TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO: 078MSDSA019

"Aspect Based Sentiment Analysis of Nepali Text using BERT based Model"

SUBMITTED BY

Suprabha Regmi

A THESIS REPORT

SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER ENGINEERING SPECIALIZATION IN DATA SCIENCE AND ANALYTICS

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

LALITPUR, NEPAL

June, 2024

Aspect Based Sentiment Analysis of Nepali Text using BERT based Model

Submitted By

Suprabha Regmi

(078MSDSA019)

Supervised By

Asst. Prof. Dr. Aman Shakya

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Engineering Specialization in Data Science and Analytics

Department of Electronics and Computer Engineering

Institute of Engineering, Pulchowk Campus

Tribhuvan University

Lalitpur, Nepal

June, 2024

COPYRIGHT©

DECLARATION

I declare that the work hereby submitted for Master of Science in Computer Engineering specialization in Data Science and Analytics (MSDSA) at IOE, Pulchowk Campus entitled **"Aspect Based Sentiment Analysis of Nepali Text using BERT based Model"** is my own work and has not been previously submitted by me at any university for any academic award.

I authorize IOE, Pulchowk Campus to lend this thesis to other institution or individuals for the purpose of scholarly research.

Suprabha Regmi

078MSDSA019

Date: June, 2024

RECOMMENDATION

The undersigned certify that they have read and recommended to the Department of Electronics and Computer Engineering for acceptance, a thesis entitled **"Aspect Based Sentiment Analysis of Nepali Text using BERT based Model"**, submitted by **Suprabha Regmi** in partial fulfillment of the requirement for the award of the degree of **"Master of Science in Computer Engineering Specialization in Data Science and Analytics"**.

....

Supervisor: Asst. Prof. Dr. Aman Shakya

Department of Electronics and Computer Engineering,

Institute of Engineering, Tribhuvan University.

....

External Examiner: Assoc. Prof. Birodh Rijal

Everest Engineering College

....

Committee Chairperson: Asst. Prof. Dr. Aman Shakya

Program Coordinator, Msc in Computer Engineering Specialization in Data Science and Analytics,

Department of Electronics and Computer Engineering,

Institute of Engineering, Tribhuvan University.

Date: June, 2024

DEPARTMENTAL ACCEPTANCE

The thesis entitled **"Aspect Based Sentiment Analysis of Nepali Text using BERT based Model"**, submitted by **Suprabha Regmi** in partial fulfillment of the requirement for the award of the degree of **"Master of Science in Computer Engineering Specialization in Data Science and Analytics"** has been accepted as a bonafide record of work independently carried out by her in the department.

....

Assoc. Prof. Dr. Jyoti Tandukar

Head of the Department

Department of Electronics and Computer Engineering,

Pulchowk Campus,

Institute of Engineering,

Tribhuvan University,

Nepal.

ACKNOWLEDGEMENT

ABSTRACT

Aspect based sentiment analysis (ABSA) is a fine-grained sentiment analysis which aims at extracting opinion terms, opinion targets, aspect category and the polarity associated with each aspect category. The majority of research on ABSA is in English, with a small amount of work available in Nepali. For high-resource language, ABSA seems easy as compared to low-resource one because of enough datasets and required tools. Singh et.al.(2020) created Nepali dataset for aspect based sentiment analysis . We have added 1031 nepali text comments related to news and politics category from YouTube to the dataset by previous researcher under same annotation guideline. This work focusses on target oriented opinion word extraction task (TOWE) which is not performed on previous research for ABSA in nepali along with aspect category detection (ACD) and aspect category polarity (ACP) detection task in nepali text. TOWE challenges were previously handled by lexicon-based algorithms and machine learning approaches but because of their reliance on the effectiveness of the handmade features deep learning approaches gained popularity. Bidirectional Encoder Representations from Transformers(BERT) among other pre-trained language models beats existing best results in several NLP tasks by large margin. We present convincing results using NepaliBERT for TOWE task with ROUGE-L score of 0.80 and has experimented on several models viz SVM, LSTM, BiLSTM, CNN and multilingual BERT for aspect category detection task and aspect category polarity detection task with the F1-score of 82.78% and 83.13 % for ACD and ACP respectively. Our method has outperformed current state of the art for aspect category detection task and aspect category polarity detection in nepali text.

Keywords: SA, Aspect based Sentiment Analysis, BERT, Target Oriented Opinion Word Extraction, ACD, ACP.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABSA Aspect Based Sentiment Analysis

ATE Aspect Term Extraction

ACD Aspect Category detection

ACP Aspect Category Polarity

BERT Bidirectional Encoder Representations from Transformers

Bi-LSTM Bidirectional Long Short-Term Memory

CNN convolutional neural network

DL Deep Learning

DNN Deep Neural Network

GRU Gated Recurrent Unit

IE Information Extraction

LSTM Long Short Term Memory

ML Machine Language

NER Named Entity Recognition

NN Neural Network

NLP Natural Language Processing

RNN Recurrent Neural Network

SA Sentiment Analysis

SemEval Semantic Evaluation

SVM Support Vector Machine

TD-LSTM Target Dependent-Long Short Term Memory

TOWE Target Oriented Opinion Word Extraction

CHAPTER 1

INTRODUCTION

In this chapter background related to sentiment analysis, its various variations and their importance will be discussed. The major focus will be towards aspect based sentiment analysis (ABSA) and its various subtasks like Aspect category detection (ACD), Aspect category polarity (ACP) detection, target term extraction, opinion term extraction and most important is extraction of opinion term and its corresponding target term.

1.1 Background

Others opinion have always mattered for we humans. With the growth of internet and its usage various online platforms are there to solve the problem in our daily lives. Growing reviews on such platforms and in various forms viz text, emojis, video review, audio review etc has made it easy for those who faces dilemma whether to approach that platform or not. So in general one can say that the opinions are being made available to strangers via internet and the strangers in respect with others opinion made their opinion so the opinion keeps on growing and is recorded. This growing opinion or review and its importance to all of us has emerged a field of study known as sentiment analysis also called as opinion mining. Author **Pang et al.** in (2) defines Sentiment analysis as the process of evaluating human emotions, opinions, reviews expressed in text to detect the writer's mental outlook towards a particular event, topic, product, service, etc. and assign a relevant sentiment. It is also called opinion mining or subjectivity analysis. Its main goal is to find out the opinion of people on certain topic. It not only has received attention from academic field but also from the industry because it helps in providing the feedback information of customers by online reviews. It helps various businesses in making marketing strategies that is needed to remain in the market in this highly competitive time. It helps in finding out the emotions of customers. It is used to extract opinions that is present within texts, sentences, or

documents. The opinion of the given text is classified into positive, negative, and neutral by it. It is impossible to manually process the huge data on opinions due to the enormous amount of text. **Haselmayer et al.** (3) in their paper mentioned the importance of sentiment analysis(SA) of text data in social science mainly in politics. **Rodriguez et al.** (4) in their review paper in sentiment analysis from social media platforms mentioned that active research is going on in sentiment analysis area since 2008 as huge number of papers on this area are seen. Author also mentioned that due to the ease of obtaining data from social media, SA is a growing field of study in social networks.

Sentiment analysis is the field that tries to give machines and computer software the ability to understand the emotions of the user. Sentiment analysis means to examine and determine the intent or feeling behind a text, speech, or other form of communication. SA is a field (5) that combines artificial intelligence (AI) , natural language processing (NLP) and information retrieval (IR). As per **liu et al.**(6) SA is classified into three levels namely the document level, sentence level and entity or aspect level SA. This aspect level SA performs fine-grained SA. The main idea it is based upon is that the review text consists of a opinion term also called sentiment term (negative or positive) and a target on which the opinion is targeted to. Aspect based sentiment analysis(ABSA) is a fine-grained SA that involves multiple subtasks like target term extraction, opinion term extraction, target oriented opinion word extraction, aspect category detection, aspect category polarity detection etc. SA is expanded upon to a more detailed level by ABSA. Like SA, ABSA has became popular both in industries and academic field. Targeted sentiment analysis (TSA) is another variety of SA and there are various tasks under the big umbrella of TSA. In any type of sentiment analysis the major task is to find the sentiment or polarity of the text but the level of granularity increases in ABSA, TSA, Targeted ABSA etc. In ABSA, the nature of data under work determines which task is appropriate for that data.

ABSA aims in determining the different aspects in the text review and its corresponding sentiment. Targeted aspect based sentiment analysis (TABSA) combines the principles of ABSA and also include the target associated with the aspect and for which sentiment is expressed in the text review. Targeted sentiment analysis

(TSA) investigates the classification of opinion polarities towards certain target entity mention in given sentences.

Various tasks inside ABSA are depicted in following diagram.
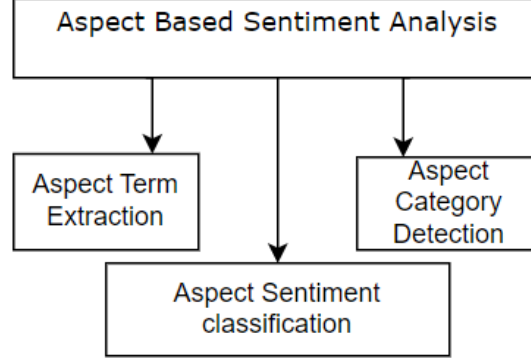
Figure 1.1: General ABSA sub-tasks

Aspect term extraction (ATE) deals with identification of aspects from a given sentence. Aspects are noun or noun phrases in a text. Some author named aspect term extraction task as opinion target extraction task, naming differs from author to author. Author **Aminu Da et al.** (7) mentions that ABSA's main task is extraction of aspects from the review. Author **Arjun Mukherjee et al.** (8) mentions that Aspect category detection(ACD) deals with clustering the similar aspect terms into categories. Aspect sentiment classification is another sub-task in ABSA. After the extraction of aspect term in a review sentence and categorization of aspect terms into categories aspect sentiment classification is done. There are various views regarding tasks of aspect based sentiment analysis as presented by various authors in their research.

Pontiki et al. (9) in 2014 presented four subtasks under ABSA task namely Aspect term extraction(SB1), Aspect term polarity(SB2), Aspect category detection(SB3) and Aspect category polarity(SB4). For SB1 and SB2 subtasks datasets on laptops and restaurants domain were provided whereas for SB3 and SB4 subtasks dataset for restaurant reviews were only provided. SB1 subtask identify all aspect terms present in a sentence, aspect terms are normally nouns or noun phrases. SB2 subtask determine polarity of each aspect term. Aspect category detection subtask identify aspect categories, these are coarser than aspect terms and can be inferred through the adjectives present in the text. The author gave example of "Expensive

but Delicious" here, the aspect categories are PRICE and FOOD but they are not instantiated through specific aspect terms rather inferred from the adjectives 'expensive' and 'delicious'. So, subtasks SB1 and SB2 were handled independently. SB4 subtask determines polarity of aspect categories .

Pontiki et al. (10) in 2015 presented subtasks for SE-ABSA15 task namely In-domain ABSA and Out-of-domain ABSA. Subtask 1 focuses on in-domain Aspect-Based Sentiment Analysis (ABSA) for laptop or restaurant reviews. Participants are required to identify opinion tuples, consisting of Aspect Category(entity and attribute pairs) , Opinion Target Expression (OTE) representing the linguistic expression referring to the reviewed entity and Sentiment Polarity (positive, negative, or neutral) assigned to each entity-attribute pair. In Subtask2 Out-of-domain Aspect-Based Sentiment Analysis (ABSA) participants need to assess their systems in an unfamiliar domain (hotel reviews) without any training data.

Pontiki et al. (11) in 2016 presented multiple subtasks and slots where participants have the freedom to choose domains, languages, and specific tasks for SE-ABSA16 task. Subtask 1 (SB1) is Sentence-level ABSA in which the participants must analyze opinionated texts about a target entity, identifying opinion tuples with Aspect category (entity and attribute pairs), Opinion target expression (OTE), and Sentiment polarity. The identification of Slot2 values is specifically required in the restaurants, hotels, museums, and telecommunications domains. SB2 is Text-level ABSA where the goal is to extract cat, pol tuples summarizing opinions from customer reviews about a target entity. cat represents Aspect Category and pol can be positive, negative, neutral or conflict. SB3 is Out-of-domain ABSA where participants test their systems in domains with no provided training data, and the specific domains are unknown until the evaluation period begins. From (9), (10), (11) we can conclude that there are three main tasks in ABSA: aspect category detection, opinion and its corresponding target extraction, and aspect category polarity detection.

Author **Zhu et al.** in (12) redivide the ABSA task into the information extraction(IE) task and the sentiment classification(SC) task.
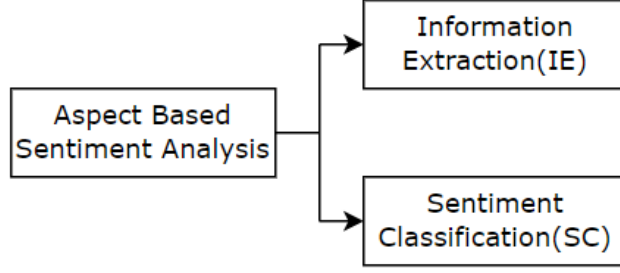
Figure 1.2: ABSA subtasks.

The author further divided IE task into various subtasks. The subtasks are depicted in following figure.
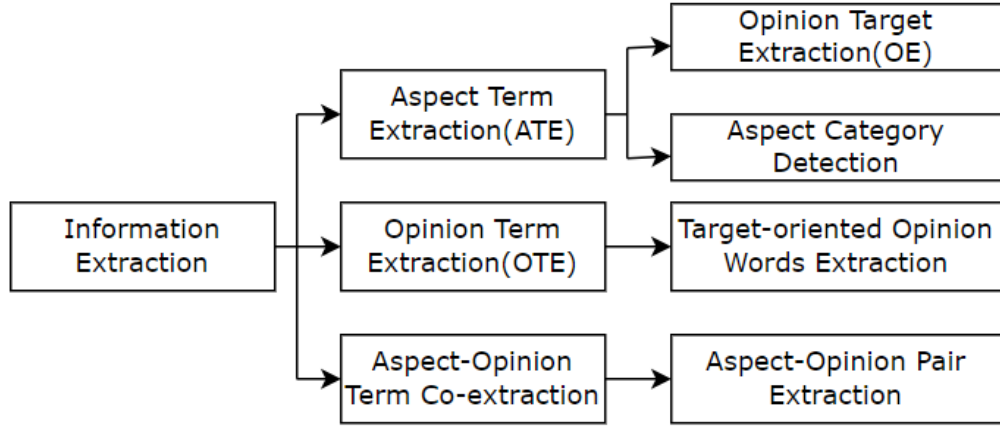
Figure 1.3: IE task classification.

In IE task, most of the research is concentrated on Aspect term extraction(ATE) task while very few on OTE task or we can say TOWE task.

Various benchmark dataset for ABSA are available in various languages and on several domains. For high-resource language, ABSA seems easy as compared to low-resource one because of enough datasets and required tools. ABSA has been extended beyond english and explored in other languages such as Hindi (13), (14) , Indonesian (15), (16) , Urdu (17) , Arabic (18), (19) etc. This paper focuses on conducting ABSA subtasks on a Nepali language text. Previous studies by (20), (1) have also explored ABSA in the Nepali language. Although started but their (1) research lag to deal with the task of determining which opinion is targeted to which target that is target oriented opinion term extraction task and that task is dealt here. Also, the classification task here has outperformed than that of author.

The tasks, naming, dataset(domain and language) differs from paper to paper, for the convenience to readers the tasks and naming will be clearly explained here.

A) Dataset: NepSA dataset by author (1) is taken as a baseline dataset along with some addition. The format of dataset as provided by original author is changed and some new data are added in the existing one. The dataset consists of 5066 sentences in which 4035 sentences are taken from (1) and 1031 are added. The source of added dataset is youtube comments from the videos under 'News and Politics' category with maximum views of the year 2023 AD . Target entities, aspect categories, opinion terms and polarity of aspect categories are annotated.

B) Naming:

   (a) Opinion Term: The words or expressions present in a text that convey sentiment or opinion are opinion terms. Opinion term usually consists of adjectives or adverbs.

   (b) Opinion target: opinion target refers to the specific entity, aspect, or topic that is the subject of an expressed opinion or sentiment in a given text. It is also called as Target term.

   (c) Target categories: The dataset consists of four target categories: PERSON, LOCATION, ORGANIZATION and MISCELLANEOUS.

   (d) Aspect categories: Aspect categories are inferred through the opinion terms present in the text(9). The dataset consists of four predefined aspect categories namely GENERAL, FEEDBACK, PROFANITY and VIOLENCE.

C) Tasks:

This thesis work will be covering following subtasks under ABSA namely

   (a) Aspect Category Detection(ACD): As mentioned in (9) Aspect categories are coarser than aspect terms and can be inferred through the adjectives present in the text. Here also they are inferred from the adjectives present in the text.

   (b) Target term extraction: Target terms are nouns or noun phrases.

(c) Opinion term extraction: Opinion terms are the words or phrases in a text that depicts the sentiment the text wants to say.

(d) Target oriented opinion word extraction task:

Opinion terms are extracted along with the target terms(which are normally noun or noun phrases) towards which the opinion is opinionated to are extracted here and in case of multiple targets present in a text the nearest one(nearest to opinion term) is assumed to be the actual target(21). Typically, this task is approached as a sequence labeling problem utilizing the BIO tagging scheme. The tags consist of a prefix : B, I, O and a label. B (Beginning) indicates the beginning of an entity. I (Inside) indicates that the token is inside an entity. O (Outside) indicates that the token is outside any entity. Following diagram shows an example of applying the BIO tagging scheme for nepali text.

Sentence:

किन होला नेपाल फेल भएको कहाँ निर पो होला नेता चुके का ?

True Labels for target नेपाल:

किन/O होला/O नेपाल/O फेल/B-OT भएको/O कहाँ/O निर/O पो/O होला/O नेता/O चुके/O का/O

True Labels for target नेता:

किन/O होला/O नेपाल/O फेल/O भएको/O कहाँ/O निर/O पो/O होला/O नेता/O चुके/B-OT का/O

Extraction Results:

{(फेल, नेपाल)
(चुके, नेता)}

Figure 1.4: Identifying target terms and opinion terms in a sentence.

(e) Aspect category Polarity(ACP) detection:

The main goal of any type of SA is to find the polarity of the text. In this task the polarity for aspect categories will be detected. Here, polarity for aspect categories will be treated as binary classification task. Two polarities are given, 0 for positive and 1 for negative.

7

Let us understand ABSA from the example.
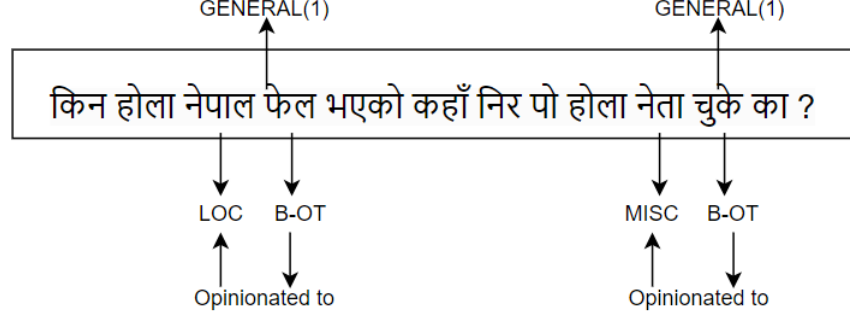
Figure 1.5: Fine-grained annotation sample

The english translation for above sentence is: *"Why is Nepal declining ? Where have the politicians lapsed?"*

In above sentence, B-OT is denotion for opinion term, LOC for location and MISC for miscellaneous. LOC and MISC are target categories. The aspect category for both opinion terms in above example is GENERAL and aspect category polarity is 1 that is negative polarity. The first opinion is opinionated to नेपाल(LOC) target, the next opinion is opinionated towards नेता(MISC) target. The task would be:

1. At first, target term would be extracted which are नेपाल (English: *Nepal*) and नेता (English: *Politician*) in above example. Then opinion terms are extracted. Then opinion term opinionated to the target is find out. That is, for the first opinion term its target नेपाल is obtained and for the second opinion term its corresponding target नेता is obtained. The distance is taken as the metrics while finding target opinion pair that is the target nearest to the opinion term is assumed to be the target towards which opinion is opinionated to.

2. Aspect category detection(ACD) task would find that the category associated with first opinion term is GENERAL and aspect associated with second opinion term is GENERAL.

3. Aspect category polarity(ACP) task would find that the polarity for first category(GENERAL) is 1 and then for second category(GENERAL) is 1.

For high-resource language like english, many datasets are available via SE (9), (10), (11) shared task. Moderate number of studies can be found on Nepali sentiment

analysis, but very few on Nepali ABSA. Papers(1), (20) performed ABSA task in nepali dataset till Feb 2024. This is due to the reason that Nepali ABSA is challenging task than Nepali SA and more challenging than ABSA in high-resource language. SA is evolving. Due to the NLP origins SA tasks seems challenging but over the last few years much progress has been made due to the high demand for SA. Companies and customers both needs SA, companies are curious about how their customers feel about the goods and services and consumers also want to hear what other people think before making purchases. For both businesses and individuals they need to concentrate on the specific aspect where their customers are complaining or making suggestions to better their product or service, ABSA will help in analyzing aspects in texts like reviews, comments, etc. This saves money and time both. ABSA on reviews, comments regarding politics, political figures is also popular and has been an effective way for making strategies during election. Here, ABSA is carried out on nepali text extracted from youtube comments on most popular youtube videos with highest views of category 'nepali news and politics'. The process of extracting aspects, targets from the textual data as well as the identification of associated sentiment words is difficult in ABSA.

Previously the target oriented opinion term extraction task used rule based approaches (22), (23), (24) and supervised based methods. (25) proposed a novel ML approach which is created using the lexicalized HMM framework, (26) used several machine learning(ML) methods like Logistic Regression,NB, SVM and KNN for opinion word extraction. In 2015, **Liu et al.** (27) suggested a general class of discriminative model using word embedding and Recurrent neural network(RNN) for opinion term extraction task. (28) also used deep learning to solve TOWE task. Due to the limitation of inner structure of RNN it cannot learn long distance dependencies in text and in the preprocessing step a dependency parser is needed. **Wang et al.** (29) in 2017 used attention mechanism for aspect term and opinion term extraction task.

The recent innovation of various deep learning techniques and transformers has removed the effort needed for feature engineering and has achieved better results in several NLP tasks. More detail regarding the methods applied is discussed in literature review part.

1.2 Problem definition

There is huge amount of data created online due to rapid growth in the usage of internet and its technologies. The internet is the home for different social sites, movie providers and various other platforms where people jot down their emotions, people spill out their anger, sorrow, happiness and many more human emotions in the form of text and many other forms as well. Since ages, Other peoples opinion, thinking, perception matters a lot to us while doing anything whether it be watching some movie in netflix or buying household goods from online store or say while buying some sorts of lentils from nearby stores also we usually took advice from others that which company's packed lentils is good. So, making decisions based on others opinion is not a new thing to people. The traditional SA techniques mostly treats the review or text as a whole that is either at document level or sentence level. The opinion term extraction, the target towards which the opinion is targeted to and sentiment towards particular aspect of a text review was not considered in document level and sentence level SA. This creates incomplete analysis of the sentiment that is present in text. Due to this some fine grained analysis is needed which can provide complete analysis about what opinion term, what target, which aspect category and what sentiment polarity the review, opinion or text has. There are several tasks within ABSA, under the subtask Target oriented opinion term extraction, it aims to find the opinion term and the target(aspect) towards which that opinion is opinionated to. Under subtask Aspect Category Detection(ACD) the aspect category is detected for the opinion term that is multi-class classification task and under Aspect category polarity(ACP) subtask the polarity of aspect category is detected and for two polarity labels it will be treated as binary classification task. In conclusion we can say that it gives more detailed result in sentiment analysis.

The main challenge in ABSA is that the sentiment expression frequently depends heavily on the particular topic under discussion. So, creating an effective ABSA system requires correctly recognizing and extracting the appropriate opinions, targets, aspect category from customer reviews as well as appropriately assigning sentiment polarity to each aspect category. Several researches on ABSA and

TABSA has been done and most of the research have been done in english dataset but very few study have been done in low resource language like nepali. Very few papers are available for nepali text based ABSA and TABSA. This research can contribute to the ABSA task in nepali language.

A thesis on ABSA may examine various NLP techniques, ML techniques, DNN techniques, word embedding techniques and meaningful insights will be obtained. Along with exploring strategies to increase the precision and effectiveness of ABSA systems, the thesis may also look into the efficacy of various methodologies for various domains and datasets.

Sentiment analysis task is focussed on extracting sentiment of whole sentence but this become insensible when particular sentence carry more than one opinion terms and more targets for which opinions are targeted to . This is where more fine-grained sentiment analysis model is needed which can extract the relevant opinion terms and their corresponding targets and sentiment polarity. This feature is posed by subtasks under aspect based sentiment analysis which will be covered in this thesis work.

1.3 Objectives

The main objectives of this project are:

1. To extend the Nepali dataset for use in various subtasks of aspect based sentiment analysis.

2. To identify opinion term, target term towards which the opinion is targeted to, aspect category and aspect category polarity of nepali text.

1.4 Scope of the thesis work

The major scope of this research work are as follows:-

1. The proposed model will be exploring state of the art to determine opinion, target, aspect category and polarity present in the review.

2. The dataset will be Nepali dataset

1.5 Thesis contribution

The major contributions of this thesis are as follows:-

1. Dataset addition in existing baseline dataset for ABSA.

2. Information Extraction tasks like Target term extraction, opinion term extraction, Target oriented Opinion term extraction are performed with acceptable results.

3. The classification tasks, Aspect category detection and Aspect category polarity detection for Nepali text have achieved state-of-the-art performance.

1.6 Organization of the thesis work

The thesis proposal is organized as:

Chapter 2: Available literature about various ABSA subtasks

Chapter 3: More theoretical background related to this study are briefly described in this chapter.

Chapter 4: The methodology used for the research of ABSA is discussed in this chapter.

Chapter 5: The Result and discussion portion.

Chapter 6: Conclusion.

CHAPTER 2

LITERATURE REVIEW

A significant amount of research has been done for aspect based sentiment analysis and its various subtasks.

2.1 Sentiment Analysis

pang et al. (2) in 2008 defines Sentiment analysis (SA) as the process of evaluating human emotions, opinions, reviews expressed in text to detect the writer's mental outlook towards a particular event, topic, product, service, etc. The phrase sentiment analysis was used here probably for the first time, although the phrase opinion mining was used previously but the author finds most of the similarity on both these phrases.

Agarwal et al. (30) in 2015 defines SA as a method that extracts, transforms, and interprets opinions from texts using Natural Language Processing (NLP) inorder to categorize them as positive, negative, or neutral sentiment .

Mantyla et al. (31) in 2018 said that nearly 7000 papers on sentiment analysis(SA) topic have been published and also mentioned that SA is one of the fields of research that is expanding very fast, with 99 percent of the articles published after 2004.

Aqlan et al. (32) in 2019 defines sentiment analysis as a process in which there is thorough examination of data that is maintained online inorder to identify and classify the opinions conveyed in a passage of text.

2.2 Aspect Based Sentiment Analysis (ABSA)

ABSA is a form of SA that gives fine-grained, detailed form of analysis of sentiment or opinion. That is apart from saying whether a piece of text depicts positive, negative sentiment, ABSA tells that in a sentence there are different opinions or sentiments expressed towards different targets and this task of finding opinion word/phrase and the target towards which opinion is expressed is known as Target-oriented opinion word extraction (TOWE) task in ABSA.

Hu et al. (22) in 2004 first introduced the concept of aspect based sentiment analysis, author named it as feature based sentiment analysis where feature resembles aspect conceptually. Paper talks about three tasks and also proposes a number of novel methods to do the tasks.

The dataset for ABSA determines the task to be followed. If there are more target terms in the available dataset then it is good to go for TABSA task but if very less or no target terms are there then TABSA is not a choice of perfect. Therefore data determines what to do. Generally speaking ABSA is a finegrained SA. In research like (9), (10), (11) aspect term extraction task extracts aspects which are noun or noun phrases whereas author in (12) further subtasked the aspect term extraction task into opinion target extraction(OE) and aspect category detection(ACD) tasks where OE task aims in extracting aspect terms and ACD task extracts aspect category.

Liu et al.(33) in 2012 defines ABSA as opinion mining from text on particular entities and their aspects, can offer both businesses and consumers insightful information.

Zhang et al. (34) in 2012 explains aspect extraction as extracting attributes of an object about which opinions are told, is one of the crucial subtask in ABSA.

pontiki et al. (9) in 2014 introduced annotated restaurant and laptop reviews datasets for aspect based sentiment analysis under task 4 in SemEval-2014 . Likewise, the dataset increases again in Sem-Eval-2015 (10). In Sem-Eval-2016 datasets became available in multi languages (11).

Huang et al.(35) in 2018 mentions that mostly the reviews on various domains not only corresponds to single aspect but the single sentence consists of seperate perception towards more than one aspects. In order to classify sentiment expressed toward certain aspects given in sentences, aspect based sentiment analysis is used.

Ilmania et al. (36) in 2018 has applied Word2vec and GRU for aspect detection and has presented 2 ways for Sentiment classification: one through Bi-GRU and another one is by using CNN.

wang et al. (37) in 2019 mentioned that at the sentence or document level, previous studies on sentiment classification have achieved outstanding results.But such classification is independent of aspects.

Zhang et al. (38) in 2022 surveyed on ABSA and its tasks, methods and challenges has presented various subtasks within ABSA, their methods using various machine learning approaches, deep learning approaches and attention based techniques and also various challenges are presented there. Four terms has been introduced in this paper: aspect term, aspect category, opinion term and sentiment polarity.

While **Zhang et al.** (39) in 2020 has focused on aspect term sentiment analysis(ATSA) step and had constructed the ATSA model using Bi-LSTM Network. ABSA has various subtasks within it, some authors has explored towards aspect term sentiment analysis and some has focused towards aspect category sentiment analysis but very few has done both.

Opinion mining needs several pre-processing steps in order to transform the text in natural language into a form that a machine can interpret. One of the most useful step is POS tagging. A word's POS tagging tell us the morphological affixes it can take. POS also assist in sentiment analysis application by helping select out nouns or other important words from a document (40).

While talking about methods, in earlier days, ML methods were focussed which focusses on handcrafted features like lexicon inorder to train sentiment classifiers(41). Although such methods seems effective but they depend on the efficiency of the handcrafted features. Nowdays most papers mentions the use of deep learning techniques on absa tasks because of its several benefit over ML techniques like no need of feature engineering etc.

liu et al. (42) in 2020 mentioned that various deep learning techniques can be applied to do aspect-based sentiment analysis.

Islam et al. in 2023 applied deep learning framework Deep-ABSA for ABSA from Bangle texts. The authors mentioned there that opinion mining is important from political point of view. The popularity, defame, name of political figures can be known from the peoples opinion in social media.

2.3 Targeted Aspect Based Sentiment Analysis

Jiang et al. (43) in 2011 demonstrated the importance of target entities for a sentiment classification task.

Saeidi et al. (44) in 2016 introduced Sentihood dataset for targeted aspect based sentiment analysis in urban neighbourhoods domain. For fine-grained opinion mining, social media platform has been used for the first time here. There are 5215 sentences, 3862 of them contains a single location and 1353 of them contains multiple locations. F1 measure is used for aspect detection task while accuracy is used for sentiment classification task. Bi-LSTM is employed inorder to learn a classifier for each of the aspects.

Ma et al. in 2018 proposed a neural architecture for the task of TABSA. Author has used sentihood and semeval-2015 dataset for the task of TABSA. Author concludes by mentioning to inquire into how commonsense knowledge fits into modeling of the relationship between targets.

Wu et al. in 2020 proposed 2 variants of Context-Guided BERT for ABSA and TABSA. SemEval-2014 (task 4) and SentiHood (T)ABSA datasets are used.

2.4 TABSA in Nepali Language

Singh et al. in 2020 introduced benchmark dataset for nepali TABSA named NEPSA dataset and evaluate by using various ML models (1). In initial days sentiment analysis was done in sentence based or document based which is still done but thesedays the focus has shifted more towards aspect based sentiment analysis and targeted sentiment analysis. That is more fine-grained variety is todays need in sentiment analysis field.

Target oriented opinion word extraction(TOWE) task is not mentioned by the author (1) so, the main concern of this thesis work would be to focus on TOWE task along with other tasks of ABSA for nepali dataset.

2.5 Target oriented opinion Term Extraction

Rule based methods for TOWE task do not need labelled data instead they require rules that are designed manually and this is hard to apply on reviews which are complex. Some works on TOWE by utilizing rule based methods are :

Hu and liu et al (22) in 2004 suggested using synonym/antonym from WordNet to find the opinion terms and association rule mining based on Apriori algorithm

inorder to find noun phrases and nouns as the aspect term under aspect term extraction task.

Bafna et al. (23) in 2013 continued the work of **Hu et al.**. To find aspect terms they first used association rules then excluded the terms which do not belong to aspect term with some probabilistic way. Then they extract opinion terms closest to aspect terms.

Htay et al. (24) in 2013 performed the extraction of opinion word using the annotated customer reviews dataset from 5 products. Author used POS tagging technique and consider the adjective/adverb as opinion word.

For supervised based methods, the TOWE task is treated as a sequence labeling problem. **Jin and Srihari et al.** (25) in 2009 proposed a novel ML approach which is created using the lexicalized HMM framework. The author at first with several words made a lexicon and their corresponding (POS)part of speech, then used lexicon made by them inorder to label manually the aspect term and opinion terms present in a text. At last used HMM for training of sequence tagging process. The author performed the task in english dataset

Chetviorkin et al. (45) in 2011 proposed a methodology that uses ML based techniques to automatically extract opinion words including adjectives and non-adjectives for any domain. This paper mainly focuses on creating Russian sentiment lexicon that is domain-specific.

Shamshurin et al. (26) in 2012 used several machine learning(ML) methods like Logistic Regression,NB, SVM and KNN for opinion word extraction.

Deep learning is being widely used in NLP very strongly.

Liu et al. (27) in 2015 suggested a general class of discriminative model using word embedding and Recurrent neural network(RNN) for opinion term extraction task. The author used review datasets from SemEval 2014 ABSA task-task 4, (9) Restaurant and Laptop domain datasets.

Veyseh et al. (28) in 2020 mentioned Target Opinion Word Extraction(TOWE) as a subtask of ABSA where opinion word for a given aspect term is find out. Author formulated TOWE problem as a sequence labelling task. Author used BIO tagging schema to encode the label. Author used deep learning paradigm to solve the task.

Due to the limitation of inner structure of RNN it cannot learn long distance dependencies in text and in the preprocessing step a dependency parser is needed.

Wang et al. (29) in 2017 used attention mechanism for aspect term and opinion term extraction task.The experiment showed that attention mechanism can improve the extraction performance effectively. Author used SemEval-2014 task 4 subtask 1(9) and SemEval-2015 task 12 subtask 1(10), it means author performed the experiment in english dataset.

Fernando et al. (46) in 2019 used attention mechanism and double embeddings for aspect term and opinion term extraction task for hotel reviews on indonesian language.

Winatmoko et al. (47) in 2019 used transfer learning using pre-trained BERT for aspect and opinion term extraction. Author experimented with dataset in bahasa Indonesia. The author discussed the advantage regarding the number of epochs required. Pretrained models at maximum needed 4 epochs inorder to train while the BiLSTM model was trained for 200 epochs.

Following table is the summary of literature review on ABSA related task.

Authors	Title	Year Pub-lished	Method Used	Data set	Evaluation Metrics	Results
Ilmania, Arfinda and Cahyawi-jaya and others.	Aspect Detection and Sentiment Classification using Deep Neural Network for Indonesian Aspect-Based Sentiment Analysis	2018	For aspect detection:Word2vec and GRU and for Sentiment classification 2 ways: one through Bi-GRU another one is CNN	9800 re-views	Precision, Recall, F-Measure	Precision= 0.8936, Recall = 0.8775, F-score = 0.8855
Ishaq, Adnan and As-ghar et al.	Aspect-based sentiment analysis using a hybridized approach based on CNN and GA	2020	Mining seman-tic features, Word2vec, CNN, hyper-parameter of CNN are tuned with GA	Hotel re-views, auto-mo-biles re-views, movie re-views	Accuracy, precision, recall, f1 measure rate	Precision=91.6%, Accu-racy=93.4%, Recall=92.2%, f1 measure=89.23%
Cahyaning-tyas, Siwi et al.	Deep learning for aspect-based sentiment analysis on Indonesian hotels reviews	2020	For aspect classification : LSTM, for sentiment classification : various deep learning meth-ods	5,387 re-views	Accuracy	Accuracy=92.6% .
Zhang, Bowen and Li et al.	Knowledge guided capsule atten-tion network for aspect-based sentiment analysis	2020	For ATSA: Bi-LSTM Network and a Capsule Atten-tion Network	SE 2014, Rest15, Rest16	Accuracy and F1score	Accuracy=82.05%, F1score=74.04%
Kothalawa-la, Malki et al.	Aspect-based sentiment analysis on hair care product reviews	2020	SVM and Multinomial NB algorithms	200 re-views on sham-poo	Accuracy	Accuracy of 85%

Authors	Title	Year Published	Method Used	Dataset	Evaluation Metrics	Results
Ye, Xingxin and Xu et al.	ALBERTC-CNN Based Aspect Level Sentiment Analysis	2021	ALBERTC-CNN	2 datasets of the SemEval-2014 open task, the laptop and restaurant datasets	Accuracy, RMacro, F1Macro	Accuracy = 80.56%, Rmacro = 79.13%, F1Macro = 76.97%
Abdelgwad et al.	Arabic aspect sentiment polarity classification using BERT	2022	BERT-Linear-pair	Arabic hotel, reviews Arabic news	Accuracy	Accuracy for Arabic hotel reviews = 89.51%, Arabic news = 85.73%
M. Melih et al.	A Dataset and BERT-based Models for Targeted Sentiment Analysis on Turkish Texts	2022	BERT-based models	Turkish Twitter dataset	F1-score	67% F1-score

Table 2.1: Summary of Literature Review for ABSA

Following table is the summary of literature review on Target oriented opinion term extraction (TOWE) task.

Authors	Title	Year Published	Method Used	Data set	Evaluation Metrics	Results
Wenya Wang, Sinno Jialin Pan and others.	Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis	2016	joint model(RNN and CRF)	SE Challenge 2014 task 4(Res and Lap)	F1scores	Res-aspect= 84.93 Res-opinion=84.11 Lap-aspect=78.42 Lap-opinion=79.44

Authors	Title	Year Published	Method Used	Data set	Evaluation Metrics	Results
Wenya Wang, Sinno Jialin Pan and others.	Coupled Multi-Layer Attentions for Co-Extraction of Aspect and Opinion Terms	2017	coupled multi-layer attentions	SE 2014-Res and Lap, SemEval 2015-Res	F1scores	For SemEval 2015-Res, for aspect = 70.73 opinion = 73.68
Zhifang Fan, Zhen Wu, Xin-Yu Dai and others.	Target-oriented Opinion Words Extraction with Target-fused Neural Sequence Labeling	2019	Neural Network	14res, 14lap-SE 2014, 15res-SE 2015, 16res-SE 2016.	Precision, Recall, F-Measure	14lapSE14- Precision= 73.24 , Recall = 69.63 , F-score = 71.35
He Zhao, Longtao Huang and others.	SpanMlt: A Span-based Multi-Task Learning Framework for Pairwise Aspect and Opinion Terms Extraction	2020	BiLSTM encoder and BERT encoder	SE 2014,15,16-RES and LAP of all	F1scores	15 RES, aspect term= 81.76, opinion term = 78.91, pair = 64.68
Shaowei Chen, Jie Liu, Yu Wang and others.	Synchronous Double-channel Recurrent Network for Aspect-Opinion Pair Extraction	2020	Synchronous Double-channel Recurrent Network (SDRN) and BERT as the encoding layer	Semeval 2014,15-RES and LAP, MPQA version 2.0 corpus and JDPA	F1scores	For aspect-opinion pair extraction, For 15-Res- = 70.94

Table 2.2: Summary of Literature Review for TOWE task

CHAPTER 3

THEORETICAL BACKGROUND

3.1 Aspect based sentiment analysis(ABSA) using deep Learning and Transformers

Sentiment refers to the positivity or negativity that is expressed in text. While coming to ABSA, it refers to sentiment of aspects, targets etc. Sentiment analysis provides a way to evaluate text in some language to determine if the expression is favourable or unfavourable and in some case neutral also and in some case the text might depict both positive and negative sentiment. Paper by author Pang et al. was one of the first research paper in sentiment analysis (SA) (48) . They classified the reviews into two classes: positive and negative by using ML methods: Naive Bayes, maximum entropy and SVM.

Deep learning methods are used mostly in various research thesedays because it can handle the complexity of NLP better than ML algorithms. Traditional ML algorithms rely mostly on hand-crafted features which can be limited in their ability to capture the full range of linguistic nuances and contexts that are present in natural language. Deep Learning shifted the feature engineering task to model rather than human which used to involve in other classical machine learning. Various deep learning techniques are CNN, RNN, LSTM, Bi-LSTM, GRU, GAN, MLPs etc.

3.2 Target oriented opinion word extraction task

TOWE is a fine grained task of target-oriented SA (6) that seeks to extract opinion terms and the opinion target(or aspect) from text. Let us take an example, "The food is nice but the service is bad", TOWE aims to extract opinion terms "nice" and "bad" and their target terms "food" and "service" respectively. TOWE task combines aspect term(target) extraction task and opinion term extraction task. Typically, the TOWE task is approached as a sequence labeling problem utilizing the BIO tagging scheme.

3.3   Bidirectional LSTM model

A type of RNN called a bidirectional LSTM (BiLSTM) model consists of two LSTMs, one of which takes input in a forward direction and the other in a backward manner. As a result, it extracts additional information from text by processing the sequence in two directions and understanding what text comes after and before a word in a sentence.(49)

1. LSTM

   LSTM stands for Long Short Term Memory. LSTM is a type of RNN which is capable of learning the long-term dependencies mostly in sequential data. LSTMs are designed to avoid the vanishing gradient problem that can occur in traditional RNNs when the gradients become too small to update the weights effectively. LSTM has feedback connections, therefore it can process the complete sequence of data because of this. In Bi-LSTM, 2 LSTMs are used.

   LSTM consists of 3 gates namely forget gate, input gate and output gate. Equations for them is shown below:

   (a) Equations of Forget Gate

   $$f_t = \sigma \, (\, W_f \, x_t + U_f \, h_{t-1} + b_f \,)$$
   $$c_t = c_{t-1} \circ f_t$$

   (b) Equations of Input gate

   $$i_t = \sigma \, (\, W_i \, x_t + U_i \, h_{t-1} + b_i \,)$$
   $$a_t = \tanh \, (\, W_a \, x_t + U_a \, h_{t-1} + b_a \,)$$
   $$c_t = c_{t-1} \circ f_t + i_t \circ a_t$$

   (c) Equations of Output gate

   $$o_t = \sigma \, (\, W_o \, x_t + U_o \, h_{t-1} + b_o \,)$$
   $$h_t = \tanh(c_t) \circ o_t$$

3.4   Transformers in NLP

Transformers were introduced by **Vaswani et al. (50)** which relies solely on self-attention mechanisms without recurrent or convolutional layers which abandons sequential processing and enables parallelization of training, leading to faster

convergence. Actually transformer consists of many encoders and many decoders within it.

Transformer

INPUT ⟶ | Encoder | Decoder | ⟶ OUTPUT
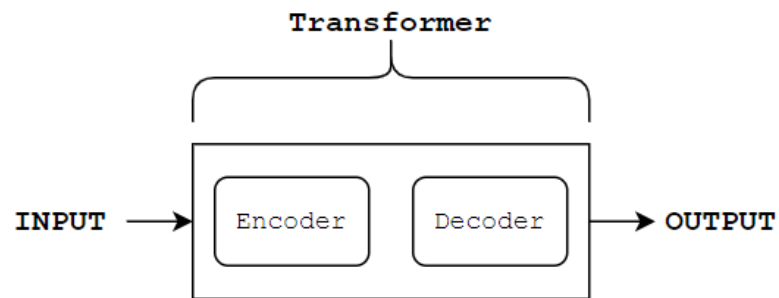
Figure 3.1: Transformer as a black box

OUTPUT
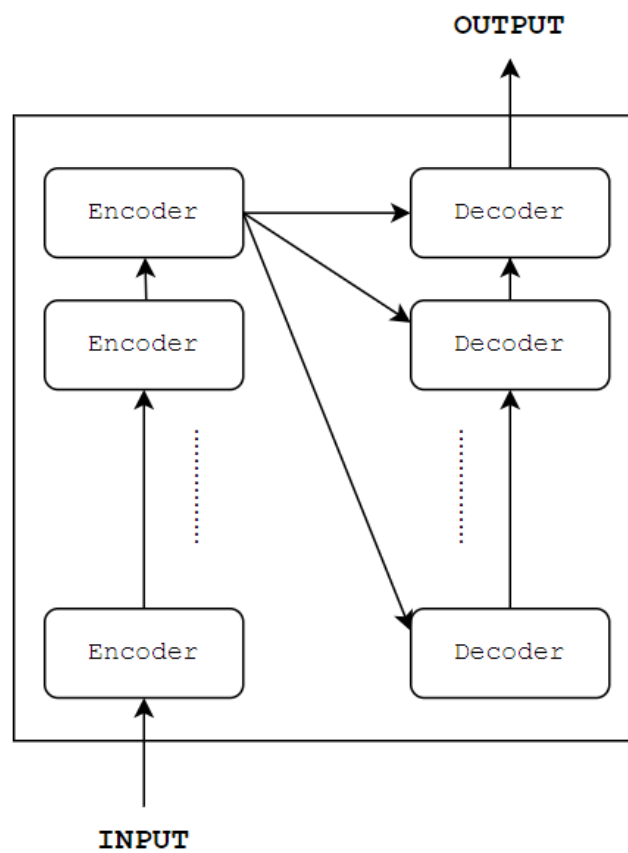
Figure 3.2: Inside transformer

There are four features in transformers:

1. Word Embedding

2. Positional Encoding

3. Self-Attention

4. Residual Connections

24

Four features allow the transformers to encode word into numbers, encode the position of the words, encode the relationships among words and relatively and easily and quickly train in parallel.

The scaled dot-product attention is a crucial component of the Transformer, defined by the following equation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where, Q, K and V are query, key, and value matrices and d_k is the dimensionality of the key vectors.

3.5   Bidirectional Encoder Representations from Transformers(BERT)

BERT(Bidirectional Encoder Representations from Transformers) is based on transformers which is a deep learning model in which every output element is connected to every input element, and the weightings between them are dynamically calculated based upon their connection. BERT is a tool that understands language more than any other tool. BERT is a revolutionary concept in natural language processing (NLP). BERT takes into account both the left and right context of each word in a sentence because at its core, it is meant to comprehend and process language in a bidirectional manner.
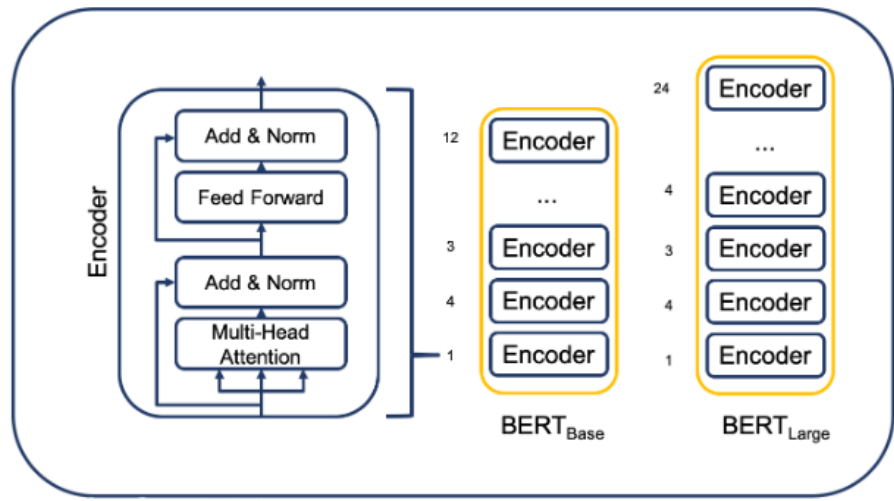
Figure 3.3: The architecture of BERT$_{base}$ and BERT$_{large}$

The foundation of BERT is a stack of encoder layers. The number of encoder layers is where $BERT_{base}$ and $BERT_{large}$ diverge. In the $BERT_{large}$ model there are 24 layers of encoders layered on top of one another, compared to 12 layers in the $BERT_{base}$ model. BERT model can also be used for NER(Named Entity Recognition) task.
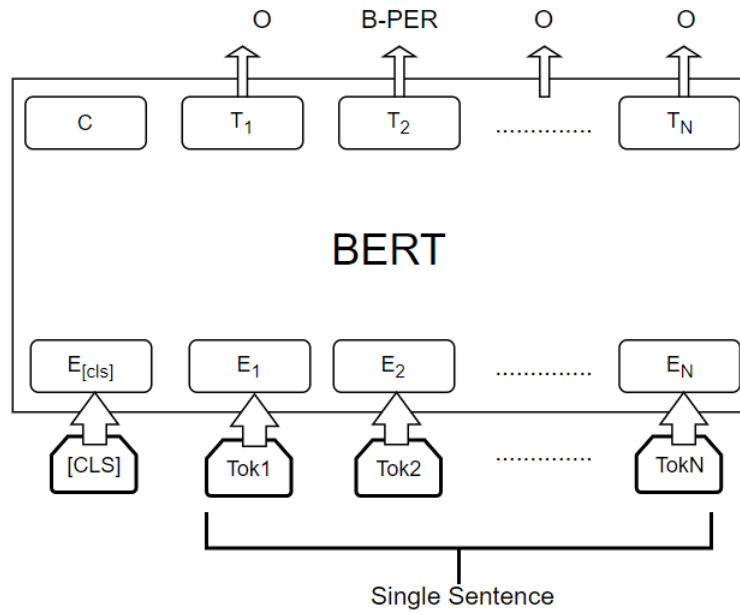
Figure 3.4: BERT for NER

CHAPTER 4

METHODOLOGY

4.1 Overview

Sentiment analysis is a task of finding opinion, sentiment, polarity of a text. There are various varieties of sentiment analysis. Sentence-based SA considers single polarity for whole sentence. Unlike sentence-based SA, in aspect based SA(another variety of SA), the opinion/ sentiment finding process gives some extra details like Target extraction task, opinion extraction task, target oriented opinion word extraction task, aspect category detection and aspect category polarity etc. Targeted sentimet analysis is another variety of ABSA. (51) .

Aspect based sentiment analysis consists of several subtasks within it. Some of the major papers (9), (10), (11) categorised tasks as: aspect term extraction, aspect category detection, the TOWE task is also seen as a vital task in ABSA. The dataset used in this research is NepSA dataset prepared by **Singh et al.** (1). About 1031 dataset was added following the same annotation guideline and in total there will be 5066 sentences with required annotations. The dataset consists of fields like opinion term which reflects the opinion, aspect categories which are inferred from the opinion terms and in the dataset there are four predefined aspect categories The dataset consists of target term which is the target towards which the opinion is targeted, target category which is inferred from the type of target, there are four predefined target categories The dataset also consists of labelling for polarity for the aspect categories or we can also say for corresponding opinion terms. 0 is for positive opinion and 1 is labelled for negative opinion. I would like to appreciate the author **Singh et al.**(1) for the great initiation of ABSA in nepali language and the availability of dataset and the annotation guideline so that new data can be added. The main contribution of this thesis work would be TOWE task, ACD task and ACP task, all of the three are performed here along with improvement in classification task. Existing paper of ABSA in nepali dataset didnot talk about TOWE task so, this would be the new thing here. Dataset has

been modified for NER task.

4.2   Block diagram of the model
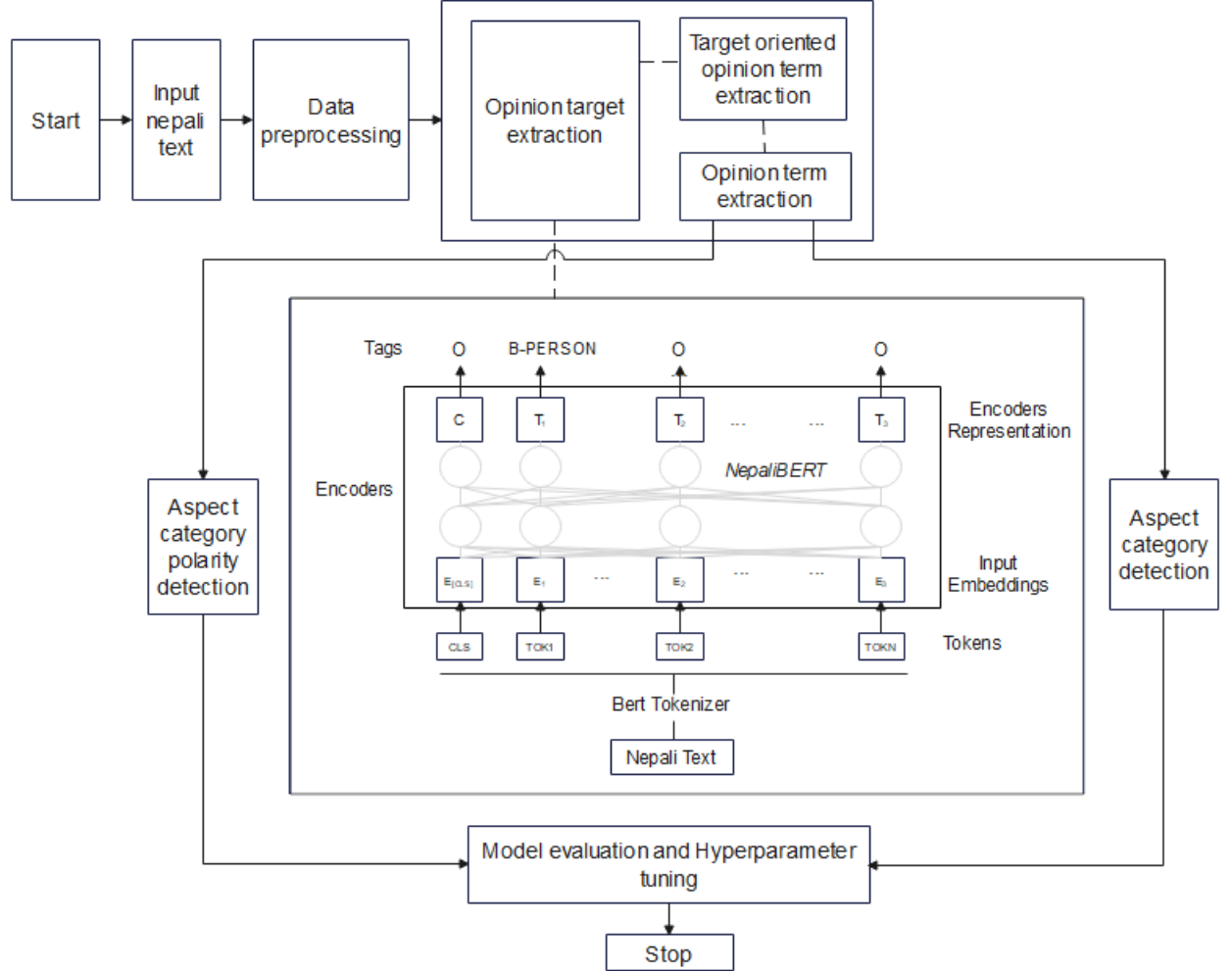
Figure 4.1: Block diagram

4.3   Steps Involved

1. **Dataset selection:**

   NepSA dataset is taken as a baseline dataset for this thesis work. 1031 datasets were added in baseline dataset following the same annotation guideline followed while preparing NepSA dataset. So in total 5066 sentences and needed labelling is there. As ABSA in low resource language is rarely popular so there is very less chance to get data in low resource language like

nepali. Natural language processing is complex but various deep learning and transformers variations tries to simplify it. Several researches are done in ABSA for english dataset but very few on nepali. So, This thesis work tries to contribute in ABSA tasks using nepali dataset.

2. **Data Preprocessing:**

Data pre-processing is performed after acquiring the dataset. In data preprocessing following steps are performed:

(a) Data Loading and Cleaning: The dataset is loaded from a CSV file using Pandas.

(b) Feature Engineering: Columns 'text' and 'opinion' are concatenated to create a new column 'Combined text' for concatenated embedding.

(c) Text Tokenization and Vectorization: Tokenization is the process of dividing text into smaller parts. Words, sentences, paragraphs, etc. can be used as tokens. Text can be divided into meaningful chunks that NLP models can parse with ease with the aid of tokenization. The NLTK library is used for word tokenization while implementing SVM model. While implementing BERT, bert tokenizer is used. The TfidfVectorizer from Scikit-learn is employed to convert text data into numerical vectors. The vectorizer is fitted on the 'Combined text' column.

3. **Target oriented opinion word extraction task**

Initially target term is extracted, for target term extraction, as this task is extraction task so, Named Entity Recognition (NER) technique is used for this. Target terms are normally nouns or noun phrases. The task is treated as a sequence labelling problem utilizing the BIO tagging scheme. The dataset is made in the format that each token of the sentence is given some label. BERT model is used for NER task as the recent trend on NLP and many papers on this task seems to apply BERT as explained in literature above. Then opinion term is extracted using the BERT model. Then, opinion word and the targets towards which the opinion is opinionated to is extracted, in

most cases the opinion terms are targeted to nearest target terms so this assumption is utilized here and when more than one opinion in a single sentence then the nearest target term is assumed to be targeted by the opinion term.

4. **Aspect Category Detection task**

Aspect Category Detection(ACD) is one of the subtask of ABSA. The responsibility of ACD is to detect which aspect categories come in a text or review. There have been several promising approaches to aspect category detection in recent years. The task of classifying an aspect term into one of the predetermined categories is known as aspect category detection, or ACD for short. In the situation when aspect term is not seen visully on sentence then aspect categories are inferred from the opinion term (9). Aspect Category Detection (ACD) is a fundamental task in sentiment analysis, which aims to identify the aspect categories mentioned in a given review sentence from a predefined aspect category set.

5. **Aspect Category Polarity**

The polarity is provided to the aspect categories on the basis of the sense it gives, here 2 classes are there, first one is positive which is '0' here and next is negative which is '1' here. This task is treated as binary classification task. Aspect Category Polarity task is also done for two cases, first is concatenation of text with opinion term and another experiment is conducted without concatenation. Several experiments on various models is performed. Due to more feature added, concatenation shows higher measures while evaluating.

6. **Model evaluation and hyperparameter tuning**

Finally Model evaluation and hyperparameter tuning is done. The appropriate optimizer, loss function appropriate as per the task will be used. Various evaluation metrices will be used as appropriate for the tasks. For classification task, confusion matrix is generated and precision, recall, F1-score, accuracy of

the model will be measured while for information extraction task, ROUGE-L score will be used.

4.4   Experimental Design

4.4.1   Dataset description

Dataset Name	Nep-SA dataset plus extra added
Aspect Categories	Feedback, General, Profanity, Violence
Target Entities/Categories	Person, Organization, Location, Miscellaneous
polarity	0 and 1
Total 0 polarity	2563
Total 1 polarity	2503
Total sentences	5066
Total Words	84651
Total B-OT	5295
Total I-OT	4875
Maximum words in a sentence	93
Average words in sentence	17

Table 4.1: Overall description of our dataset used for experiment

Aspect Categories	Count		Total
	0	1	
General	1484	2063	3547
Feedback	519	130	649
Profanity	430	119	549
Violence	130	191	321
Total	2563	2503	5066

Table 4.2: Total count of sentences

Number of words	Count of opinion terms
One word	2071
Two words	1862
Three words	792
Four words	242
Five words	62
Six words	22
Seven words	10
Eight words	2
Nine words	2
Ten words	1

Table 4.3: Total count of number of words in opinion terms

The sample dataset is shown below:

0	GENERAL	जोगाउन को लागि	छाया सेन्टर	ORG	गुठी विधेक ल्याएर ठमेल मा राज गुठि को जग्गा मा बने को छाया सेन्टर जस्ता लाई जोगाउन को लागि ल्याउदैछ विधेक ।
1	GENERAL	लखेटनु पछ	NA	NA	दले ले देश सकेछन सबै बेचे र खान सुरू गरेछन अब दले लाई लखेटनु पछ ।
1	GENERAL	ससकृती ध्वस्त पार्ने	नेपाल	LOC	नेपाल को ससकृती ध्वस्त पार्ने योजना हो यो !
1	GENERAL	भुमाफिया	NA	NA	मठ मन्दिर गुम्बा का जग्गा हरु मा भुमाफिया को नजर परे को हुना ले यो विधेक आए को हो !
1	GENERAL	बेची सके	NA	NA	नेपाल का कल कर्खाना र नदि नाला बेची सके अब मठ मन्दीर लाई पनि बेची खाने नियत देखिन्छ ।
1	GENERAL	बेची खाने नियत	NA	NA	नेपाल का कल कर्खाना र नदि नाला बेची सके अब मठ मन्दीर लाई पनि बेची खाने नियत देखिन्छ ।
1	GENERAL	कालोमोसो लगाएर भगउनु पर्छ	भणडारि	PERSON	यो भणडारि लई कालोमोसो लगाएर भगउनु पर्छ तर्क चै नदीने खालि राम्रो छ मात्र भन्ने ।
1	GENERAL	झोले , झन्डे	NA	NA	अन्य देश मा जमिन राज्य को हुन्छ , , जनता रदेश मा घात हुदैन , तर नेपाल मा , जमिन झोले , झन्डे ले घात गरे अ॰पने बनाउ छन ।
0	GENERAL	सनातन को सम्राट	राजा	MISC	सरकार बिदेशि को डिजाइन मा छ सनातन को सम्राट राजा लाइ हटाए पछि सास्कृतिक मास्स सजिलो हुन्छ भनेर तिब्र रुप मा अगाडि गणतन्त्र बड्दै छ ।

Figure 4.2: Sample dataset

For target term identification, opinion term extraction task and for the identification of which opinion term is targeted to which target term, the task is treated as sequence labelling task so, dataset was splitted into words and BIO tagging scheme is used.

4.4.2   Task Definition

There are various tasks under aspect based sentiment analysis. pontiki et.al. (9) mentioned four tasks under ABSA. pontiki et.al. (10) again in 2015 introduced new task as opinion target expression which is the target term in text along with previous tasks as of 2014.

If S is a sentence that consists of several words: w1, w2, ..., wn. The task is to detect following things:

1. Opinion term, O_t , Where $O_t \in S$

2. Target term, T_t , where $T_t \in S$

3. (O_t , T_t)

4. Aspect category(A_c)
   where $A_c \in (GENERAL, FEEDBACK, PROFANITY, VIOLENCE)$

5. Polarity of aspect category (P_{ac})
   where $P_{ac} \in (0, 1)$

4.4.3 Methodology for Target term extraction, Opinion term extraction and Target Oriented Opinion Word Extraction task
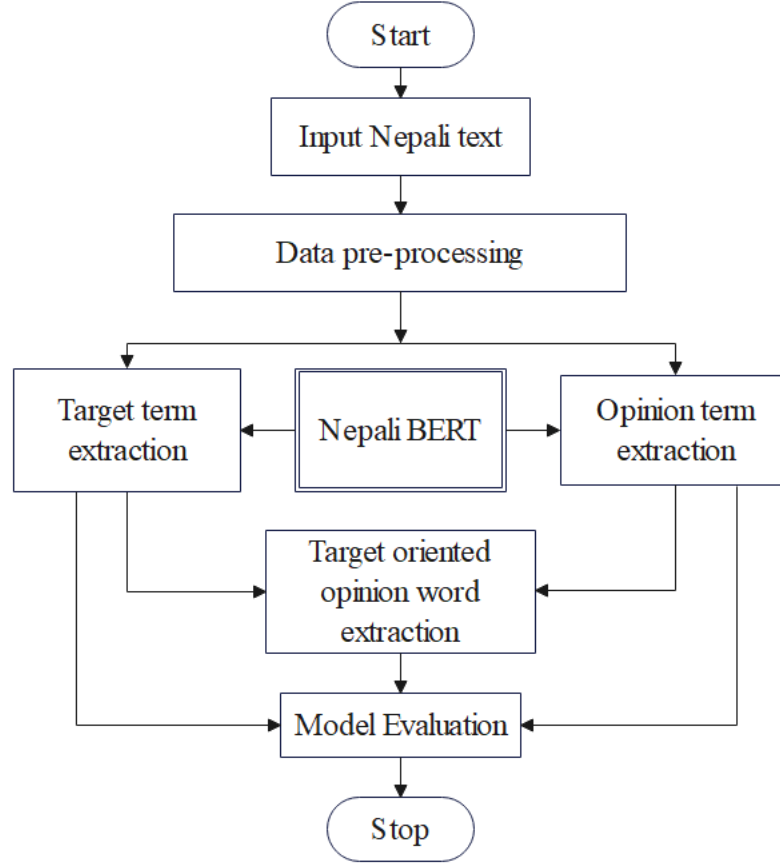
Figure 4.3: Flowchart for Extraction task

Initially target term is extracted, for target term extraction, as this task is extraction task so, Named Entity Recognition (NER) technique is used for this. Target terms are normally nouns or noun phrases. The task is treated as a sequence labelling problem utilizing the BIO tagging scheme.

Following steps are taken to perform the task:

1. **Dataset loading, preprocessing and preparing the dataset for the use:** The sequence length is limited to 128 tokens here since the maximum size of the dataset is 93 tokens. As per the suggestion of the Bert paper batch size of 32 will be used here. Bert can handle token sequences up to 512 in length. For tokenization, BertTokenizer class is imported from the hugging face transformers library.

2. **Loading BERT model:** NepaliBERT model by the author Rajan Ghimire

is used for this task. Based on the BERT model, NepaliBERT is a cutting-edge language model for Nepali. 67 millions lines of unprocessed Nepali text data were used to train NepaliBERT. A LARGE SCALE NEPALI TEXT CORPUS and the Oscar dataset were combined to create the final data set. Additional fine-tuning was done on existing model. The model is loaded with following hyperparameters:

Overall description of the hyperparameters

Hyperparameters	Value
Attention Dropout Probability	0.1
Hidden Dropout Probability	0.1
Hidden Size	768
Num Attention Heads	12
Num Hidden Layers	6
Hidden Activation Function	gelu
Intermediate Layer Size	3072
Vocabulary Size	50000
Max Position Embeddings	512

Table 4.4: Summary of Hyperparameters and their value

3. **Train model:** While training pytorch library is used and the steps in the workflow are:

   (a) Forward pass: Model learns the pattern here.

   (b) Calculating loss values

   (c) Zero the gradients of the optimizer

   (d) Perform backpropagation on the loss function

   (e) Step the optimizer

4. **Evaluate model:**

   (a) Iterate through test data batches and make predictions on the test data via forward pass.

(b) Calculate the loss and other evaluation metrics.

(c) No need for backpropagation here.

(d) Aggregate metrics across test set and calculate averages.

5. **Applying the model for new sentences:** We first tokenize the sentence then we run the sentence through the model.

The above task is treated as NER task.

After target term extraction, opinion term is extracted. The task involves extracting opinion terms from text data. Opinion terms are words or phrases that express sentiments or attitudes towards specific targets or aspects. A BERT-based model is chosen as the state-of-the-art best model for performing the opinion term extraction task. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model known for its effectiveness in various NLP tasks. The original dataset provided by the author consists of 4035 annotations. Additionally, extra data was manually annotated, resulting in 1031 additional annotations. So in total 5066. Annotations involve labeling specific spans of text as opinion terms. The opinion term extraction task is treated as a sequence labeling task. In sequence labeling, each token or word in the input text is assigned a label that indicates its role or category. In this case, the labels likely indicate whether a token is part of an opinion term or not. BIO tagging scheme is used here. The dataset is formatted to suit the sequence labeling task. This formatted dataset is then used for training the BERT-based model for opinion term extraction. Overall, the process involves using a BERT-based model to extract opinion terms from text data by treating the task as a sequence labeling task. The dataset is annotated accordingly and formatted to facilitate training of the model, incorporating both the original annotations and additional manually annotated data.

Now the final task here is:

After the extraction of target terms and opinion terms individually, the final task here is to detect which opinion is opinionated to which target term. For this work, the assumption made here is that the opinion term is targeted to nearest target term. There are several papers saying several approaches for this task but here the approach used is that target term nearby the opinion term is the probable target.

BERT model seems to perform better in this task. If the label is opinion term then the nearest labeled target term is treated as the exact one towards which the opinion is opinionated to. The distance here is the spatial gap measured in terms of token positions between the target and the nearest opinion term found.

4.4.4 Methodology for Aspect Category detection
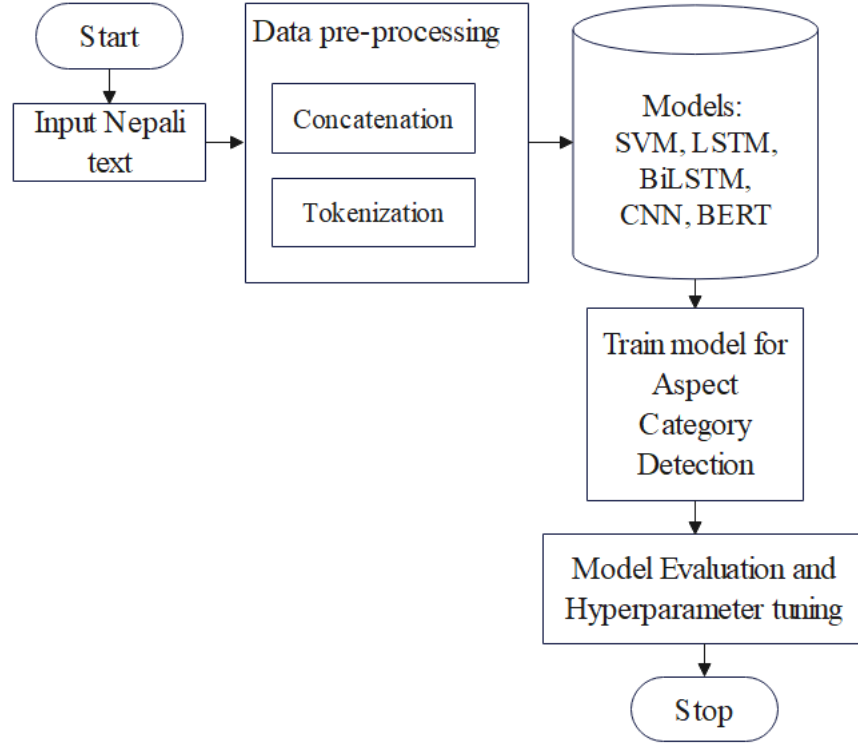
Figure 4.4: Flowchart for ACD task

1. **Data Preprocessing** Initially data is loaded and cleaned then feature engineering is done- in which Columns 'text' and 'opinion' are concatenated to create a new column 'Combined text'. Then text tokenization and vectorization is done.

2. **Model training** Model is trained, various models are explored with concatenated and non concatenated text.

   (a) SVM(Support Vector Machine): The supervised (feed-me) machine learning method SVM is useful for problems involving both regression and classification. TfidfVectorizer was used to tokenize and vectorize the combined text data. SVM classifier was created with a linear kernel

37

and trained it using the training data. Finally the trained classifier was used to make predictions on the test set and calculated evaluation measures of the model.

(b) LSTM(Long Short Term Memory): Long Short-Term Memory (LSTM) neural network was build for aspect categorization in a text dataset. Keras Tokenizer is used to convert text data into sequences of integers. LabelEncoder was used to encode categorical target variable 'aspect' into numerical format. Then dataset was splitted into training and testing sets for the 'aspect' variable. Then a sequential model was defined with an embedding layer, an LSTM layer, and a dense output layer for the 'aspect' variable. Then model was compiled with categorical cross-entropy loss and the Adam optimizer. Then the LSTM model was trained using the training data, specifying the number of epochs, batch size, and validation data. Finally predictions were made on the test set for the 'aspect' variable.

(c) BiLSTM: Bi-directional Long Short-Term Memory (BiLSTM) neural network was build for aspect categorization in a text dataset.

(d) Convolutional Neural Network(CNN): Convolutional Neural Network (CNN) model was trained for aspect classification using TensorFlow and scikit-learn metrics.

(e) Bidirectional Encoder Representations From Transformers(BERT): BERT based model was trained for aspect classification using TensorFlow and the Transformers library. Steps involved are explained below:

Data Preprocessing: Unnecessary columns were dropped and remaining columns were renamed. BERT tokenizer was used to tokenize and convert text data into input IDs and attention masks. One-hot encoding was performed for aspect labels. Then tensorFlow datasets was created.

Building BERT Model for Aspect Classification: BERT model-was loaded from the Transformers library. Model was compiled with Adam optimizer, categorical cross-entropy loss, and accuracy metric.

Training the Model.

Evaluation of Model Performance:scikit-learn's confusion matrix and classification report was used to evaluate the model's performance.

Model	Hyperparameters	Values
SVM	Kernel	Linear
LSTM	Optimizer	Adam
BiLSTM	Optimizer	Adam
BERT	Epochs	8
	Batch size	32
	Learning rate	2e-5 (0.00002)
	Loss function	CrossEntropyLoss
	Tokenizer	BERT Tokenizer

Table 4.5: Hyperparameter settings Used in Different Models

4.4.5 Methodology for Aspect Category Polarity detection
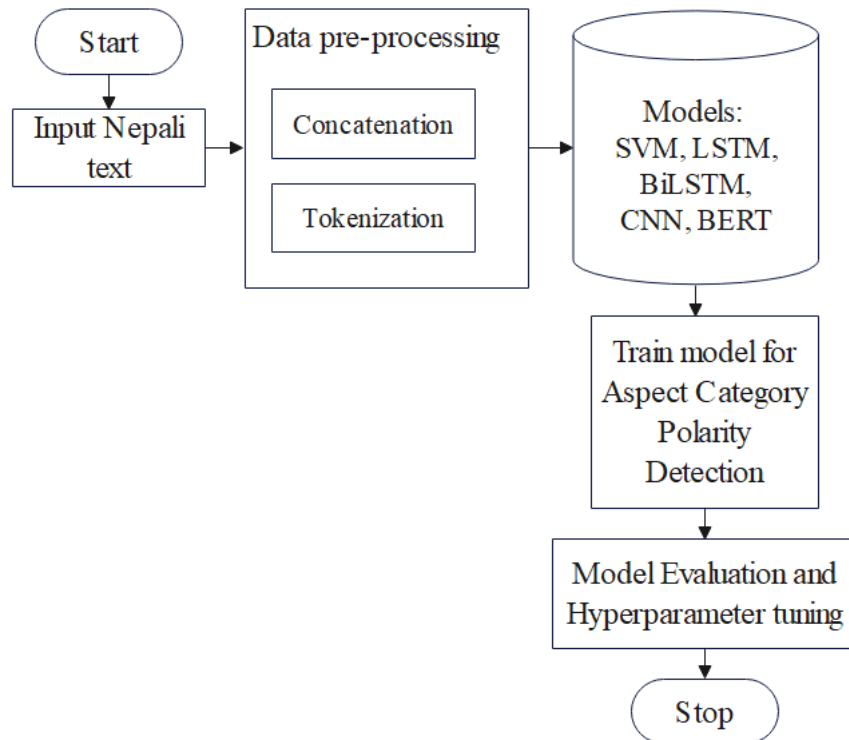
Figure 4.5: Flowchart for ACP detection task

This task is treated as binary classification task. SVM, LSTM, BiLSTM, CNN and

BERT-based models were trained for the classification of aspect category polarity. The methods applied for this task are explained below:

1. **Data Preprocessing** Initially data is loaded and cleaned then feature engineering is done in which the text is concatenated with the opinion term. The ACP detection task is carried out both for concatenated and non-concatenated embedding of text. Then text tokenization and vectorization is done.

2. **Model training** Model is trained, various models are explored with concatenated and non concatenated embedding of text.

   (a) SVM(Support Vector Machine): The supervised (feed-me) machine learning method SVM is useful for problems involving both regression and classification. TfidfVectorizer was used to tokenize and vectorize the combined text data. SVM classifier was created with a linear kernel and trained it using the training data. Finally the trained classifier was used to make predictions on the test set and calculated evaluation measures of the model.

   (b) LSTM(Long Short Term Memory): Long Short-Term Memory (LSTM) neural network was build for polarity categorization in a text dataset. Keras Tokenizer is used to convert text data into sequences of integers. LabelEncoder was used to encode categorical target variable 'polarity' into numerical format. Then dataset was splitted into training and testing sets for the 'polarity' variable. Then a sequential model was defined with an embedding layer, an LSTM layer, and output layer for the 'polarity' variable. Then model was compiled. Then the LSTM model was trained using the training data, specifying the number of epochs, batch size, and validation data. Finally predictions were made on the test set for the 'polarity' variable.

   (c) BiLSTM: Bi-directional Long Short-Term Memory (BiLSTM) neural network was build for polarity categorization in a text dataset.

   (d) Convolutional Neural Network(CNN): Convolutional Neural Network (CNN) model was trained for polarity classification using TensorFlow

and scikit-learn metrics.

(e) Bidirectional Encoder Representations From Transformers(BERT): BERT based model was trained for polarity classification using Pytorch and the Transformers library. steps involved are explained below:

Data Preprocessing: Unnecessary columns were dropped and remaining columns were renamed. BERT tokenizer was used to tokenize and convert text data into input IDs and attention masks. One-hot encoding was performed for polarity labels. Then tensors were created.

Building BERT Model for polarity Classification: BERT model was loaded from the Transformers library. Multilingual BERT was fine-tuned for ACP classification task. Model was compiled with Adam optimizer, categorical cross-entropy loss, and accuracy metric.

Training the Model.

Evaluation of Model Performance: scikit-learn's confusion matrix and classification report was used to evaluate the model's performance.

Hyperparameters	Value/Selection
Epochs	10
Optimizer	AdamW
Loss Function	CrossEntropyLoss
Batch size	32
Learning rate	1×10^{-5}

Table 4.6: Hyperparameter Settings for BERT model in ACP detection task.

4.5 Evaluation Metrics

To calculate the performance of the proposed system, ROUGE-L score, Confusion Matrix, Accuracy, Precision, Recall, F1-Score are used.

- Confusion Matrix: Confusion matrix gives matrix as output. In confusion matrix 4 important terms are there:

  1. True Positive(TP) : The model predicted YES and the actual output

was also YES.

2. True Negative(TN) : The model predicted NO and the actual output was NO.

3. False Positive(FP) : The model predicted YES and the actual output was NO.

4. False Negative(FN) : The model predicted NO and the actual output was YES. That is the model falsely predicts NO when it was YES in actual.

- Accuracy is the closeness of measurement to a specific value.
  Formula:
  Accuracy= $\dfrac{(TP + TN)}{(TP + FN + FP + TN)}$

- Precision is the fraction of correctly classified positive examples divided by the number of examples labeled by the system.
  Formula:
  Precision= $\dfrac{(TP)}{(TP + FP)}$

- Recall, also called true positive rate(TPR) or sensitivity. Recall provides us with information on the proportion of positive instances the model correctly identified out of all the positive instances that are present in the dataset.
  Recall= $\dfrac{(TP)}{(TP + FN)}$

- Out of all the negative instances that are actually present in the dataset, specificity indicates the proportion of negative instances that the model was able to correctly identify.
  Specificity = $\dfrac{(TN)}{(TN + FP)}$

- F1-Score is the harmonic mean of precision and recall. It gives the single measure of comparison and higher F1-score is better.
  F1-Score= $\dfrac{(2 * Precision * Recall)}{(Precision + Recall)}$

- ROUGE (Recall-Oriented Understudy for Gisting Evaluation) score is used to evaluate the quality of summaries generated by automatic summarization systems in information retrieval(IR) and NLP. The longest common subsequence (LCS) between the reference summary and the generated summary is measured by ROUGE-L. It uses the LCS's length to calculate F1-score, precision, and recall. ROUGE scores typically range from 0 to 1, with higher values indicating better agreement. If the ROUGE score is 0, it indicates no overlap whereas ROUGE score 1 indicates perfect overlap.

4.6 Tools and techniques

For this thesis work, Google Colab serves as the primary development environment, leveraging the versatility of Python and its powerful libraries. Key Libraries such as NLTK for natural language processing, Scikit-Learn for machine learning algorithms, and deep learning frameworks including Tensorflow, PyTorch, and Keras are employed extensively to train and evaluate several models. Additionally, Pandas and NumPy are utilized for efficient data manipulation, while Matplotlib aids in visualizing the results of data preprocessing and model evaluations. This comprehensive toolkit ensures robust experimentation and insightful analysis throughout the research process.

CHAPTER 5

RESULTS AND DISCUSSION

The main task here are the Information extraction task and classification task. Under IE task, target term extraction, opinion term extraction and target oriented opinion word extraction task are performed. Under classification task, binary classification for aspect category polarity detection and multi-class classification for aspect category detection is done. Let us see the results obtained and discuss different scenerios.

5.1   Results of Information Extraction Task

Information extraction task: IE task consists of target term extraction, opinion term extraction and extraction of opinion term and the target towards which opinion is opinionated to that is pair of opinion term and target term. At the start dataset was made in the new format as shown in Dataset description section. Every word was annotated and given a tag following the BIO tagging scheme as most papers doing TOWE task follow this. The result obtained applying BERT is shown below:

[CLS]	नेपाल	राम्रो	छ	तर	बिडम्बना	त्यहाँ	को	नेता	घुसखोरी	छ	सरकार	नै	तेसैको	छ	[SEP]
O	B-LOC	O	O	O	O	O	O	B-MISC	O	O	B-MISC	O	O	O	O

Figure 5.1: Output for target term extraction task

[CLS]/O   नेपाल/O   राम्रो/B-OT   छ/O   तर/O   बिडम्बना/O   त्यहाँ/O   को/O
नेता/O   घुसखोरी/B-OT   छ/O   सरकार/O   नै/O   तेसैको/B-OT   छ/O   [SEP]/O

Figure 5.2: Output for opinion term extraction task

The opinion term: राम्रो.. is targeted towards the location: नेपाल

The opinion term: घुसखोरी.. - is targeted towards the misc: नेता

The opinion term: तेस्रैको.. - is targeted towards the misc: सरकार

Figure 5.3: Opinion term and target pair

	Precision	Recall	F1-score
ROUGE-L	0.9513	0.9561	0.9537

Table 5.1: ROUGE-L Scores for target term extraction task

	Precision	Recall	F1-score
ROUGE-L	0.8466	0.8680	0.8572

Table 5.2: ROUGE-L Scores for opinion term extraction task

	Precision	Recall	F1-score
ROUGE-L	0.8041	0.8225	0.8132

Table 5.3: ROUGE-L Scores for opinion term and target term extraction task

5.1.1 Discussion of IE task

The ROUGE-L score for target term extraction task is high because of using NepaliBert for this NER task. 67 millions lines of unprocessed Nepali text data were used to train NepaliBert. Despite its high performance, some errors were observed. Let us discuss first error case scenerio. In case when words may have multiple meanings depending on the context then the NER model faced challenges in correctly classifying entities.

Sentence	Actual Label	Predicted Label
बर्दी फुकालेर फाल नेपाल पुलिस ।	['O', 'O', 'O', 'B-ORG', 'I-ORG', 'O']	['O', 'O', 'O', 'B-LOC', 'B-MISC', 'O']

Figure 5.4: Error example

The word नेपाल(Nepal) is used to refer to both a location and part of the name of organization(Nepal Police). In above example it refers to part of the name of organization but model predicts it as location. This example shows that the model struggled with correctly identifying "Nepal Police" as an organization, possibly due to insufficient contextual understanding or training data bias. Improving the model might involve using more diverse and contextually rich training data. Also, several other errors were primarily attributed to the presence of unknown words and incorrect spellings, which the model had difficulty processing accurately. Additionally, the morphological complexity and regional dialect variations of Nepali might have contributed to these discrepancies. The model's performance could potentially be improved by incorporating more diverse training data, including texts from different dialects and genres, and by employing advanced preprocessing techniques to handle spelling variations and rare words more effectively.

The ROUGE-L score for opinion term extraction is low as compared to that of target term extraction task. This is because opinion terms often consist of adjectives, adverbs, and subjective expressions that are more context-dependent and varied than the more straightforward nouns and noun phrases that constitute target terms. Figures of speech such as metaphor and sarcasm are the source of the inaccuracy.

The ROUGE-L for opinion term and target term pair extraction task is lesser than above tasks. The task here is to detect which opinion is opinionated to which target term. For this work, the assumption made here is that the opinion term is targeted to nearest target term. This is not the case always as there might be the scenerio when long term dependencies need to be considered, this depends upon the nature of data we are working on. To improve performance in this paired extraction task, future work could focus on enhancing the model's ability to understand and interpret complex sentence structures and context.

5.2   Results of Classification Tasks

1. **Aspect category detection (ACD) task:** This task is treated as multi-class classification task. Best result for this task was obtained through BERT.

The experiment comparing concatenated and non-concatenated embeddings is displayed in the following table. Concatenated features are feature vectors of words from opinion terms and sentences combined, whereas not concatenated features just show sentence embeddings.

Features	Concatenated				Not Concatenated			
Models	P	R	F1	Acc.	P	R	F1	Acc.
SVM	0.7230	0.7357	0.6960	0.7375	0.6060	0.6819	0.6073	0.6819
LSTM	0.7278	0.7239	0.7238	0.7239	0.6394	0.6755	0.6507	0.6755
BiLSTM	0.7397	0.7416	0.7396	0.7416	0.6494	0.6824	0.6594	0.6824
CNN	0.7576	0.7584	0.7560	0.7584	0.6638	0.6854	0.6700	0.6854
BERT	0.8289	0.8304	**0.8278**	0.8304	0.7562	0.7583	0.7572	0.7602

Table 5.4: Evaluation measures for Aspect category detection task when concatenated vs not concatenated
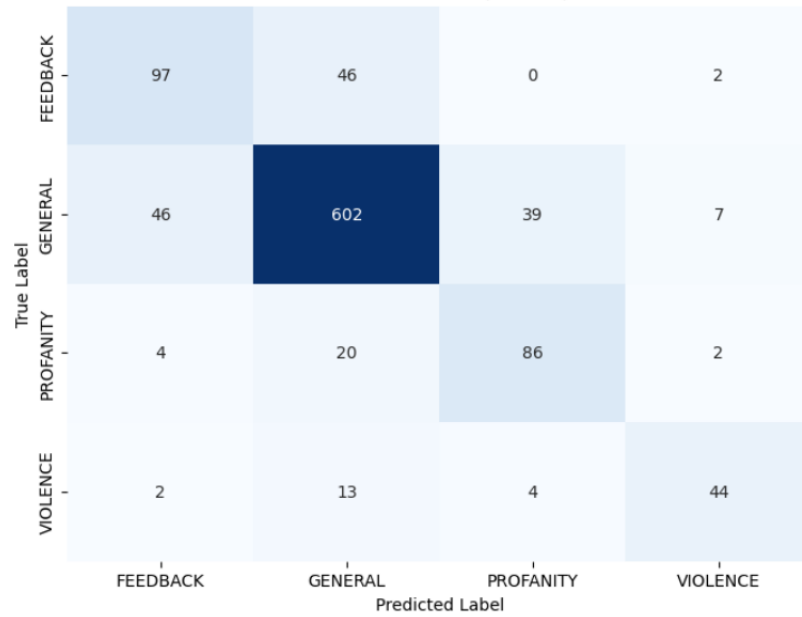
Figure 5.5: Confusion matrix for Aspect Category detection task when concatenated using BERT.

The output of ACD task is shown below:

S.No.	Text	Actual	Predicted
1	रातो कोट लगाउने सर ले राम्रो कुरा गर्नु भयो निलो कोट ले गोबर कुरा गर्यो निलो कोटे छि छि छि ।	GENERAL	GENERAL
2	माडबारि कुनै हालात मा नेपाली होइन भतुवा हो ।	GENERAL	PROFANITY
3	यिन लाई मार्न को निम्ति छिट्टै भगवान श्री कृष्ण को जन्म हुन्छ ।	VIOLENCE	VIOLENCE
4	राम बहादुर बोम्जान माथि छिटो भन्दा छिटो छानबिन सुरु गरियोस सरकार ।	FEEDBACK	FEEDBACK
5	कोभिड मा नि घुस खाए चुतिया हरुले।	PROFANITY	PROFANITY

Figure 5.6: Few output of Aspect categories detection

2. **Aspect Category Polarity Detection:** Third is polarity detection task. This task is treated as binary classification task. Best result for this task was obtained through BERT. The experiment comparing concatenated and non-concatenated embeddings is displayed in the following table. Concatenated features are feature vectors of words from opinion terms and sentences combined, whereas not concatenated features just show sentence embeddings. The results obtained are shown below:

Features	Concatenated				Not Concatenated			
Models	P	R	F1	Acc.	P	R	F1	Acc.
SVM	0.7544	0.7525	0.7520	0.7525	0.7329	0.7308	0.7301	0.7308
LSTM	0.7312	0.7298	0.7293	0.7298	0.6896	0.6893	0.6892	0.6893
BiLSTM	0.7446	0.7436	0.7440	0.7436	0.6957	0.6943	0.6937	0.6943
CNN	0.7464	0.7436	0.7429	0.7436	0.7052	0.7032	0.7024	0.7032
BERT	0.8325	0.8313	**0.8311**	0.8313	0.756	0.75	0.76	0.766

Table 5.5: Evaluation measures in aspect category polarity detection task of a given text when concatenated vs not concatenated
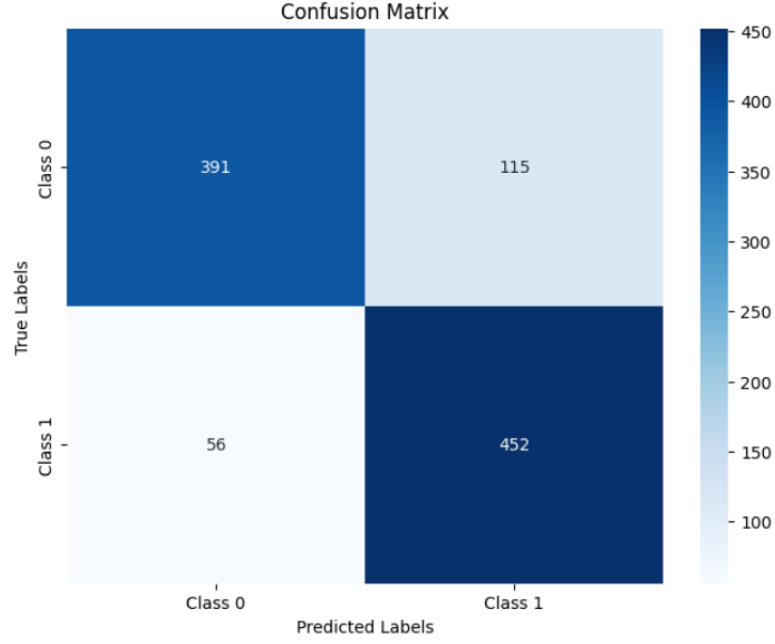
Figure 5.7: Confusion matrix for Aspect Category polarity detection task when concatenated using BERT.

5.2.1 Discussion of Classification task

The classification task is applied for ACD and ACP tasks. Let us discuss the results of ACD task. Several models were implemented for these task but the best was seen with multilingual BERT. The performance of the BERT model on our classification task was evaluated using a confusion matrix. The model was highly effective in correctly identifying GENERAL instances. The high performance on this class may be attributed to the larger number of instances (3547) in the training data, which likely provided the model with sufficient examples to learn from. But GENERAL class has considerable misclassification into FEEDBACK and PROFANITY, indicating that the model struggles to differentiate between these classes in some cases.

Also, FEEDBACK class has considerable misclassification into GENERAL. This is due to the reason that FEEDBACK and GENERAL class mostly for positive cases share same words although they have different contextual meanings. This issue prevails due to less training data so there is room for improvement for further researcher that they can increase the performance by increasing more instances of dataset.

The PROFANITY class has a notable degree of misclassification into the GENERAL

category. This issue arises because less severe profane terms and general negative opinions often share similar vocabulary, albeit used in different contexts. More diverse set of examples can tackle this issue in future.

In VIOLENCE class, among other classes few more are misclassified as GENERAL. This is mostly seen when text consists of several opinion carrying words and some word overlap with words under general negative opinions.

Now, let us discuss the results of ACP tasks. This task is also treated as a classification task. Several models were implemented for this task but the best was seen with Multilingual BERT. From several experiments it was seen that Multilingual BERT performs better than others. The performance of the BERT model was evaluated using a confusion matrix. There are more False Negatives than False Positives. The reason behind this was when sentence consists of more than one aspect category and when one aspect category's polarity is positive and anothers is negative then in such scenerio the positive examples seems to be predicted as negative due to which there are more false negatives. There are several other reasons also like less training data, use of sarcasm, subjective words etc.

5.3   Overall Discussion

Examples from the Information Extraction task is shown below:

Sentence	Actual target	Output target	Actual opinion	Output opinion
माडबारि कुनै हालात मा नेपाली होइन भतुवा हो । *Madbari is not nepali under any circumstances, instead a cadger.*	माडबारि *(Madbari)*	माडबारि *(Madbari)*	भतुवा *(Cadger)*	भतुवा *(Cadger)*
रबिसर लाई रिहाइ गरियोस । *Rabi sir need to be released.*	रबिसर *(Rabi sir)*	रबि *(Rabi)*	रिहाइ गरियोस *(to be released)*	रिहाइ गरियोस *(to be released)*
यो डकैति गर्ने लालिगुरास सहकारी लाई चाडो कारबाही होस्, रामकृष्ण ढकाल लाई चाडो न्याय चाहियो । *The robber Laligurans Cooperative should be punished quickly. Ramakrishna Dhakal wants justice.*	लालिगुरास सहकारी *(Laligurans cooperative)*	लालिगुरास सहकारी *(Laligurans cooperative)*	कारबाही होस् *(should be punished)*	कारबाही होस् *(should be punished)*
	रामकृष्ण ढकाल *(Ramkrishna Dhakal)*	रामकृष्ण ढकाल *(Ramkrishna Dhakal)*	न्याय चाहियो *(wants justice)*	न्याय *(Justice)*
सागर जि गुड दनक दिनु भयो गौतमे झोले लाई , सहि जबाफ दिनु भए को मा thq हजुर लाई । *Sagar ji you gave good logic to gautamey sycophant, thankyou for giving correct answer.*	सागर *(Sagar)*	सागर *(Sagar)*	गुड *(good)*	गुड दनक *(good logic)*
	गौतामे *(Gautamey)*	गौतामे *(Gautamey)*	झोले *(sycophant)*	झोले *(sycophant)*
	NA	*NA*	सहि जबाफ *(correct answer)*	*NA*
यो पत्रकार लाई लियर कसै ले भिर बाट झारिदेउ यार । *(Somebody take this journalist and throw from the cliff, man.)*	पत्रकार *(journalist)*	पत्रकार *(journalist)*	भिर बाट झारिदेउ *(throw from the cliff)*	भिर बाट झारिदेउ *(throw from the cliff)*

Figure 5.8: End to end inference

In above figure, few accurate examples of IE tasks are shown. In this thesis work, data is increased to 5066 that is 1031 datasets were added in 4035 datasets from previous author along with that the NepaliBERT model is used for NER task. Previously NepaliBERT has not been used in ABSA tasks in nepali text. Previously ROUGE-L score has not been used as a evaluation measure in IE task in Nepali due to which it became hard to make comparison of our obtained values. However, for the purpose of comparison we have calculated precision, recall and f1score for target term identification task.

Model	Precision	Recall	F1-score
BiLSTM + CRF	0.790	0.800	0.794
NepaliBERT	0.958393	0.961416	**0.959499**

Table 5.6: Performance comparison for Target term identification task

In above table BiLSTM + CRF model is used by author of (1) although not clearly mentioned in their paper about target term extraction as a separate task under ABSA. We use named entity recognition method for target term extraction task

using NepaliBERT model. The score is quite nice but inaccuracy occurs in this task due to several reasons such as text normalization, incorrect spelling, lack of enough data to train, use of local words (not present in vocab) and unknown words. Results obtained is compared with existing papers . As very few papers are there for nepali dataset, comparison for aspect polarity detection task is done with the result of author **Singh et al.**(1)

Following table shows the comparison of evaluation measures in aspect category polarity detection task of a given text when concatenated vs not concatenated with author, the best performing model in our dataset is used for comparison.

Features	Models	Performance			
		P	R	F1	Acc.
Concatenated	BERT + LSTM	0.804	0.80	0.79	0.80
	BiLSTM	0.816	0.816	0.816	0.815
	BERT	0.8325	0.8313	**0.8311**	0.8313
Not Concatenated	BERT + LSTM	0.782	0.779	0.782	0.789
	BiLSTM	0.806	0.805	0.805	0.805
	BERT	0.75	0.76	0.756	0.766

Table 5.7: Comparison of evaluation measures in aspect category polarity task of a given text when concatenated vs not concatenated with author(1).

Above table shows the comparison of evaluation meacures in ACP detection task between our result and that of author of (1). (BERT + LSTM) model and BiLSTM models result is that of author whereas BERT models result is obtained here. Authors best result was seen with BiLSTM. We obtained an increase in F1 score by almost around 2 % for concatenated case. As compared to that of author, our dataset size is more and the class imbalance is less severe here as count of 0(positive) is 2563 and that of 1(negative) is 2503 whereas the count of 0 in that of author is 1899 and that of 1 is 2136. This reason is one of the contributing factor in the increase in evaluation measure in our model. In overall this thesis work can contribute in knowing the political perspective of represented towards their representatives as data is from news and politics domain. In future other researchers can increase the size of data, can normalize the data and can work on

various other reasons mentioned here that contribute in less accuracy of model and can pave their way towards new findings.

CHAPTER 6

CONCLUSION AND FUTURE WORK

The primary focus of this thesis is to undertake the Target-Oriented Opinion Word Extraction (TOWE) task alongside other Aspect-Based Sentiment Analysis (ABSA) tasks within the context of a dataset in the Nepali language. This paper represents a significant contribution to Aspect-Based Sentiment Analysis (ABSA) within the realm of social domain datasets in nepali. Notably, the TOWE task has not been previously explored for Nepali datasets. Our approach employs BERT based model along with position metrics, although we acknowledge the potential application of various other techniques. The ROUGE-L score for target term extraction task is quite high than that of opinion term extraction task and TOWE task, this is due to the reasons like spelling variations, highly subjective nature of words etc. Despite several challenges, our method has obtained convincing results using NepaliBERT for TOWE task with ROUGE-L score of 0.80 and has outperformed current State-of-the-art for classification tasks achieving F1-scores of 82.78% and 83.13% for ACD and ACP detection task respectively. We envisage expanding our dataset in the future. We plan to annotate data for additional domain also like telecom domain. Utilizing BERT addresses scalability concerns, given its capability to handle large datasets effectively. However, fine-tuning on domain-specific data remains crucial for different domains.

REFERENCES

[1] O. M. Singh, S. Timilsina, B. K. Bal, and A. Joshi, "Aspect based abusive sentiment detection in nepali social media texts," in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).* IEEE, 2020, pp. 301–308.

[2] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in information retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[3] M. Haselmayer and M. Jenny, "Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding," *Quality & quantity*, vol. 51, pp. 2623–2646, 2017.

[4] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Systems with Applications*, p. 119862, 2023.

[5] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE transactions on knowledge and data engineering*, vol. 28, no. 3, pp. 813–830, 2015.

[6] B. Liu, *Sentiment analysis and opinion mining.* Springer Nature, 2022.

[7] A. Da'u and N. Salim, "Aspect extraction on user textual reviews using multi-channel convolutional neural network," *PeerJ Computer Science*, vol. 5, p. e191, 2019.

[8] A. Mukherjee and B. Liu, "Aspect extraction through semi-supervised modeling," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2012, pp. 339–348.

[9] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," in *International Workshop on Semantic Evaluation*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:61955135

[10] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, 2015, pp. 486–495.

[11] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *ProWorkshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, 2016, pp. 19–30.

[12] L. Zhu, M. Xu, Y. Bao, Y. Xu, and X. Kong, "Deep learning for aspect-based sentiment analysis: a review," *PeerJ Computer Science*, vol. 8, p. e1044, 2022.

[13] M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, "Aspect based sentiment analysis in hindi: resource creation and evaluation," in *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, 2016, pp. 2703–2709.

[14] A. Pathak, S. Kumar, P. P. Roy, and B.-G. Kim, "Aspect-based sentiment analysis in hindi language by ensembling pre-trained mbert models," *Electronics*, vol. 10, no. 21, p. 2641, 2021.

[15] P. R. Amalia and E. Winarko, "Aspect-based sentiment analysis on indonesian restaurant review using a combination of convolutional neural network and contextualized word embedding," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, pp. 285–294, 2021.

[16] S. Cahyaningtyas, D. H. Fudholi, and A. F. Hidayatullah, "Deep learning for aspect-based sentiment analysis on indonesian hotels reviews," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 2021.

[17] S. Rani and W. Anwar, "Resource creation and evaluation of aspect based sentiment analysis in urdu," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the*

10th International Joint Conference on Natural Language Processing: Student Research Workshop, 2020, pp. 79–84.

[18] A.-S. Mohammad, M. M. Hammad, A. Sa'ad, A.-T. Saja, and E. Cambria, "Gated recurrent unit with multilingual universal sentence encoder for arabic aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 261, p. 107540, 2023.

[19] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels' reviews," *Journal of computational science*, vol. 27, pp. 386–393, 2018.

[20] S. Tamrakar, B. K. Bal, and R. B. Thapa, "Aspect based sentiment analysis of nepali text using support vector machine and naive bayes," *Technical Journal*, vol. 2, no. 1, pp. 22–29, 2020.

[21] G. Cai, H. Li, and T. Lan, "Opinion targets and sentiment terms extraction based on self-attention," in *2021 11th International Conference on Information Science and Technology (ICIST)*. IEEE, 2021, pp. 132–138.

[22] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.

[23] K. Bafna and D. Toshniwal, "Feature based summarization of customers' reviews of online products," *Procedia Computer Science*, vol. 22, pp. 142–151, 2013.

[24] S. S. Htay, K. T. Lynn *et al.*, "Extracting product features and opinion words using pattern knowledge in customer reviews," *The Scientific World Journal*, vol. 2013, 2013.

[25] W. Jin, H. H. Ho, and R. K. Srihari, "Opinionminer: a novel machine learning system for web opinion mining and extraction," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 1195–1204.

[26] I. Shamshurin, "Extracting domain-specific opinion words for sentiment analysis," in *Advances in Computational Intelligence: 11th Mexican International Conference on Artificial Intelligence, MICAI 2012, San Luis Potosí, Mexico, October 27–November 4, 2012. Revised Selected Papers, Part II 11.* Springer, 2013, pp. 58–68.

[27] P. Liu, S. Joty, and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 1433–1443.

[28] A. P. B. Veyseh, N. Nouri, F. Dernoncourt, D. Dou, and T. H. Nguyen, "Introducing syntactic structures into target opinion word extraction with deep learning," *arXiv preprint arXiv:2010.13378*, 2020.

[29] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[30] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Computational intelligence and neuroscience*, vol. 2015, pp. 30–30, 2015.

[31] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—a review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16–32, 2018.

[32] A. A. Q. Aqlan, B. Manjula, and R. Lakshman Naik, "A study of sentiment analysis: concepts, techniques, and challenges," in *Proceedings of International Conference on Computational Intelligence and Data Engineering: Proceedings of ICCIDE 2018.* Springer, 2019, pp. 147–162.

[33] B. Liu, "Sentiment analysis: A fascinating problem," in *Sentiment Analysis and Opinion Mining.* Springer, 2012, pp. 1–8.

[34] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data.* Springer, 2012, pp. 415–463.

[35] B. Huang, Y. Ou, and K. M. Carley, "Aspect level sentiment classification with

attention-over-attention neural networks," in *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11.* Springer, 2018, pp. 197–206.

[36] A. Ilmania, S. Cahyawijaya, A. Purwarianti *et al.*, "Aspect detection and sentiment classification using deep neural network for indonesian aspect-based sentiment analysis," in *2018 International Conference on Asian Language Processing (IALP).* IEEE, 2018, pp. 62–67.

[37] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Tree-structured regional cnn-lstm model for dimensional sentiment analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 581–591, 2019.

[38] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: tasks, methods, and challenges," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[39] B. Zhang, X. Li, X. Xu, K.-C. Leung, Z. Chen, and Y. Ye, "Knowledge guided capsule attention network for aspect-based sentiment analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2538–2551, 2020.

[40] D. Jurafsky and J. H. Martin, "Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition."

[41] R. Sunitha *et al.*, "Survey on aspect based sentiment analysis using machine learning techniques," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 10, pp. 1664–1684, 2021.

[42] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu, and A. Abusorrah, "Aspect-based sentiment analysis: A survey of deep learning methods," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 6, pp. 1358–1375, 2020.

[43] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th annual meeting of the*

association for computational linguistics: human language technologies, 2011, pp. 151–160.

[44] M. Saeidi, G. Bouchard, M. Liakata, and S. Riedel, "Sentihood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods," *arXiv preprint arXiv:1610.03771*, 2016.

[45] I. Chetviorkin and N. Loukachevitch, "Extraction of domain-specific opinion words for similar domains," in *Proceedings of the RANLP 2011 Workshop on Information Extraction and Knowledge Acquisition*, 2011, pp. 7–12.

[46] J. Fernando, M. L. Khodra, and A. A. Septiandri, "Aspect and opinion terms extraction using double embeddings and attention mechanism for indonesian hotel reviews," in *2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*. IEEE, 2019, pp. 1–6.

[47] Y. A. Winatmoko, A. A. Septiandri, and A. P. Sutiono, "Aspect and opinion term extraction for hotel reviews using transfer learning and auxiliary labels," *arXiv preprint arXiv:1909.11879*, 2019.

[48] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," *arXiv preprint cs/0205070*, 2002.

[49] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on bilstm," *Ieee Access*, vol. 7, pp. 51 522–51 532, 2019.

[50] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[51] J. Pei, A. Sun, and C. Li, "Targeted sentiment analysis: A data-driven categorization," *arXiv preprint arXiv:1905.03423*, 2019.

APPENDIX A : Additional results

Few examples with our top model

Sentence	Target term	Opinion term	ACD	ACP
माडबारि कुनै हालात मा नेपाली होइन भतुवा हो । *Madbari is not nepali under any circumstances, instead a cadger.*	माडबारि *(Madbari)*	भतुवा *(Cadger)*	GENERAL	1
रबि सर लाई रिहाइ गरियोस् । *Rabi sir need to be released .*	रबि *(Rabi)*	रिहाइ गरियोस *(to be released)*	FEEDBACK	0
यो डकैति गर्ने लालिगुरास सहकारी लाई चाडो कारबाही होस् रामकृष्ण ढकाल लाई चाडो न्याय चाहियो । *The robber Laligurans Cooperative should be punished quickly. Ramakrishna Dhakal wants justice quickly.*	लालिगुरास सहकारी *(Laligurans cooperative)*	कारबाही होस् *(should be punished)*	GENERAL	1
	रामकृष्ण ढकाल *(Ramkrishna Dhakal)*	न्याय *(Justice)*	GENERAL	1
सागर जि गुड दनक दिनु भयो गौतमे झोले लाई , सहि जबाफ दिनु भए को मा thq हजुर लाई । *Sagar ji you gave good logic to gautamey sycophant, thankyou for giving correct answer.*	सागर *(Sagar)*	गुड दनक *(good logic)*	GENERAL	0
	गौतामे *(Gautamey)*	झोले *(sycophant)*	GENERAL	1
यो पत्रकार लाई लियर कसै ले भिर बाट झारिदेउ यार । *(Somebody take this journalist and throw from the cliff, man.)*	पत्रकार *(journalist)*	भिर बाट झारिदेउ *(throw from the cliff)*	VIOLENCE	0

Figure 6.1: End-to-end inference examples with our top model from the Information Extraction and Classification task

Confusion matrices for Aspect category Detection task when concatenated
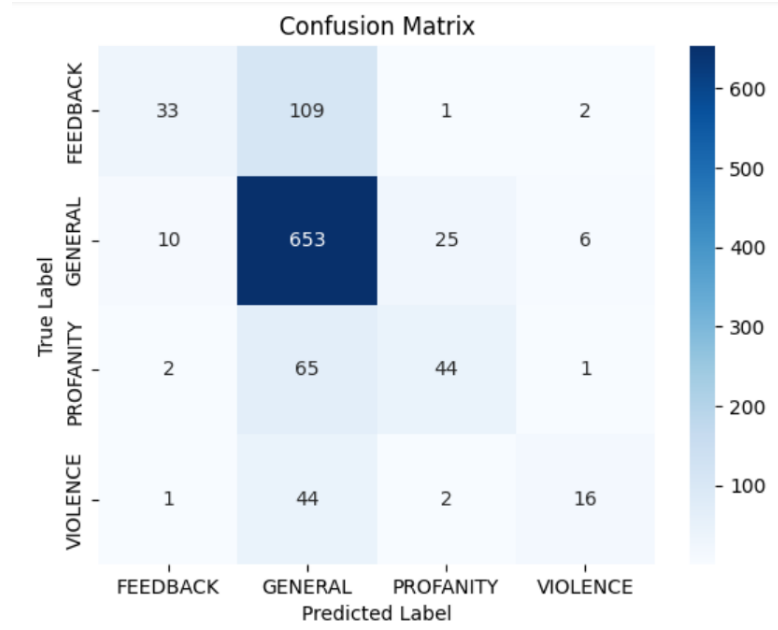
Figure 6.2: Confusion matrix for ACD task using SVM.
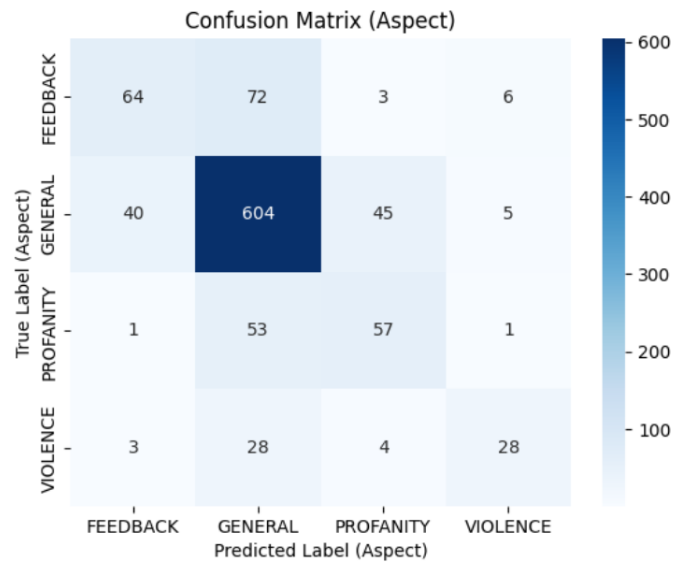
Figure 6.3: Confusion matrix for ACD task using LSTM.

Figure 6.4: Confusion matrix for ACD task using BiLSTM.

Figure 6.5: Confusion matrix for ACD task using CNN.

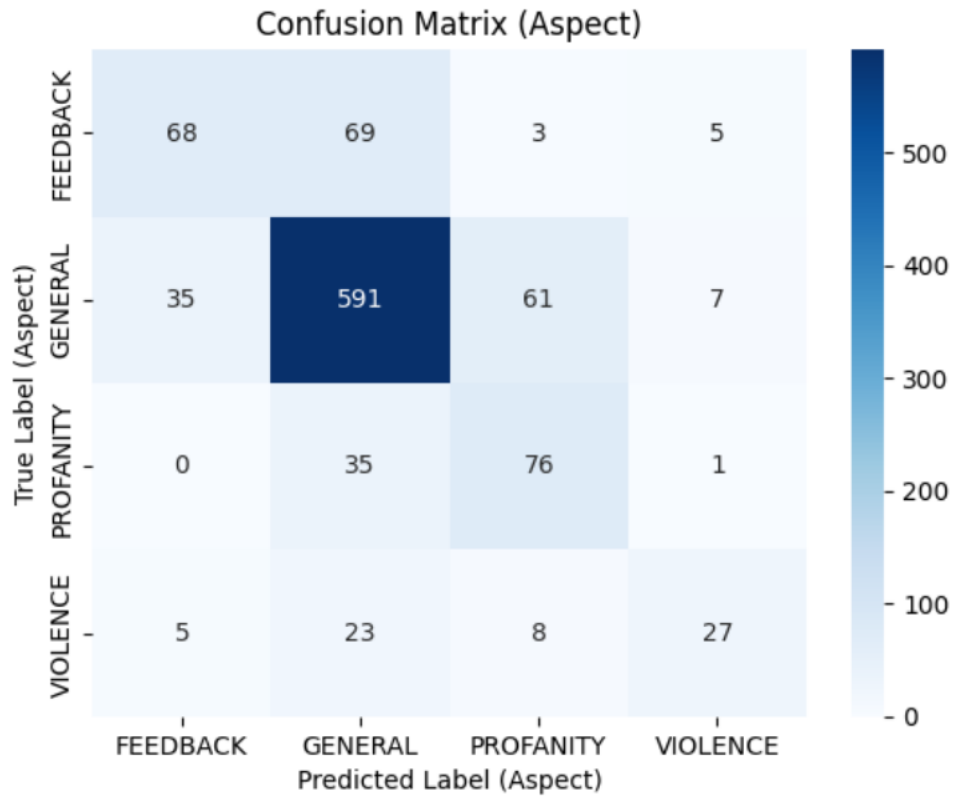**Figure 6.6:** Confusion matrix for ACD task using BERT.

Confusion matrices for Aspect category polarity detection task when concatenated
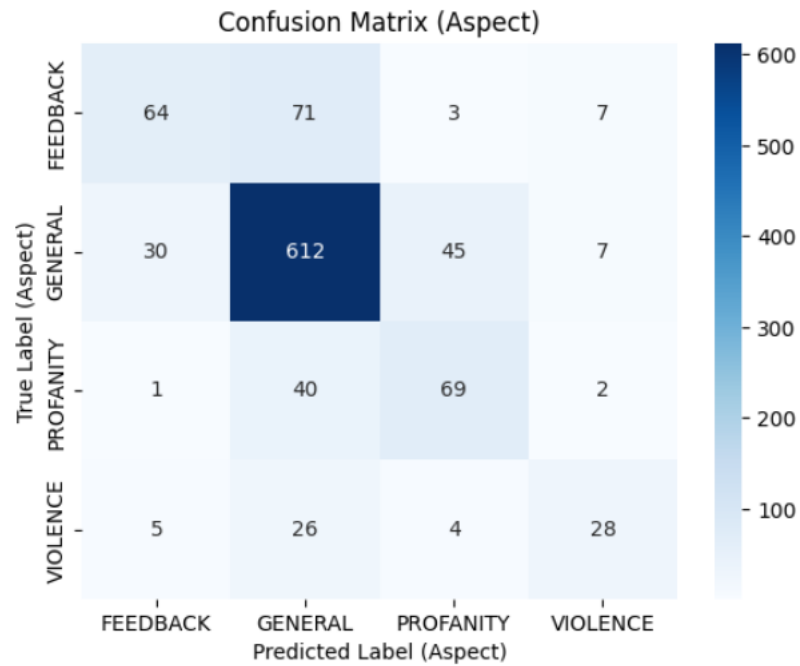

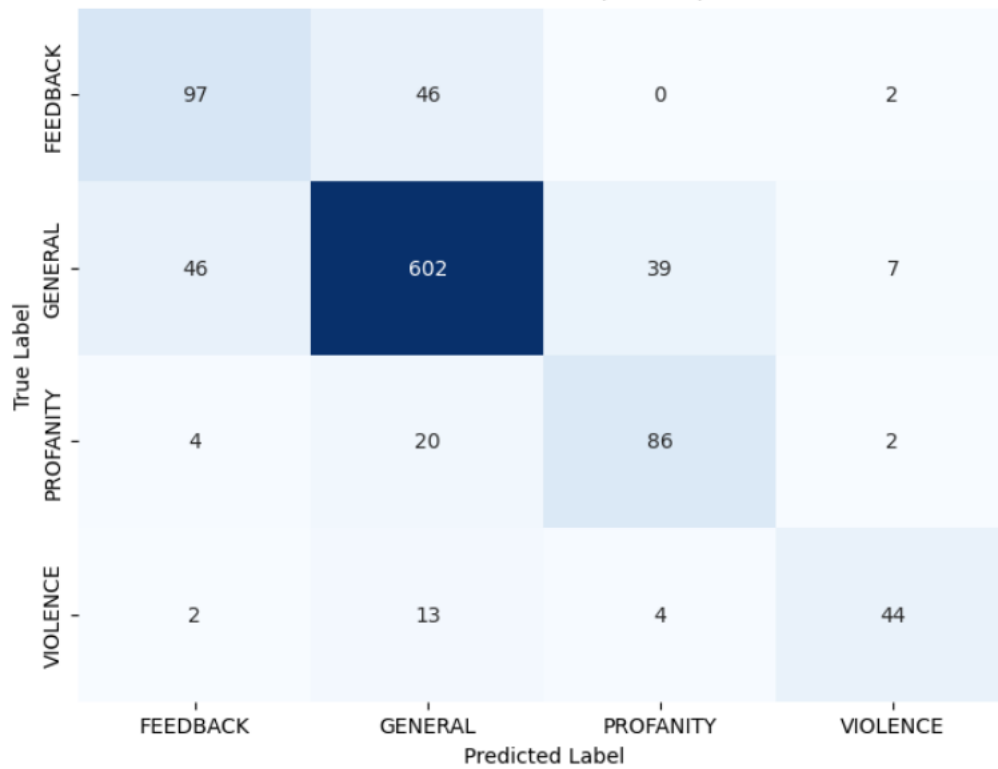**Figure 6.7:** Confusion matrix for ACP task using SVM.

Figure 6.8: Confusion matrix for ACP task using LSTM.

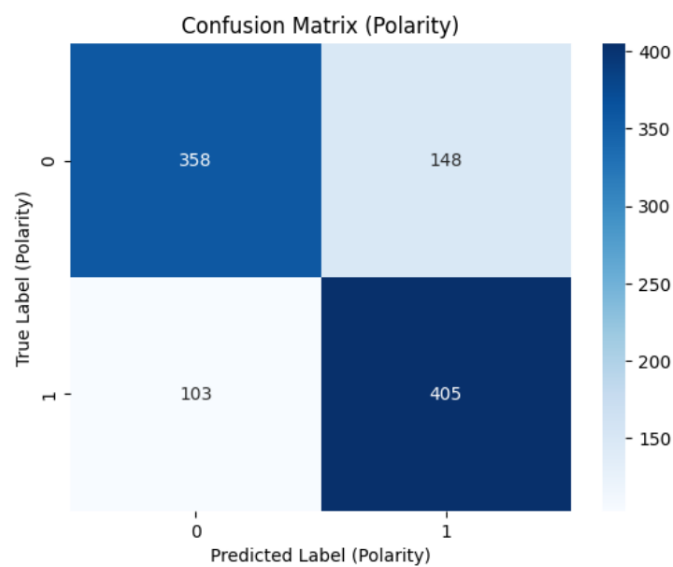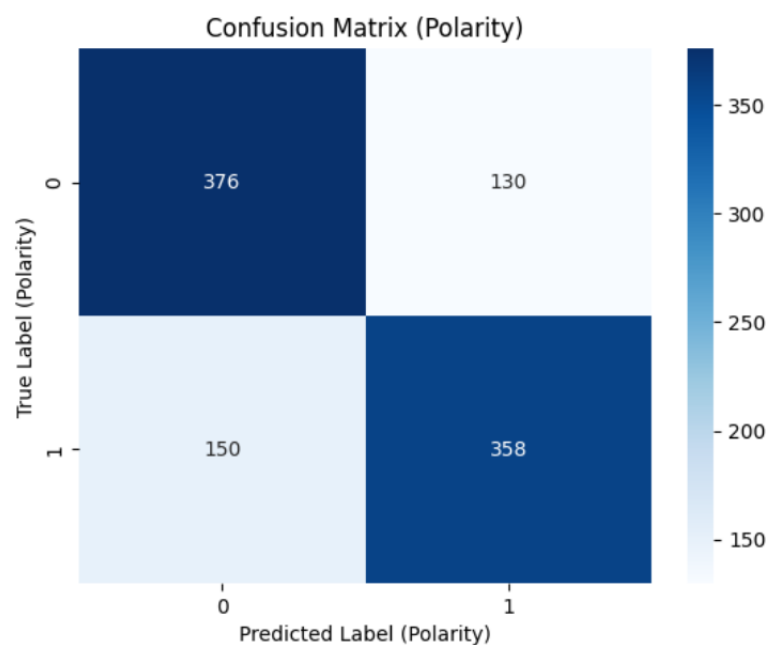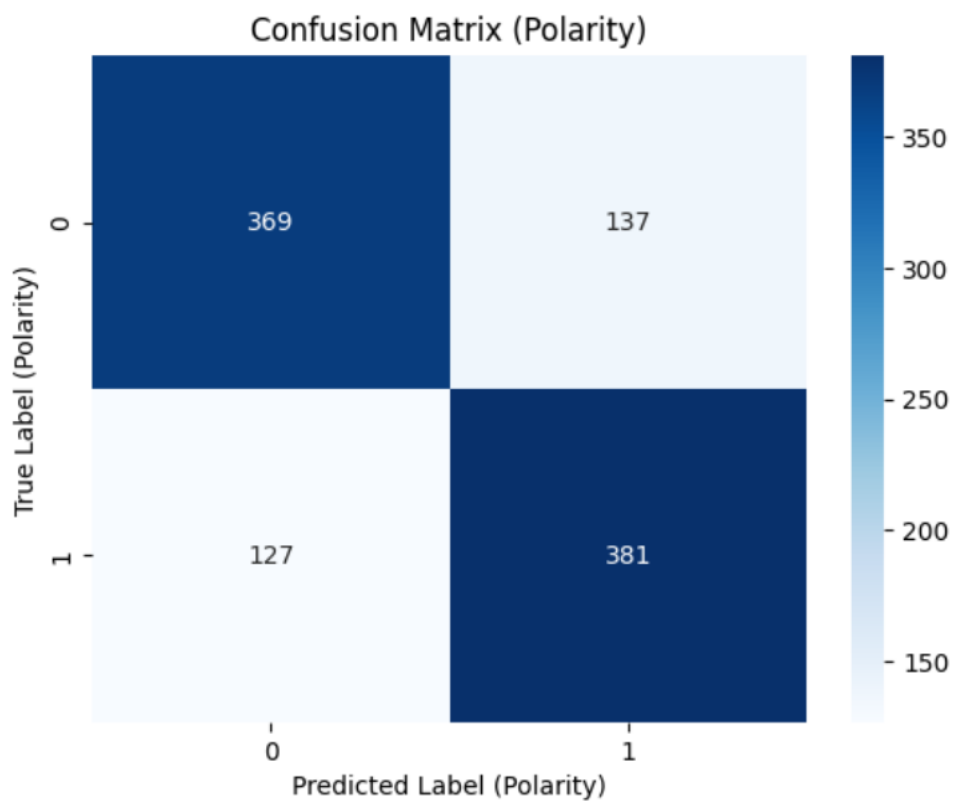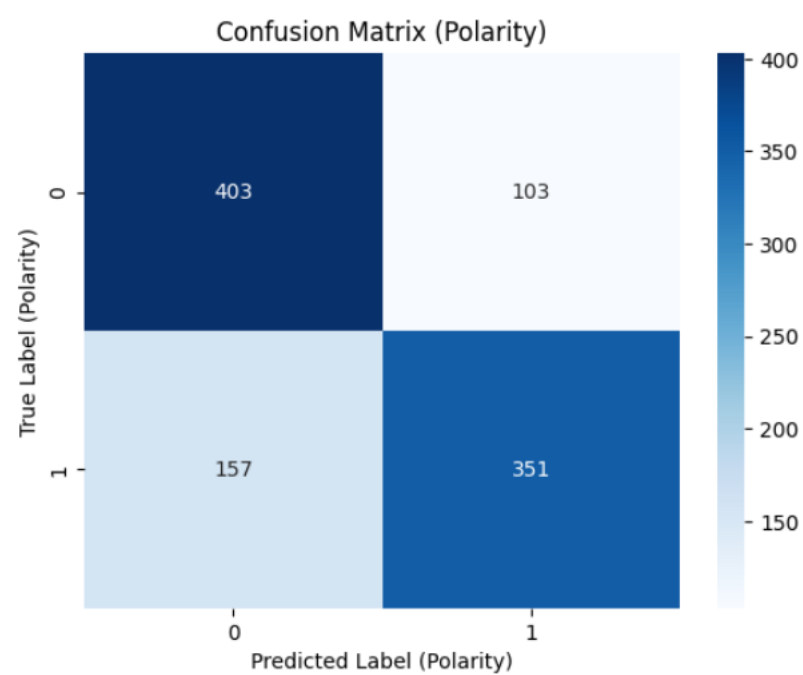**Figure 6.9:** Confusion matrix for ACP task using BiLSTM.

Figure 6.10: Confusion matrix for ACP task using CNN.

Email regarding Paper Acceptance and publication details

ISBM 2024 notification for paper 27   External   Inbox ×

ISBM 2024 <isbm2024@easychair.org>   Apr 11, 2024, 1:03 PM
to me

Dear Suprabha Regmi,

Paper ID : 27

Title : Aspect based sentiment analysis of nepali text using BERT based model

Congratulations! On behalf of the Program Committee of ISBM 2024 - Bangkok, Thailand, I am happy to inform you that your above-mentioned paper has been ACCEPTED for oral presentation in ISBM 2024 and publication in Springer LNNS series subject to fulfillment of Guidelines by Springer. An accepted paper will be published in the Springer proceedings (LNNS) only if the final version is accompanied by the payment information (i.e transaction reference number) subject to quality check as per Springer Guidelines.

Aspect Based Sentiment Analysis of Nepali Text using BERT based Model

Suprabha Regmi, Aman Shakya, Basanta Joshi

Tribhuvan University, Institute of Engineering, Pulchowk, Lalitpur, Nepal
078msdsa019.suprabha@pcampus.edu.np, aman.shakya@ioe.edu.np,
basanta@ioe.edu.np

Abstract. Aspect based sentiment analysis (ABSA) is a fine-grained sentiment analysis which aims in extracting opinion terms, opinion targets, aspect category and the polarity associated with each aspect category. Most ABSA research is in English with minimal work in Nepali. We have added 1031 nepali comments related to news and politics category from YouTube to the dataset by previous researcher under same annotation guideline to facilitate the research of ABSA in social media domain. This paper introduces target-oriented opinion term extraction (TOWE) in Nepali ABSA, alongside aspect category detection (ACD) and aspect category polarity (ACP) detection tasks. We present convincing results using NepaliBERT for TOWE with ROGUE-L score of 0.80. We experiment with various models including SVM, LSTM, BiLSTM, CNN, and multilingual BERT for ACD and aspect category polarity achieving F1-scores of 82.78% and 83.13% for ACD and ACP respectively. Our method has outperformed current State-of-the-art for classification tasks in Nepali.

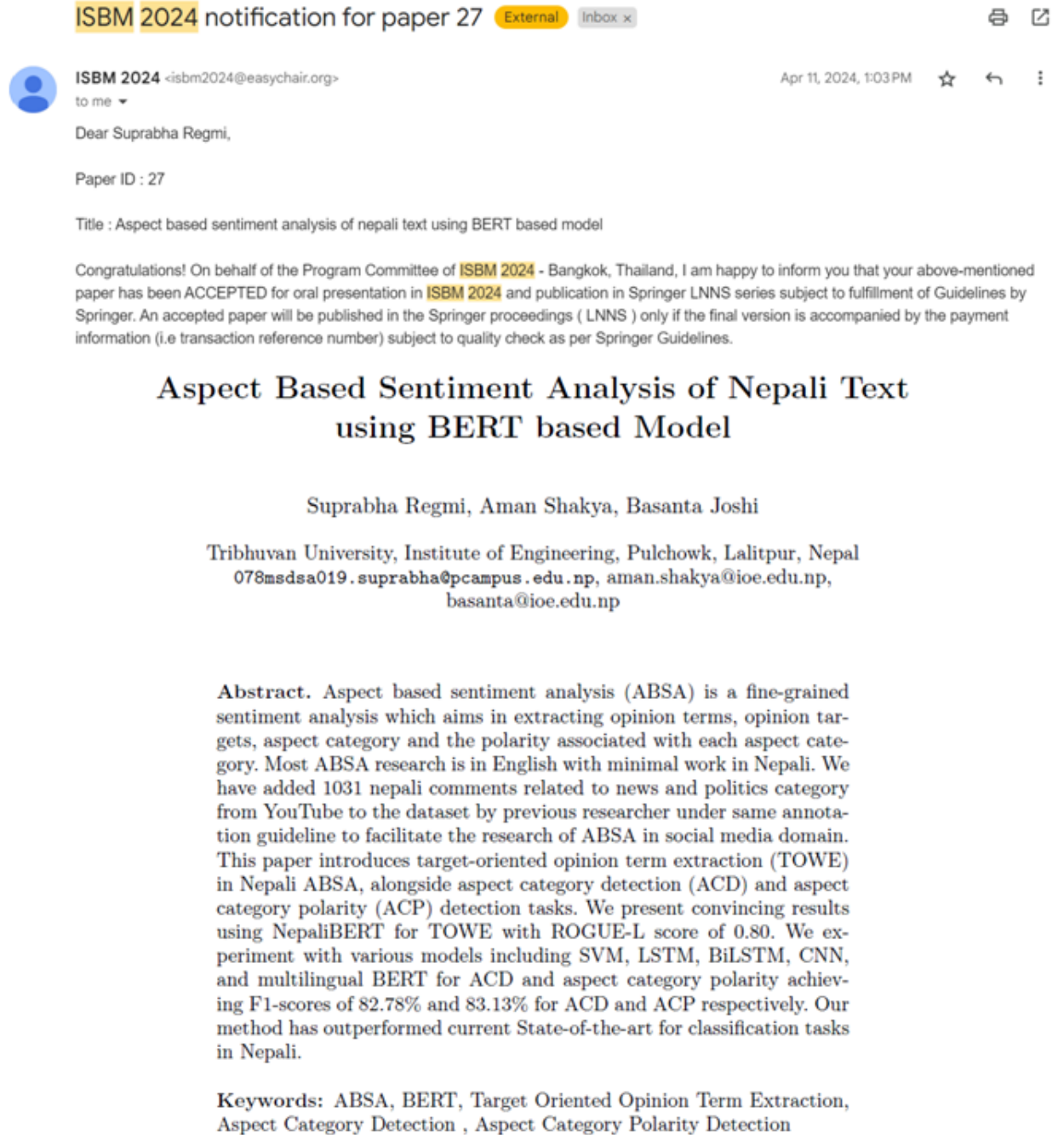Keywords: ABSA, BERT, Target Oriented Opinion Term Extraction, Aspect Category Detection , Aspect Category Polarity Detection

Figure 6.11: Accepted paper along with Abstract