraction is realized from manually constructed seed lexicon of opinion words which are also classify in positive or negative hotel customer opinions about our domain ‘Hotel’ at bootstrapping semi supervised approach. In supervised Bayes approach we were prepared training dataset which contains positive, negative and neutral training set for classify hotel reviews. Aspects also extracted by categorizing amharic opinionated text documents within ‘titles’ and those titles are considered as hotel aspects that customer opinions or reviews was given towards. After extracting aspects and opinions towards each aspect using as input, the

next serious work is summarizing opinions towards all aspects by counting the aggregated positive and negative opinion seed word lexicons for bootstrap method and training data for supervised Naïve Bayes method within visualization. So we use Bar chart for generating summary.

After we employed our proposed system we were evaluated with performance measurements. We attempt user centered and system centered evaluation metrics for semi supervised bootstrap approach. Eight evaluators are selected to evaluate Aspect based Amharic opinion summarization based on customer satisfaction by distributed nine (9) questions for each evaluator. For example fourth question says that, “*How is our system efficiency in case of time*”. For this question all evaluators agreed on time efficiency of the system is 100% as outstanding. For question “*Is the system giving the right opinion polarity classification based on aspects*”, the result of evaluation also indicates 12.5% as fair, 25% as good, 37.5% as very good and 25% as outstanding. Finally the user centered average evaluation performance of bootstrap approach is 91.38% or 4.569 out of 5 weigh as we realized. Consequently the result is very interesting and greets achievement for such aspect based customer review summarization. We also conduct system centered evaluation metric for this bootstrap semi supervised method. For this purpose we were do experiment and measure with major information retrieval (IR) evaluation parameters precision, recall, and F-measure. Positive and negative polarity of hotel reviews score their own performance by using 799 opinions. In the experiment the system performance achieved **92%** precision, **72.8%** recall and **81.40%** F-measure which are the averages of positive and negative effectiveness. In the case of hotel aspects/features identification we use the category title. This title called gestita (ገጽታ) uses to realize to retrieve all aspects without miss any more.

We also conduct experiment for Naive Bayes supervised approach performance evaluation. Here we prepare train (690 opinions) and test (107 opinions) data based on 80% and 20% of collected hotel review respectively. The performance evaluation for positive, negative and neutral opinions was performed. Therefore generally average performance of test result for this naïve Bayes hotel review classification is that **75.68%** precision, **87.05%** recall and **77.86%** F-measure. Hence less performance than bootstrapping approaches.

6.2. Recommendation as Future Work

Aspect Based Opinion Summarization is very complex task which needs knowledge's like Part-of Speech Tag (POS), syntactic, semantic, language grammar knowledge, information retrieve, information extraction, different algorithms (dictionary based, rule based, supervised, semi supervised, unsupervised and etc.) and other type of knowledge's which are complex to deal with. Applying these all kind of knowledge's needed for aspect based opinion summarization is labor intensive and time consuming, and as a result hard to achieve. Therefore there are many things needed to improve our system efficiency and effectiveness due to the reason that Amharic language has not enough resourced language. The following are some of the recommendations we suggest to be done in the future to increase the performance of the proposed system.

In the bootstrapping semi supervised method we were examine some problems such as 'seed selection' and 'semantic change'. Choosing 'seeds' are critical step in bootstrap. Some are chose frequently occurring words in their corpus and others pick up randomly or select by linguistic analyzer. None of these methods is quite satisfactory. The impact of seed set noise is reflects on the final performance. No general agreement regarding exactly how many seeds are necessary for a given task. Because of these reason in the future the 'seed set' selection should be moderate either by using the combination linguistic expert and frequently occurring word or discovery other seed choice mechanisms. For supervised naive Bayes method we use less training data due to different constraints even if it asks huge amount of training data. Therefore the domain specific training data should be constructed to achieve better results by Naive Bayes algorithm.

In the future comparative opinion mining shall be addressed to satisfy the opinion holders or customers. This Amharic aspect opinion summarization is focus only presenting the summary of customer reviews given towards to particular product aspects/features within our domain 'hotel'. When we evaluate this system by evaluators, they talked to us they also wants knowing best hotel from others. For such comparative purpose, comparative opinion summarizing should be studied and developed in the future.

In this research we employed Bootstrapping and Naïve Bayes methods to classify Amharic opinionated text document and hotel aspects will identify by categorizing. In the upcoming it shall be done by other unsupervised machine learning algorithms for short best solution. To detect aspects categorized by titles method or approach is not good method. Because it asks the

data should be organized before given to the system. So this is time consuming and tedious work. Therefore it also needed to improve by other aspect detection methodology in the next coming researches. The other things to be worked for the future in Ethiopic (Amharic language) opinion mining and summarizing must address opinion question answering and Multi lingual opinion mining and summarization. It also should be done for different domains like politics, tourism, sport, electronic device, agriculture, trade, health, education, and others burning issue for our country which will facilitate development of Ethiopia as I recommend.

References

- [1] B. Liu and S. M. Street, "Opinion mining," Unpublish, University of Illinois, no. 1, pp. 1–7, 2009.
- [2] H. D. U. K. Kim, "Comprehensive Review of Opinion Summarization," Unpublish, University of Illinois, pp. 1–30, 2012.
- [3] K. Lerman, S. Blair-goldensohn, and R. Mcdonald, "Sentiment Summarization : Evaluating and Learning User Preferences," 2005.
- [4] H. D. Kim et al., "Opinion Summarization: Opinion Mining -The Area of Study" 2010.
- [5] K. Khan, "Mining opinion components from unstructured reviews : A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 3, pp. 258–275, 2014.
- [6] B. Liu, and Lei Zhang, " A SURVEY OF OPINION MINING AND SENTIMENT ANALYSIS," Unpublished, University of Illinois, pp. 1–49, 2012.
- [7] Y. Kim and S. R. Jeong, "Opinion-Mining Methodology for Social Media Analytics," vol. 9, no. 1, pp. 391–406, 2015.
- [8] K. V. P. D, "Aspect-based Opinion Mining : A Survey," vol. 106, no. 3, pp. 21–26, 2014.
- [9] H. D. Kim, D. H. Park, V. G. V. Vydiswaran, and C. Zhai, "Opinion Summarization Using Entity Features and Probabilistic Sentence Coherence Optimization : UIUC at TAC 2008 Opinion Summarization Pilot Data preprocessing Document processing," 2008.
- [10] L. V Avanc *et al.*, "A Qualitative Analysis of a Corpus of Opinion Summaries based on Aspects," pp. 62–71, 2015.
- [11] C. Zong et al., "Random Walks for Opinion Summarization," *Natural language processing Lab, School of Computer science and Technology, Soochow University,China*, pp. 430–437, 2014.
- [12] V. Ravi and K. Raghuvver, "Web User Opinion Analysis for Product Features Extraction and Opinion Summarization," *International Journal of Web & Semantic Technology*, vol. 3, no. 4, pp. 69–82, 2012.
- [13] N. Kurian and S. Asokan, "Summarizing User Opinions : A Method for Labeled-Data Scarce Product Domains," *Procedia - Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 93–100, 2015.
- [14] N. Makadia, A. Chaudhuri, and S. Vohra, "ASPECT-BASED OPINION SUMMARIZATION FOR DISPARATE FEATURES," *IJARIIIE*, no. 3, pp. 3732–3739, 2016.
- [15] P. Beineke, T. Hastie, C. Manning, S. Vaithyanathan, and S. Jose, "An exploration of sentiment summarization," *American Association for Artificial Intelligence*, 2003.
- [16] W. Kasper and M. Vela, "Sentiment Analysis for Hotel Reviews," *proceedings of the computational Linguistics Applicaions Conference*, vol. 231527, pp. 45–52, 2011.
- [17] A. Gupta, T. Tenneti, and A. Gupta, "Sentiment based Summarization of Restaurant Reviews Final Project Report CS 224N," pp. 1–11, 2009.
- [18] E. Marrese-taylor, J. D. Velásquez, and F. Bravo-marquez, "Expert Systems with Applications A novel deterministic approach for aspect-based opinion mining in tourism products reviews," *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7764–7775, 2014.
- [19] N. Uvarova, "Abstractive microblogs summarization", Master Thesis, Unpublished, Gjovik University, Norway, 2015.
- [20] W.Maharani et al., "ASPECT-BASED OPINION SUMMARIZATION : A SURVEY," *Journal of*

- theoretical and Applied of Information Technology*, vol. 95, no. 2, pp. 448–456, 2017.
- [21] N. Makadia, "Feature-based Opinion Summarization : A Survey," *International Journal of Engineering Development and Reaserch*, vol. 4, no. 2, pp. 669–673, 2016.
- [22] M. A. Ibrahim and N. Salim, "Sentiment Analysis of Arabic Tweets : With Special Reference Restaurant Tweets," *International Journal of Computer science Trends and Techology*, vol. 4, no. 3, pp. 173–179, 2016.
- [23] D. Tapucu, M. Husani, A. Kocyigit, and H. Lee, "OBOME - ONTOLOGY BASED OPINION MINING IN," vol. 2012, pp. 134–145, 2012.
- [24] V. S. Rajput, " An Overview of Use of Natural Language Processing in Sentiment Analysis based on User Opinions," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 6, no. 4, pp. 594–598, 2016.
- [25] M. Hu, B. Liu, and S. M. Street, "Mining and Summarizing Customer Reviews," 2004.
- [26] E. Marrese-taylor and J. D. Vel, "A Novel Deterministic Approach for Aspect-Based Opinion Mining in Tourism Products Reviews".
- [27] Y. M. Li and T. Y. Li, "Deriving marketing intelligence over microblogs," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, pp. 1–10, 2011.
- [28] V. Hangya, "SZTE-NLP : Aspect Level Opinion Mining Exploiting Syntactic Cues," no. SemEval, pp. 610–614, 2014.
- [29] P. a Rangari and P. K. C. Waghmare, "A Survey on Aspect based Opinion Mining," vol. 4, no. 3, pp. 517–520, 2015.
- [30] P. Zhao, X. Li, and K. Wang, "Feature extraction from micro-blogs for comparison of products and services," *Lect. Notes Comput. Sci.*, vol. 8180 LNCS, no. PART 1, pp. 82–91, 2013.
- [31] S. Poria, E. Cambria, A. Gelbukh, and C. Gui, "A Rule-Based Approach to Aspect Extraction from Product Reviews".
- [32] M. Qiu, L. Yang, and J. Jiang, "Mining User Relations from Online Discussions using Sentiment Analysis and Probabilistic Matrix Factorization," no. June, pp. 401–410, 2013.
- [33] S. S. Htay and K. T. Lynn, "Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews," vol. 2013, no. 3, 2013.
- [34] Keshav R, et al., "Content based Recommender System on Customer Reviews using Sentiment Classification Algorithms," *International Journal of Computer Science and Information Technologies*, Vol. 5 (3), Pp. 4782-4787, 2014.
- [35] H. Kansal and D. Toshniwal, "Aspect based summarization of context dependent opinion words," *Procedia - Procedia Comput. Sci.*, vol. 35, pp. 166–175, 2014.
- [36] Li Z huang, Feng J ing and Xiao-Yan Z hu, "Movie Review Mining and Summarization," ACM, 2006.
- [37] M. K. Dalal and M. A. Zaveri, "Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews," vol. 2013, 2013.
- [38] L. Zhuang, "Movie Review Mining and Summarization," pp. 43–50.
- [39] A. K. Samha, "Aspect-Based Opinion Mining Using Dependency Relations," vol. 4, no. 1, pp. 113–123, 2016.
- [40] Q. Li, Z. Jin, C. Wang, and D. Dajun, "Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems," *Knowledge-Based Syst.*, vol. 107, pp. 289–300, 2016.

- [41] B. Cristian, "Opinion Summarization for Hotel Reviews of Bucharest," ACM proceeding.
- [42] B. Samei, "Multi-Document Summarization Using Graph-Based Iterative Ranking Algorithms and Information Theoretical Distortion Measures," pp. 214–218, 2014.
- [43] S. Gerani, Y. Mehdad, G. Carenini, R. T. Ng, and B. Nejat, "Abstractive Summarization of Product Reviews Using Discourse Structure," pp. 1602–1613, 2014.
- [44] D. Dhanush, A. K. Thakur, and N. P. Diwakar, "Aspect-based Sentiment Summarization with Deep Neural Networks," vol. 5, no. 5, pp. 371–375, 2016.
- 45 A. Alemu, and S. Eyassu, "Classifying *Amharic Webnews*," to appear in Information Retrieval, springer verlag.
- 46 Ephrem Alamerew, "Automatic Annotation of Opinionated Amharic Text for Opinion Mining," unpublished thesis Debre Berhan University.
- 47 S. Brody and N. Elhadad, "An Unsupervised Aspect-Sentiment Model for Online Reviews," Human Language Technologies: The 2010 Annual Conference of the North American, Los Angeles, California, Pp. 804–812, June 2010.
- 48 L.Zhang and B.Liu, "Aspect and Entity Extraction for Opinion Mining," In Proceedings of IEE, 2013.
- 49 H.Armstrong, "Machines *THAT Learn in the Wild*," Nesta, 2015.
- 50 O. Simeone, "A Brief Introduction to Machine Learning for Engineers," Pp.1–201, 2017.
- 51 S.Gebremeskel, "Sentiment Mining Model for Opinionated Amharic Texts," Master Thesis, Unpublished, Addis Ababa University, 2010.
- 52 C.Linand P.Chao, "Tourism-Related Opinion Detection and Tourist-Attraction Target Identification," The Association for Computational Linguistics and Chinese Language Processing, Vol. 15, No. 1, pp. 37-60, March 2010.
- 53 D.Maynard, G.Gossen, A.Funk and M.Fisichella, "Should I Care about Your Opinion? Detection of Opinion Interestingness and Dynamics in Social Media," Future Internet, Vol. 6, Pp.457-481, 2014.
- 54 B.Liu, "Sentiment Analysis and Opinion Mining," Morgan Claypool Publishers, 2012.
- 55 ኔታሁን አማራ, "ዘመናዊ የአማርኛ ስዋሰኖች በቀላል አቅራቢ" አልፋ አሳታሚ ድርጅት, አዲስ አበባ, 2004.
- 56 L. Ku, Y. Liang and H. Chen, "Opinion Extraction, Summarization and Tracking in News and Blog Corpora," American Association for Artificial Intelligence, National Taiwan University, 2006.
- 57 O. Appel et al., "Main Concepts, State of the Art and Future Research Questions in Sentiment Analysis," Acta Polytechnica, Hungarica, Vol. 12, No. 3, 2015.
- 58 I. Matsuno et al., "Aspect-based Sentiment Analysis using Semi-supervised Learning in Bipartite Heterogeneous Networks," Journal of Information and Data Management, Vol. 7, No. 2, Pp.141–154, August 2016.
- 59 Md Shad Akhtar, Asif Ekbal, and Pushpak Bhattacharyya, "Aspect Based Sentiment Analysis: Category Detection and Sentiment Classification for Hindi," Department of Computer Science & Engineering, Indian Institute of Technology, Patna, India-801103, 2016.
- 60 D. Manning, P. Raghavan, and H. Schutze, "Introduction to Information Retrieval," Published in the United States of America by Cambridge University Press, New York, 2008.

- 61 Wegderes Tariku, "ASPECT BASED SUMMARIZATION OF OPINIONATED AFAAN OROMOO NEWS TEXT," Master Thesis, Unpublished, Debre Birhan University, August 2017.
- 62 ባዩ ይማም, "የአማርኛ ሰዋሰድ," ኦዲዮ ኦቢቲቪ, 1987.
- 63 Alemebante Mulu and Vishal Goyal, "Amharic Text Predict System for Mobile Phone," International Journal of Computer Science Trends and Technology (IJCST) – Volume 3, Issue 4, Jul-Aug 2015.
- 64 Ding Ying and Jing Jiang, "Towards Opinion Summarization from Online Forums," Proceedings of Recent Advances in Natural Language Processing: 10th RALP, Hissar, Bulgaria, September, 2015.
- 65 Sonal Singhal, "COMPREHENSIVE REVIEW OF OPINION SUMMARIZATION," International Journal of Computer Engineering and Applications, Volume VI, Issue II, May 2014.
- 66 Ahmad Kamal, "Review Mining for Feature based Opinion Summarization and Visualization," International Journal of Computer Applications, Vo. 119, No.17, Pp. 0975 – 8887, June 2015.
- 67 Martin Potthast and Steffen Becker, "Opinion Summarization of Web Comments," Proceedings of the 32nd European Conference on Information Retrieval, Milton Keynes, UK, pp. 668-669, 2010.
- 68 Hitoshi Nishikawa, Takaaki Hasegawa, Yoshihiro Matsuo and Genichiro Kikui, "Opinion Summarization with Integer Linear Programming Formulation for Sentence Extraction and Ordering," Coling 2008: Poster Volume, Pp.910–918, Beijing, August, 2010.
- 69 G. Somprasertsri and P. Lalitrojwong, "Mining Feature Opinion in Online Customer Reviews for Opinion Summarization," Journal of Universal Computer Science, vol.16, no. 6, Pp. 938-955, 2010.
- 70 S.Poria et al., "Aspect extraction for opinion mining with a deep convolution neural network," Knowledge-Based Systems 108, Pp. 42–49, 2016.
- 71 M. Asghar et al., "A Review of Feature Extraction in Sentiment Analysis," Journal of Basic and Applied Scientific Research, Vol.4, Pp.181-186, 2014.
- 72 Chinsha T C and S. Joseph, "Aspect based Opinion Mining from Restaurant Reviews," Advanced Computing and Communication Techniques for High Performance Applications ICACCTHPA, 2014.
- 73 M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," ACM, KDD, August 22–25, 2004.
- 74 Suganya S, Sureka K, and Vishnupriya P., "INTEGRATING ASPECTS BASED ON OPINION MINING FOR PRODUCT REVIEW," International Conference on Information Engineering, Management and Security, Pp.01-06, 2015.
- 75 Alemebante Mulu and Vishal Goyal, "Amharic Text Predict System for Mobile Phone," International Journal of Computer Science Trends and Technology (IJCST) – Volume 3, Issue 4, Pp.113-118, Jul-Aug 2015.
- 76 Wondwossen Mulugeta and Michael Gasser, "Learning Morphological Rules for Amharic Verbs Using Inductive Logic Programming," Workshop on Language Technology for Normalisation of Less-Resourced Languages (SALTMIL8/AfLaT2012), 2012.
- 77 Seid Muhie and Mulugeta Libsie, "TETEYEQ: Amharic Question Answering For Factoid Questions", In Proceedings of Information Retrieval and Information Extraction for Less Resourced Languages Conference (IEIR-LRL), Donostia, Spain, September 7th, 2009.

- 78 Mulualem Wordofa ,"*SEMANTIC INDEXING AND DOCUMENT CLUSTERING FOR AMHARIC INFORMATION RETRIEVAL*,"Master Thesis, Unpublished, Addis Ababa University, 2010.
- 79 Addis Ashagrye ,"*AUTOMATIC SUMMARIZATION FOR AMHARIC TEXT USING OPEN TEXT SUMMARIZER*," Master Thesis, Unpublished, Addis Ababa University, 2013.
- 80 Mersehaizen Wolde Kirkos, "*የአማርኛ ሰዋሰድ*," Berihanena Selam Printing Press, Addis Ababa, 1934.
- 81 Baye Yimam, "*የአማርኛ ሰዋሰድ*", Ethiopian Materials Production and Distribution Agency (E.M.P.D.A), Addis Ababa, 1986.
- 82 Charles William Isenberg, "*GRAMMER OF THE AMHARIC LANGUAGE*" university microfilm, INC. a subsidiary of Xerox corporation Ann Arbor, 1965.
- 83 Bo Pang and Lillian Lee, "*Opinion mining and sentiment analysis*" Foundations and Trends in Information Retrieval, Vol. 2, Pp.1–135, No 1-2 (2008).
- 84 Z. Guan et al., "Weakly-Supervised Deep Learning for Customer Review Sentiment Classification," *Proceeding of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, Pp.3719-3725, 2016.
- 85 S.Guha, A.Joshi, and V.Varma, "*SIEL: Aspect Based Sentiment Analysis in Reviews*," *4th Joint Conference on Lexical and Computational Semantics*, Denver, Colorado, USA, may 2015.
- 86 M. jiang, P.Cui and C. faloutsos, "*Suspicious Behavior Detection: Current Trends and Future Directions*," *Published by the IEEE Computer Society*, 2016.
- 87 K.Bhattacharjee and Li.Petzold, "*Detecting Opinions in a Temporally Evolving Conversation on Twitter*," 2015.
- 88 T.tlahun, "*Opinion Mining From Amharic Blog*,"Master Thesis, Unpublished, Addis Ababa University, 2013.
- 89 Dim En Nyaung and Thin Lai Lai Thein, "*Feature-Based Summarizing and Ranking from Customer Reviews*," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 9, no. 3, 2015.
- 90 Amani K Samha, Yuefeng Li And Jinglan Zhang, "*Aspect-Based Opinion Extraction From Customer Reviews*," Unpublished, Queensland University Of Technology, Brisbane, Australia, 2013.
- 91 Shu Zhang, Yingju Xia, Yao Meng, and Hao Yu, "*A Bootstrapping Method for Finer Grained Opinion Mining Using Graph Model*," *23rd Pacific Asia Conference on Language, Information and Computation*, Pp. 589–595, 2009.
- 92 Zhen Hai, Kuiyu Chang and Gao Cong, "*One Seed to Find Them All: Mining Opinion Features via Association*," *ACM, CIKM'12*, Maui, HI, USA, October 29–November 2, 2012.

Appendix

A. Sample of Amharic Review on Hotel Domain Collected From Customer

የጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
የደንበኞች ጥያቄዎች
የጉዞ ተሰጪ አገልግሎት ለማረጋገጥ

1. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
2. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
3. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
4. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
5. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ

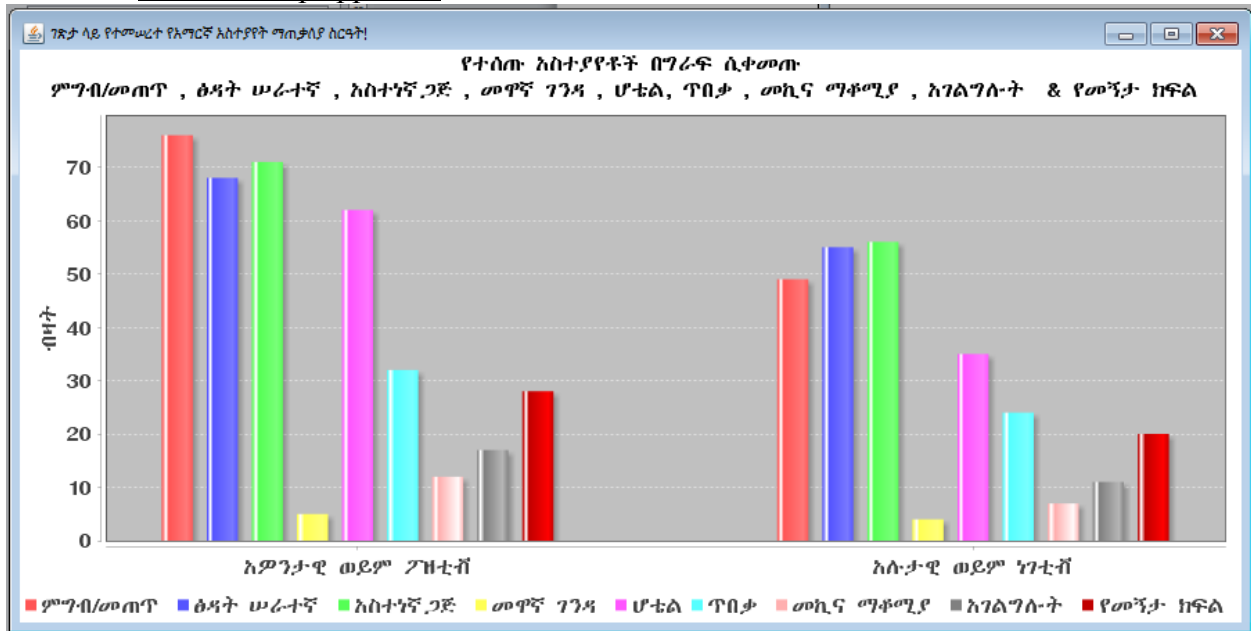
2

የጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
የደንበኞች ጥያቄዎች
የጉዞ ተሰጪ አገልግሎት ለማረጋገጥ

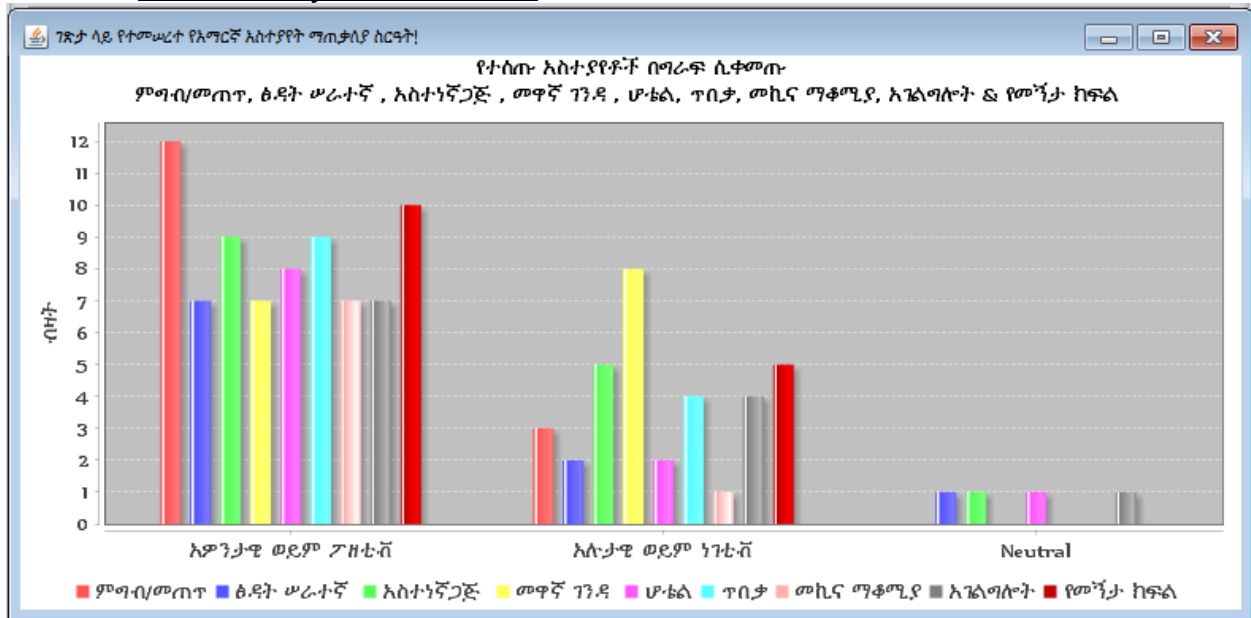
1. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
2. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
3. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
4. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ
5. ለጉዞ ተሰጪ አገልግሎት ለማረጋገጥ

B. Sample of result visualization with graph

➤ For Bootstrap approach



➤ For Naive Bayes Classification



No	Opinion	Category	Aspect
1	ፅዳት አጠበቀች ደስ ይላል::	Positive	ምግብ/መጠጥ
2	ሁሉም የምግብ ዓይነቶች አሉ ደስ ይላል::	Positive	ምግብ/መጠጥ
3	ከምግባቸው ጥብስና ከትፎ ተመችቶኛል	Positive	ምግብ/መጠጥ
4	ጫው አታብዙ ደስ ይላል::	Positive	ምግብ/መጠጥ
5	መስተንግዳቹ ተመችቶኛል::	Positive	ምግብ/መጠጥ
6	ሁሉም የምግብ ዓይነቶች ስላላቹ ሳላደንቅ አላልፍም::	Positive	ምግብ/መጠጥ
7	ከሌሎች ሆቴልች ምግባቸው አረፍ ነው ::	Positive	ምግብ/መጠጥ
8	ስጋቸው አረፍ ነው::	Positive	ምግብ/መጠጥ
9	የጫው ነገር ግን ብታሻሽሉ::	Negative	ምግብ/መጠጥ
10	የበግ አፍሰቶው ዐረፍ ነው ::	Positive	ምግብ/መጠጥ
11	ፒዳችሁን ውድጄላችቸለሁ::	Negative	ምግብ/መጠጥ
12	ብፈያችሁ ጣት ያስቆሮጥማል::	Negative	ምግብ/መጠጥ
13	ምግቡ አረፍ ነው ::	Positive	ምግብ/መጠጥ

C. Sample of source code Amharic Aspect Opinion Summarization Using Bootstrap on Hotel Domain

For GUI

```

javax.swing.GroupLayout jPanel3Layout = new javax.swing.GroupLayout(jPanel3);
jPanel3.setLayout(jPanel3Layout);
jPanel3Layout.setHorizontalGroup(
    jPanel3Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
        .addComponent(jFileChooser1, javax.swing.GroupLayout.PREFERRED_SIZE, 0, Short.MAX_VALUE)
);
jPanel3Layout.setVerticalGroup(
    jPanel3Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
        .addComponent(jFileChooser1, javax.swing.GroupLayout.PREFERRED_SIZE, 0, Short.MAX_VALUE)
);
jPanel4.setBorder(javax.swing.BorderFactory.createTitledBorder(null, "አስተያየቱን ይክፈቱ",
javax.swing.border.TitledBorder.DEFAULT_JUSTIFICATION, javax.swing.border.TitledBorder.DEFAULT_POSITION, new
java.awt.Font("Power Geez Unicode1", 3, 14))); // NOI18N
jPanel4.setFont(new java.awt.Font("Nyala", 1, 14)); // NOI18N
btnBrowse.setFont(new java.awt.Font("Power Geez Unicode1", 1, 14)); // NOI18N
btnBrowse.setText("ፋይሉን ይክፈቱ");
btnBrowse.setCursor(new java.awt.Cursor(java.awt.Cursor.HAND_CURSOR));
btnBrowse.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        btnBrowseActionPerformed(evt);
    }
});
btnSummarize.setFont(new java.awt.Font("Power Geez Unicode1", 1, 14)); // NOI18N
btnSummarize.setText("በግራፍ ያሳጥፍ");
btnSummarize.setCursor(new java.awt.Cursor(java.awt.Cursor.HAND_CURSOR));
btnSummarize.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        btnSummarizeActionPerformed(evt);
    }
});
btnClear.setFont(new java.awt.Font("Power Geez Unicode1", 1, 14)); // NOI18N
btnClear.setText("ያጥፍ");
btnClear.setCursor(new java.awt.Cursor(java.awt.Cursor.HAND_CURSOR));
btnClear.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        btnClearActionPerformed(evt);
    }
});
}
});
javax.swing.GroupLayout jPanel4Layout = new javax.swing.GroupLayout(jPanel4);
jPanel4.setLayout(jPanel4Layout);
jPanel4Layout.setHorizontalGroup(
    jPanel4Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
        .addComponent(btnBrowse, javax.swing.GroupLayout.DEFAULT_SIZE, 513, Short.MAX_VALUE)

```



```
.addComponent(btnSummarize, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE,
Short.MAX_VALUE)
.addComponent(btnClear, javax.swing.GroupLayout.Alignment.TRAILING, javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE)
);
jPanel4Layout.setVerticalGroup(
jPanel4Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
.addComponent(btnBrowse, javax.swing.GroupLayout.PREFERRED_SIZE, 50, javax.swing.GroupLayout.PREFERRED_SIZE)
.addPreferredGap(javax.swing.LayoutStyle.ComponentPlacement.RELATED)
.addComponent(btnSummarize, javax.swing.GroupLayout.PREFERRED_SIZE, 50, javax.swing.GroupLayout.PREFERRED_SIZE)
.addPreferredGap(javax.swing.LayoutStyle.ComponentPlacement.UNRELATED)
.addComponent(btnClear, javax.swing.GroupLayout.PREFERRED_SIZE, 50, javax.swing.GroupLayout.PREFERRED_SIZE)
.addGap(0, 0, Short.MAX_VALUE)
);
javax.swing.GroupLayout layout = new javax.swing.GroupLayout(getContentPane());
getContentPane().setLayout(layout);
layout.setHorizontalGroup(
layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
.addGroup(layout.createSequentialGroup()
.addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.TRAILING, false)
.addComponent(jPanel2, javax.swing.GroupLayout.Alignment.LEADING, javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE)
.addComponent(jPanel1, javax.swing.GroupLayout.Alignment.LEADING, javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE))
.addPreferredGap(javax.swing.LayoutStyle.ComponentPlacement.RELATED)
.addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
.addComponent(jPanel4, javax.swing.GroupLayout.Alignment.TRAILING, javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE)
.addComponent(jPanel3, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE,
Short.MAX_VALUE)))
);
layout.setVerticalGroup(
layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
.addGroup(javax.swing.GroupLayout.Alignment.TRAILING, layout.createSequentialGroup()
.addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING, false)
.addComponent(jPanel1, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE,
Short.MAX_VALUE)
.addComponent(jPanel3, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE,
Short.MAX_VALUE))
.addGap(18, 18, 18)
.addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING, false)
.addComponent(jPanel2, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE,
Short.MAX_VALUE)
.addComponent(jPanel4, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE,
Short.MAX_VALUE))
.addContainerGap(javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE)
);
pack();
} // </editor-fold> // GEN-END: initComponents
private void btnClearActionPerformed(java.awt.event.ActionEvent evt) // GEN-FIRST: event_btnClearActionPerformed
// TODO add your handling code here:
if (!(txtDisplay.getText().equals("")) {
txtDisplay.setText("");
} else {
JOptionPane.showMessageDialog(null, "ፋይል ጠፍቷል!", "ማረጃ", 1);
}
} // GEN-LAST: event_btnClearActionPerformed
private void btnSummarizeActionPerformed(java.awt.event.ActionEvent evt) // GEN-FIRST: event_btnSummarizeActionPerformed
// TODO add your handling code here:
btnSummarize.disable();
btnClear.enable(false);
try{
```

For seed opinion word lexicon as bootstrap

```

try(BufferedReader br = new BufferedReader(new
FileReader("C:\\Users\\ABU ABDELAH\\Desktop\\amharic_opSummary_Datasets\\modifiedAmharicOpinionSummarization\\Opinion
Summarize\\src\\opinion\\summarize\\positiveKey.txt")) {
    StringBuilder sb = new StringBuilder();
    String line = br.readLine();
while (line != null) {
    sb.append(line);
    sb.append(System.lineSeparator());
    line = br.readLine();
    }
    String textPositive = sb.toString();
    positiveKey = textPositive.split("\n"); //includes '\0' at the end
    }
    try(BufferedReader br = new BufferedReader(new
FileReader("C:\\Users\\ABU ABDELAH\\Desktop\\amharic_opSummary_Datasets\\modifiedAmharicOpinionSummarization\\Opinion
Summarize\\src\\opinion\\summarize\\negativeKey.txt")) {
    StringBuilder sb = new StringBuilder();
    String line = br.readLine();
    while (line != null) {
        sb.append(line);
        sb.append(System.lineSeparator());
        line = br.readLine();
    }
    String textNegative = sb.toString();
    negativeKey = textNegative.split("\n"); //includes '\0' at the end
    }
    if (!txtDisplay.getText().equals("")) {
        String[] titles = txtDisplay.getText().split("ገጽ:ጋ-: ");
        category = new String[titles.length];
        pos = new int [titles.length];
        neg = new int [titles.length];
        for(int t=1;t<titles.length;t++){ //titles
            negative=0;
            positive=0;
            String[] opinion = titles[t].split("\n");
            title = opinion[0]; //subtile
            category[t]=opinion[0];
            for (int opn= 1; opn < opinion.length; opn++) {
                keyWords = opinion[opn].split(" ");
                for (int i = 0; i < keyWords.length; i++) {
                    for (int j = 0; j < positiveKey.length; j++) {
                        if((positiveKey[j]).contains(keyWords[i])) {
                            positive = positive + 1;
                        } } }
                for (int k = 0; k < keyWords.length; k++) {
                    for (int l = 0; l < negativeKey.length; l++) {
                        if(negativeKey[l].contains(keyWords[k])) {
                            negative = negative + 1;
                        } } }
            }
        neg[t]=negative;
        pos[t]=positive;
    }
    DrawBarChart chart = new DrawBarChart("ገጽ:ጋ ላይ የተመሠረተ የአማርኛ አስተያየት ማጠቃለያ ስርዓት!");
    chart.setFont(new java.awt.Font("Power Geez Unicode1", 1, 14));
    chart.pack();
    RefineryUtilities.centerFrameOnScreen(chart);
    chart.setVisible(true);
    chart.setAlwaysOnTop(rootPaneCheckingEnabled);
    chart.setLocationRelativeTo(jPanel2);
    } else {
        JOptionPane.showMessageDialog(null, "ፋይል አልተከፈተም ስለዚህም አንድ ለገጽ ይጻፍ!", "መረጃ", 1);
    }
}

```

For summarizing aspect opinion visualizing with graph (use jFree chart)

```
public class DrawBarChart extends ApplicationFrame {
```

```

private static final long serialVersionUID = 1L;
public static String topic="";
static {
    ChartFactory.setChartTheme(new StandardChartTheme("JFree/Shadow",
        true));
}
public DrawBarChart(String title) {
    super(title);
    CategoryDataset dataset = createDataset();
    JFreeChart chart = createChart(dataset);
    ChartPanel chartPanel = new ChartPanel(chart);
    chartPanel.setFillZoomRectangle(true);
    chartPanel.setFont(new java.awt.Font("Nyala", 1, 14));
    chartPanel.setMouseWheelEnabled(true);
    chartPanel.setPreferredSize(new Dimension(580, 270));
    setContentPane(chartPanel);
}
private static CategoryDataset createDataset() {
    DefaultCategoryDataset dataset = new DefaultCategoryDataset();
    String [] category=OpinoiSummarize.category;
    int [] pos=OpinoiSummarize.pos;
    int [] neg=OpinoiSummarize.neg;
    String pun="";
    for(int i=1;i<category.length;i++){
        if(i+2==category.length){
            pun = " & ";
            topic=topic+category[i]+pun;
        }
        else if(i+2<category.length){
            pun = ", ";
            topic=topic+category[i]+pun;
        }
        else if(i+1==category.length){
            topic=topic+category[i];
        }
        dataset.addValue(pos[i], category[i], "ፖዘቲቭ");
        dataset.addValue(neg[i], category[i], "ነገረብ");
    }
    return dataset;
}
private static JFreeChart createChart(CategoryDataset dataset) {
    Font font= new java.awt.Font("Nyala", 1, 14);
    JFreeChart chart = ChartFactory.createBarChart(
        "የተሰጡ አስተያየት በግራፍ", null /* x-axis label*/,
        "ብዛት" /* y-axis label */, dataset);
    chart.addSubtitle(new TextTitle(topic, font));
    chart.setBackgroundPaint(Color.white);
    chart.getTitle().setFont(font);
    CategoryPlot plot = chart.getCategoryPlot();
    LegendTitle legend = chart.getLegend();
    legend.setItemFont(font);
    ValueAxis axis2 = plot.getRangeAxis();
    CategoryAxis axis = plot.getDomainAxis();
    axis.setTickLabelFont(font);
    axis2.setTickLabelFont(font);
    legend.setItemFont(font);
    plot.getDomainAxis().setLabelFont(font);
    plot.getRangeAxis().setLabelFont(font);
    NumberAxis rangeAxis = (NumberAxis) plot.getRangeAxis();
    rangeAxis.setStandardTickUnits(NumberAxis.createIntegerTickUnits());
    BarRenderer renderer = (BarRenderer) plot.getRenderer();
    renderer.setDrawBarOutline(false);
    chart.getLegend().setFrame(BlockBorder.NONE);
    return chart;
}
public static void main(String args[]) {

```

```

try {
for (javax.swing.UIManager.LookAndFeelInfo info : javax.swing.UIManager.getInstalledLookAndFeels()) {
    if ("Nimbus".equals(info.getName())) {
        javax.swing.UIManager.setLookAndFeel(info.getClassName());
        break;
    }
}
} catch (ClassNotFoundException ex) {
    java.util.logging.Logger.getLogger(OpinoiSummarize.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
} catch (InstantiationException ex) {
    java.util.logging.Logger.getLogger(OpinoiSummarize.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
} catch (IllegalAccessException ex) {
    java.util.logging.Logger.getLogger(OpinoiSummarize.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
} catch (javax.swing.UnsupportedLookAndFeelException ex) {
    java.util.logging.Logger.getLogger(OpinoiSummarize.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
}
}
java.awt.EventQueue.invokeLater(new Runnable() {
    public void run() {
        new OpinoiSummarize().setVisible(true);
    }
});
}

```

D. List of positive and negative seed opinion word lexicons

Positive aspect opinion word lexicon seeds

No	Hotel aspects	Amharic positive opinion words
1	ምግብ መጠጥ	ውሃው ጥሩ አይደለም
2	አገልግሎት	ሰጋቸው አሪፍ ነው።
3	መዋኛ ገንዳ	ገንዳው በወይቱ ምርጥ ነው
4	ፅዳት ሰራተኛ	ፅዳት ሰራተኛ ጎበዝ ናቸው
5	መኪና ማቆሚያ	መኪና ማቆሚያ ሰፊ በመሆኑ ደስ በሎኛል

Negative hotel aspect opinion word lexicon seeds

No	Hotel aspects	Amharic negative opinion words
1	ጥበቃ	መጥፎ ጥበቃ ነው ያላቸው ቢያሰቡበት
2	መኝታ ክፍል	ጠባብ ነው
3	ሰራተኛ	ሀኒ ላንድ ሆቴል የሚሰሩ ፅዳት ሰራተኞች ሰዎችን አያከብሩም
4	አስተናጋጅ	አስተናጋጆቹ በጣም ስልቹ ናቸው
5	ምግብ	የምግብ ዋጋ በጣም ወድ ነው

Authorization

I authorize the School of Computing, JiT, under Jimma University to lend this thesis for the other institution or individuals for the purpose of scholarly research. I further authorize the Department of Information Technology, Jimma University to reproduce the thesis by photocopy or by other means, in total or in part of at the request of other institutions or individuals for the purpose of their scholarly research work.

Declaration

I, the undersigned, announce that this “**Amharic Aspect Opinion Summarization Using Bootstrap on Hotel Domain**” thesis work is my original work and has not been presented for MSc degree in any other university, and all sources of materials for the thesis have been acknowledged.

SEID HUSSEIN

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



DEBELA TESFAYE (Ph. D.)

Jimma, Ethiopia

June, 15, 2019