



**ARTICLE**

# A Semi-Supervised Approach for Aspect Category Detection and Aspect Term Extraction from Opinionated Text

Bishrul Haq<sup>1</sup>, Sher Muhammad Daudpota<sup>1</sup>, Ali Shariq Imran<sup>2</sup>, Zenun Kastrati<sup>3,\*</sup> and Waheed Noor<sup>4</sup>

<sup>1</sup>Department of Computer Science, Sukkur IBA University, Sukkur, 65200, Pakistan

<sup>2</sup>Department of Computer Science, Norwegian University of Science and Technology (NTNU), Gjøvik, 2815, Norway

<sup>3</sup>Department of Informatics, Linnaeus University, Växjö, 35195, Sweden

<sup>4</sup>Department of Computer Science & IT, University of Balochistan, Quetta, 87300, Pakistan

\*Corresponding Author: Zenun Kastrati. Email: zenun.kastrati@lnu.se

Received: 26 March 2023 Accepted: 13 June 2023 Published: 31 October 2023

## ABSTRACT

The Internet has become one of the significant sources for sharing information and expressing users' opinions about products and their interests with the associated aspects. It is essential to learn about product reviews; however, to react to such reviews, extracting aspects of the entity to which these reviews belong is equally important. Aspect-based Sentiment Analysis (ABSA) refers to aspects extracted from an opinionated text. The literature proposes different approaches for ABSA; however, most research is focused on supervised approaches, which require labeled datasets with manual sentiment polarity labeling and aspect tagging. This study proposes a semi-supervised approach with minimal human supervision to extract aspect terms by detecting the aspect categories. Hence, the study deals with two main sub-tasks in ABSA, named Aspect Category Detection (ACD) and Aspect Term Extraction (ATE). In the first sub-task, aspects categories are extracted using topic modeling and filtered by an oracle further, and it is fed to zero-shot learning as the prompts and the augmented text. The predicted categories are the input to find similar phrases curated with extracting meaningful phrases (e.g., Nouns, Proper Nouns, NER (Named Entity Recognition) entities) to detect the aspect terms. The study sets a baseline accuracy for two main sub-tasks in ABSA on the Multi-Aspect Multi-Sentiment (MAMS) dataset along with SemEval-2014 Task 4 sub-task 1 to show that the proposed approach helps detect aspect terms via aspect categories.

## KEYWORDS

Natural language processing; sentiment analysis; aspect-based sentiment analysis; topic-modeling; POS tagging; zero-shot learning

## 1 Introduction

Technological advancement has enabled countless ways to write, read, and share opinions, wherein web technologies play a significant role in everyday life. It has paved the way for the swift growth of many communication mediums such as social networks, online forums, product review sites, blogging sites, website reviews, etc. [1–4]. These mediums are primarily used to express the opinions or sentiment embedded within the text, which can be identified with the help of sentiment analysis, a sub-field of



Natural Language Processing (NLP). The sentiment can be represented with polarity being positive, negative, or neutral [5,6] or into several point scales ranging from very good to very bad [7,8].

Sentiment analysis is considered to be a dominant part in many areas like Machine Learning, Deep Learning, and Rule-based approaches to categorize text into their relevant classes (e.g., positive, negative, neutral) [9,10]. Although sentiment analysis is an important task, whereas the traditional approaches are mainly dependent on detecting sentiments based on the whole text however, it is more effective when coupled with Aspect Based Sentiment Analysis (ABSA) to extract the correct sentiments based on the aspects, which refers to the extraction of aspects from the opinionated text. ABSA can be broken down into Aspect Term Extraction and Aspect Polarity Detection [11,12].

The present literature for ABSA is dominantly centered around sentiment detection on opinionated text. Still, it is also imperative to extract the aspect categories and terms about which the sentiment has been written [1]. The literature proposed many techniques for extracting different utterances from the text, primarily focused on supervised learning-based approaches to extract aspects from the text. The aspect from the text can be further identified under three main categories: lexicon-based, machine learning, and hybrid-based approaches [13].

### *1.1 Approaches in ABSA*

Lexicon-based approaches are widely used across many studies under ABSA in which SentiWordNet is generally used to allocate lexicon scores to each word in the sentence [14,15]. Moreover, Term Frequency-Inverse Document Frequency (TF-IDF), Part-of-speech (POS) tagging along with aspect extraction have been applied to differentiate the characteristics or highlight concluding words from the sentences to extract aspects that lead to good performance in aspect extraction [16,17]. POS tagging refers to labeling each word in a sentence with its relevant part of speech, such as noun, adverb, verb, etc. [18,19], a vital pre-processing step in NLP. It represents the linguistic rules, random patterns, or a combination of both [20].

Machine learning is another area widely seen in the ABSA domain to find the aspects of an opinionated text. It makes use of two main sub-categories in machine learning as supervised (including Support Vector Machine (SVM), Naive Bayes (NB), Logistic Regression (LR), and K-Nearest Neighbor (KNN) algorithms) and unsupervised learning [21–23]. The recent advancements in ABSA have also seen many approaches based on deep neural networks. Several algorithms are employed to find the best-performing classifier on ABSA sub-tasks, including Fully Connected Deep Neural Networks (FC-DNN), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM). In recent studies, Transformer based approaches [24–27] are also used to produce better accuracy by using an attention mechanism [28].

The combination of lexicon and machine learning-based approaches are rising in ABSA, which uses hybrid approaches to detect the aspects. Several studies have applied lexicon for sentiment identification [29], and Machine Learning and Deep Learning techniques [4] to extract the aspects, whereas both supervised and unsupervised techniques are used alongside topic modeling [30] and Named Entity Recognition [31]. In recent years, many state-of-the-art Transfer Learning methods have been proposed along with zero-shot learning [32], XLM-R [33], BERT [34], etc., in the hybrid domain under ABSA.

Moreover, prompt-based approaches like Few Shots or Zero Shot have recently been applied in ABSA [35], whereas prompts are given to the pre-trained model to complete the task. The usage of pre-trained models for Zero-Shot Learning (ZSL) has significantly improved ABSA due to its ability to understand linguistic features and possess better text encoders [36].

There are several challenges found in the present state of ABSA, whereas most studies are centered around supervised approaches that demand labeled datasets. Moreover, unsupervised approaches mainly depend on broader linguistic patterns, seed words, etc. Hence it is more challenging for the model to extract aspects regardless of the domain [1]. Hence, this study formulates a robust process to specifically address present challenges via a semi-supervised approach which helps the oracle (human) only to choose categories suggested via the study, which then detects categories, and further, the categories are used to extract the aspect terms.

### *1.2 Study Objective & Research Questions*

The study helps to formulate the validity and the enhancement of ABSA along with a domain-independent semi-supervised approach. The focused questions based on the study objectives serve to detect aspect categories and aspect terms from users' textual reviews with minimal human supervision. The study addresses the following research questions (RQ).

- **RQ1:** Can a semi-supervised approach facilitate the extraction of aspect terms and categories from text reviews with minimal human supervision?
- **RQ2:** How far domain independent semi-supervised techniques can be exploited to extract meaningful aspect categories and terms from the opinionated text?
- **RQ3:** What is the performance of the models in extracting aspect categories and terms from the opinionated text?

### *1.3 Study Contribution*

The study majorly contributes in the following ways:

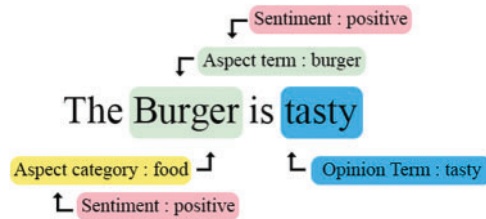
- A domain-independent semi-supervised approach to detect aspect terms and categories from the opinionated text.
- Detect aspect terms based on aspect categories by employing a novel two-step approach.
- Report baseline accuracy on the MAMS dataset for ACD and ATE subtasks.

This study employs two main steps to detect aspect terms and categories. Categories are derived by topic modeling, which serves as the input to the second approach for extracting aspect terms. Before the approaches, the study implements pre-processing steps with several techniques, such as removing stop words and special characters, filtering texts based on English, and then applying tokenization on the textual dataset. The study recognizes the main aspect categories in the **first approach** by employing topic modeling techniques, mainly using LDA with Trigram and TF-IDF settings and BERTopic Modeling. The topic models aid in extracting meaningful topics and then are used as the prompts for zero-shot learning with augmented texts, whereas contextual augmentation helps extract the aspect categories. In the **second approach**, the predicted categories from the first approach are used as input to find similar phrases from the extracted phrases (e.g., Compound Nouns, Nouns, Proper Nouns, NER entities) in order to evaluate the performance of aspect terms extraction and validate our approach.

The remaining part of the paper is structured as follows. The related works section describes similar works carried out within ABSA. The complete methodology of the proposed approach is presented in the methodology section. The study results are discussed in the results and discussion section. Finally, the conclusion section concludes the paper.

## 2 Related Works

Natural Language Processing (NLP) is one of the areas that have been applied in various domains to produce much-needed applications [37,38]. Sentiment analysis is a subbranch of NLP [39] in which aspects are identified. ABSA includes several essential sentiment elements such as aspect term, opinion term, polarity, and category [1,40]. Further, ABSA is recognized into several sub-tasks, as shown in Fig. 1, including Aspect Category Detection (ACD), Aspect Term Extraction (ATE), Opinion Term Extraction (OTE), and Aspect Sentiment Classification (ASC) [41].



**Figure 1:** ABSA sub task

The subtasks in ABSA define the detection of an aspect or an entity. For example, in the sentence ‘**The Burger is tasty**’, the aspect term is defined as *Burger*, which serves the subtask ATE. The term falls under the category of *food*, which the category ACD represents. The OTE specifies the opinion word *tasty* [41]. The sentiment resonated from the opinion or term accompanied by a sentiment polarity is addressed as ASC.

The literature refers to ASC with many other terms, including ACSA (Aspect Category Sentiment Analysis) [26], ATSA (Aspect Term Sentiment Analysis) [26], ALSA (Aspect Level Sentiment Classification) [42], ABSC (Aspect Based Sentiment Classification) [43]. In comparison, a limited number of studies have been reported on ACD and ATE subtasks. Literature on ABSA reports several techniques, such as Lexicon-based, Machine Learning, and Hybrid-based approaches [44], whereas only a few studies have been reported based on unsupervised approaches targeting ACD and ATE.

Garc’ia-Pablos et al. [45] used an unsupervised approach that takes a single seed word for each aspect and their sentiment with several unsupervised learning approaches. It uses Latent Dirichlet Allocation (LDA) as topic modeling to assess performance on SemEval 2016 dataset. A similar study by [46] took seed words for each category while focusing on the subtasks of ATE using Guided LDA. It also uses a BERT-based semantic filter on the restaurant domain under SemEval 2014, 2015, and 2016. The study reported significant performance improvement compared to other supervised and unsupervised approaches.

Purpura et al. [47] proposed a weekly supervised setting to classify the aspects and sentiments. The study uses Non-negative Matrix Factorization (NMF) topic Modeling coupled with short seed lists based on each aspect. The study compares the performance of SemEval 2015 and 2016 datasets against various weekly and semi-supervised models. Most studies are centered around human supervision based on seed words with categories or aspects for each category to perform ATE, ACD, or ASC under unsupervised settings. We have not found any study related to MAMS dataset for the subtasks ACD and ATE as per our knowledge.

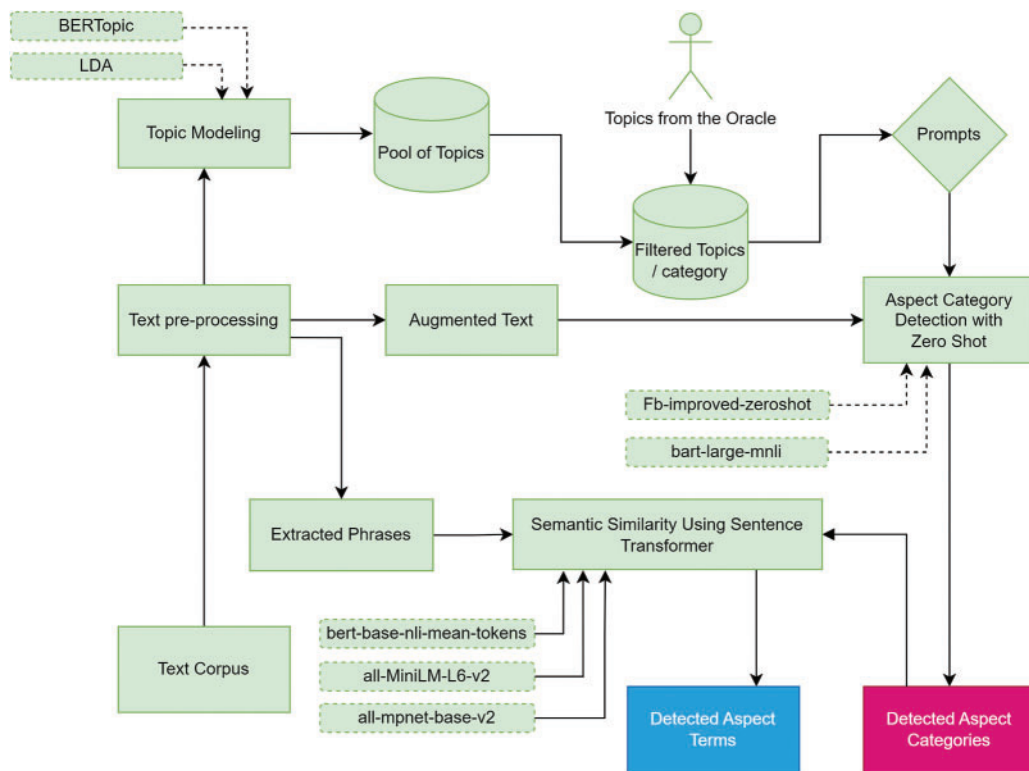
Zero-Shot Learning (ZSL) can predict unseen classes, which has been commonly used in computer vision [48]. Kumar et al. [32] have implemented a three-step approach for detecting the aspects and their sentiments. They use Bidirectional Encoder Representations from Transformers (BERT) and construct vocabularies by using part-of-speech (POS) tagging to label the text in the subset by employing a

Deep Neural Network (DNN). Another study [49] has applied for a zero-shot transfer with few shots learning. It focuses on three ABSA tasks: Aspect Extraction, End-to-End Aspect-Based Sentiment Analysis, and Aspect Sentiment Classification. The study is based on the post-training of several pre-trained models. Similarly, another study [50] has applied zero-shot learning to classify sentiments with enhanced settings under SemEval 2014 dataset with few shot settings. The study has outperformed other supervised learning approaches.

Data Augmentation in NLP is vital in generating new data, which is especially useful when data is scarce. It adds more contextual information by increasing the data size [9]. Contextual augmentation is a novel method applied in many recent studies under NLP. It uses BERT models to stochastically change words with different words as per the contextual surroundings [51]. Contextual augmentation tends to give a more robust output than EDA (Easy Data Augmentation) [52]. To the best of our knowledge, no study has been conducted under contextual augmentation for ABSA.

### 3 Methodology

The methodology presented in Fig. 2 explains the overall aspect detection and extraction process from the opinionated text. Initially, the study commences with the pre-processing step. Further, the study continues with the two main sub-tasks: Aspect Category Detection (ACD) and Aspect Term Extraction (ATE).



**Figure 2:** Overall methodology of the study

### 3.1 Text Corpus

The study uses the MAMS (Multi Aspect Multi-Sentiment) dataset consisting of restaurant reviews, a challenge dataset for Aspect Term Sentiment Analysis (ATSA) and Aspect Category Sentiment Analysis (ACSA) [53] and also SemEval 2014 dataset subtask 1 addresses the Aspect Term Extraction (ATE). The MAMS dataset holds separate XML files, whereas this study uses train XML files, excluding test and validation XML files for each sub-task.

Moreover, we have found that MAMS has not been applied in ACD or ATE sub-tasks. The MAMS dataset has multiple aspects and longer text compared to SemEval 2016 Restaurant domain [54]. Hence, this study sets a baseline score for the subtask via extracting the aspect categories and terms from train XML files, as shown in Fig. 3. The SemEval 2014 dataset subtask 1 is based on the laptop domain and has been applied in many studies for ATE hence, we are comparing the performance of our approach against the reported studies.



**Figure 3:** Converting XML files into CSV files

#### 3.1.1 ATSA and ACSA in MAMS Dataset

The ATSA in train XML files contains more than 4000 sentences with aspect terms and their sentiment polarities. Similarly, ACSA in train XML file contains more than 3000 sentences with eight aspect categories (e.g., ‘staff’, ‘menu’, ‘food’, ‘price’, ‘service’, ‘miscellaneous’, ‘ambiance’, ‘place’) with their polarities. Sample data capture of ATSA and ACSA are given below:

//Aspect Category Detection

<sentence>

<text>I like the smaller portion size for dinner.</text>

<aspectCategories>

<aspectCategory category="miscellaneous" polarity="negative"/>

<aspectCategory category="food" polarity="neutral"/>

</aspectCategories>

</sentence>

//Aspect Term Extraction

<sentence>

<text>Food is pretty good but the service is horrific.</text>

<aspectTerms>

<aspectTerm from="0" polarity="positive" term="Food" to="4"/>

<aspectTerm from="28" polarity="negative" term="service" to="35"/>

</aspectTerms>

</sentence>

### 3.1.2 ATE and ATSC in SemEval 2014 Dataset Subtask 1

SemEval 2014 consists of manually annotated reviews of laptop and restaurant domains, whereas this study explores Laptop reviews which include two subtasks as ATE and ATSC. The dataset holds test and training sets in separate XML files. This study uses training data that holds more than 3000 sentences. The capture of the training sample is given below:

```
//Aspect Term Extraction & Aspect Term Sentiment Analysis
<sentences>
  <sentence id="2339">
    <text>I charge it at night and skip taking the cord with me because of the good battery
life.</text>
    <aspectTerms>
      <aspectTerm term="cord" polarity="neutral" from="41" to="45"/>
      <aspectTerm term="battery life" polarity="positive" from="74" to="86"/>
    </aspectTerms>
  </sentence>
</sentences>
```

### 3.1.3 Text Pre-Processing

The XML files were converted into data frames, and several steps were followed to pre-process the data, such as removing stop words and special characters, filtering texts based on English, and then applying tokenization to the textual dataset, as shown in the example below:

- **Text:** Food is pretty good but the service is horrific
- **Pre-processed text:** Food pretty good service horrific
- **Tokenized text:** ['food', 'pretty', 'good', 'service', 'horrific']

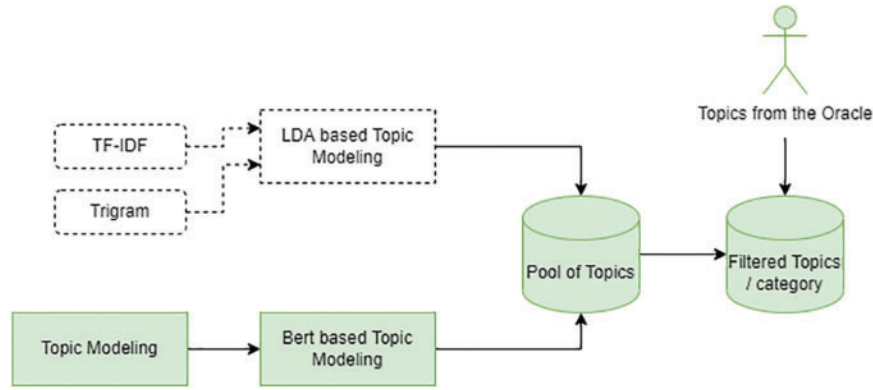
## 3.2 Topic Modeling

Topic Modeling is an analytical technique that helps to filter and determine the correlation between frequent and relevant topics, which helps to sort the most common keywords from the text [55]. It has been widely applied in many studies to gather meaningful words within the text corpus [56]. This study employed the most frequent topic modeling methods, Bidirectional Encoder Representations from Transformers (BERT) and Latent Dirichlet Allocation (LDA), to extract the most relevant topics as shown in Fig. 4. Table 1 shows the topics gathered from each sub-task and the filtered topics via the aid of Oracle.

### 3.2.1 LDA Topic Modeling

Latent Dirichlet Allocation (LDA) is one of the topic modeling methods frequently applied in ABSA to extract topics by estimating a probabilistic word model over the corpus. LDA uses the Dirichlet process to model latent topics, representing a topic by the distribution of words occurring in the corpus. A given corpus represented by  $X$  consists of  $L$  texts where each text represented by  $x$  consists of  $N_x$  words ( $x \in 1, \dots, L$ ). The probability distribution of  $X$  is computed with the following Equation [57,58] whereas this study applies LDA with TF-IDF and Trigram Linguistic features to extract the topics.

$$p(X | \alpha, \beta) = \prod_{d=1}^L \int p(\theta_x | \alpha) \left( \prod_{n=1}^{N_x} \sum_{z_{xn}} p(Z_{xn} | \theta_d) p(w_{xn} | z_{xn}, \beta) \right) x \theta_x \tag{1}$$



**Figure 4:** Topic modeling

**Table 1:** Filtered topics through topic modeling with the involvement of oracle

BERTopic	<p><b>Filtered topics (NOUN AND PROPER NOUN)</b>                      ‘saturday’, ‘ambience’, ‘waitress’, ‘manager’, ‘hostess’, ‘sushi’, ‘price’, ‘time’, ‘waiter’, ‘reservation’, ‘minutes’, ‘bartender’, ‘beers’, ‘reservations’, ‘space’, ‘service’, ‘fun’, ‘server’, ‘place’, ‘portions’, ‘staff’, ‘night’, ‘food’, ‘list’, ‘scene’, ‘table’, ‘tax’, ‘wine’, ‘bartenders’, ‘bar’, ‘ones’, ‘pizza’, ‘atmosphere’, ‘fixe’, ‘noise’, ‘décor’, ‘menu’, ‘chefs’, ‘dessert’, ‘order’, ‘coffee’, ‘chef’, ‘dinner’, ‘mind’, ‘water’, ‘room’, ‘prices’, ‘steaks’, ‘prix’, ‘steak’, ‘friends’, ‘selection’</p> <p><b>Selected topics</b>                      ‘staff’, ‘menu’, ‘food’, ‘price’, ‘service’, ‘ambiance’, ‘place’, ‘miscellaneous’</p>
LDA + Trigram	<p><b>Filtered topics (NOUN AND PROPER NOUN)</b>                      ‘server’, ‘waitress’, ‘dessert’, ‘sauce’, ‘sushi’, ‘table’, ‘reservation’, ‘food’, ‘drinks’, ‘bar’, ‘price’, ‘dining’, ‘wine’, ‘lunch’, ‘manager’, ‘fish’, ‘bill’, ‘entrée’, ‘time’, ‘cheese’, ‘staff’, ‘waiter’, ‘french’, ‘items’, ‘menu’, ‘décor’, ‘appetizers’, ‘dinner’, ‘place’, ‘service’, ‘restaurant’, ‘steak’, ‘room’, ‘portions’, ‘area’, ‘plate’, ‘minutes’, ‘order’</p> <p><b>Selected topics</b>                      ‘staff’, ‘menu’, ‘food’, ‘price’, ‘service’, ‘ambiance’, ‘place’, ‘miscellaneous’</p>
LDA + TF-IDF	<p><b>Filtered topics (NOUN AND PROPER NOUN)</b>                      ‘staff’, ‘sushi’, ‘appetizers’, ‘restaurant’, ‘manager’, ‘wine’, ‘lot’, ‘order’, ‘chicken’, ‘menu’, ‘bottle’, ‘bill’, ‘dinner’, ‘service’, ‘server’, ‘drinks’, ‘dessert’, ‘bar’, ‘time’, ‘minutes’, ‘reservation’, ‘sauce’, ‘price’, ‘place’, ‘chef’, ‘noodles’, ‘portions’, ‘waitress’, ‘food’, ‘atmosphere’, ‘table’, ‘décor’, ‘waiter’, ‘cheese’, ‘plate’, ‘restaurants’</p> <p><b>Selected topics</b>                      ‘staff’, ‘menu’, ‘food’, ‘price’, ‘service’, ‘ambiance’, ‘place’, ‘miscellaneous’</p>



### 3.2.2 BERTopic Modeling

Recent development in Pre-trained Language Models (PLMs) has played an essential role in topic modeling. BERTopic Modeling uses word embedding along with class-based TF-IDF to create dense clusters with Uniform Manifold Approximation and Projection (UMAP) to lower the dimension of embedding, which yields promising results [59]. The word with frequencies  $l$  is used for each class denoted with  $m$  which is then divided by the number of total words shown as  $x$ . The number of words as average per class represented with  $o$  is divided by the total frequency of word  $l$  in all  $q$  classes as shown below:

$$c - TF - IDF = \frac{l_m}{x_m} \times \log \frac{o}{\sum_j l_j} \quad (2)$$

### 3.3 Aspect Category Detection

ACD is one of the sub-tasks in ABSA, which categorizes the text based on its features. The study assigns categories for the text from the filtered topics, used as the prompts for zero-shot learning (ZSL). Contextual augmentation is used to augment the text, which serves as the input for ZSL. Fig. 5 shows complete process of ACD.

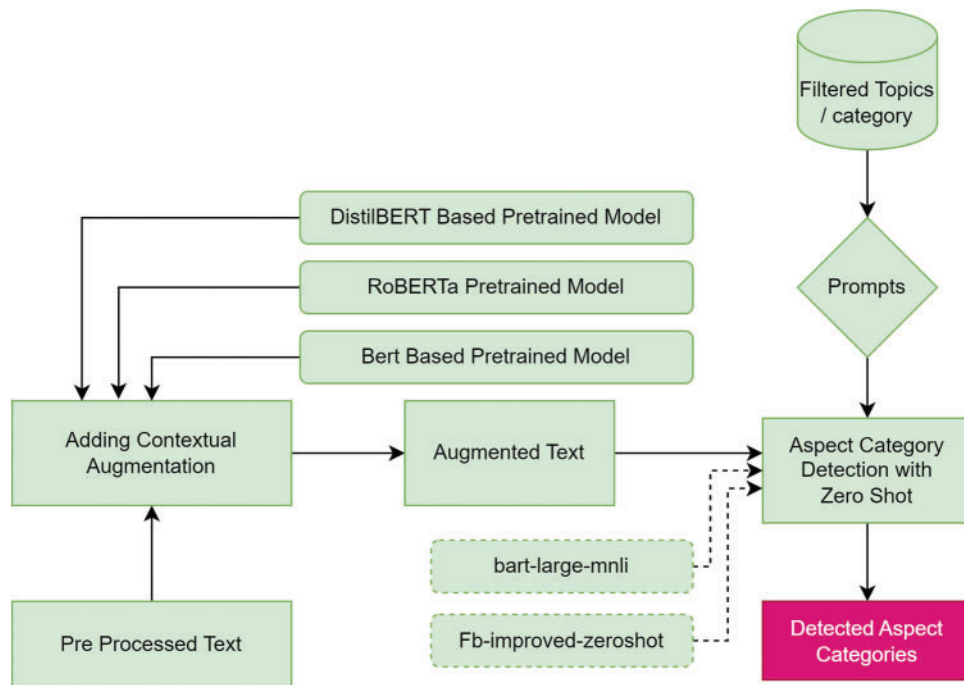


Figure 5: Aspect category detection

#### 3.3.1 Contextual Augmentation

Data augmentation (DA) is used to avoid overfitting the data to build more robust models [60]. Due to its success in computer vision, DA has also been adapted to many NLP tasks [61]. Hence, this study applies contextual augmentation via randomly replacing words with different predictions based on substitutions and insertion techniques [51]. This study uses several pre-trained models such as **roberta-base**, **bert-base-uncased**, **distilbert-base-uncased** against the text, yielding the results for augmented text. A sample is shown in Table 2 (for ZSL with ‘**bart-large-mnli**’ model).

**Table 2:** Contextual augmentation with different pre-trained models

Algorithm	Method	Raw text	Augmented text
RoBERTa [62]	roberta-base	Though the service might be a little slow the waitresses are very friendly	<b>Insert:</b> Though initially, the service might be a little bit slow the two waitresses are very friendly <b>Substitution:</b> Though the service can be slow the waitresses are very friendly
BERT [63]	bert-base-uncased	Though the service might be a little slow the waitresses are very friendly	<b>Insert:</b> Even though the food service itself might be a little slow the two waitresses are very friendly <b>Substitution:</b> Although the food might be slow the waitresses are very friendly
DistilBERT [64]	distilbert-base-uncased	Though the service might be a little slow the waitresses are very friendly	<b>Insert:</b> Though the service might little slow down the waitresses who are very user friendly <b>Substitution:</b> Though the food might little help the waitresses are very busy

### 3.3.2 Zero-Shot Learning (ZSL)

Recently, several studies have applied ZSL on ABSA, specifically on Sentiment Classification [32,33]. ZSL is represented with  $X = \{C_x^b | b = 1, \dots, N_x\}$  as the the set of seen labels where  $C_x^b$  denotes a seen label and the set of unseen labels denoted by  $Y = \{C_y^b | b = 1, \dots, N_y\}$  where  $C_y^b$  is an unseen label. Whereas  $X \cap Y = \phi$  and  $L$  is a  $K$ -dimensional feature space such that  $K \in \mathbb{R}^K$  where  $\mathbb{R}$  is real. Hence,  $K^{tr} = \{(n_i^{tr}, m_i^{tr}) \in L \times X\}_{i=1}^{N^{tr}}$  denotes the set that is part of seen labeled; for each labeled instance  $(n_i^{tr}, m_i^{tr})$ ,  $n_i^{tr}$  is the instance from the feature space, where  $m_i^{tr}$  represents as the corresponding label.  $N^{te} = \{n_i^{te} \in L\}_{i=1}^{N^{te}}$  denotes as the set of test instances where each  $n_i^{te}$  is represents a test instance in the feature space.  $M^{te} = \{m_i^{te} \in Y\}_{i=1}^{N^{te}}$  denotes the labels for  $N^{te}$  that are to be predicted [48].

This study takes the prompts for ZSL via topic modeling, and the input text was given via contextual augmentation by augmenting the data. We have tested two different zero shot transformer models such as ‘**bart-large-mnli**’, ‘**Fb-improved-zeroshot**’ with insert and substitution settings available at Hugging Face library<sup>1</sup>. Moreover, this study sets a threshold of 0.93 with some extended selection, as shown in Algorithm 1.

- The bart-large-mnli model trained on MultiNLI (MNLI) dataset comprises spoken and written text based on ten sources and genres [59].

<sup>1</sup><https://huggingface.co/models>

- The Fb-improved-zeroshot model is trained using the bart-large-mnli for English and German to classify academic search logs [60].

---

**Algorithm 1:** Setting Threshold for ZSL Model
 

---

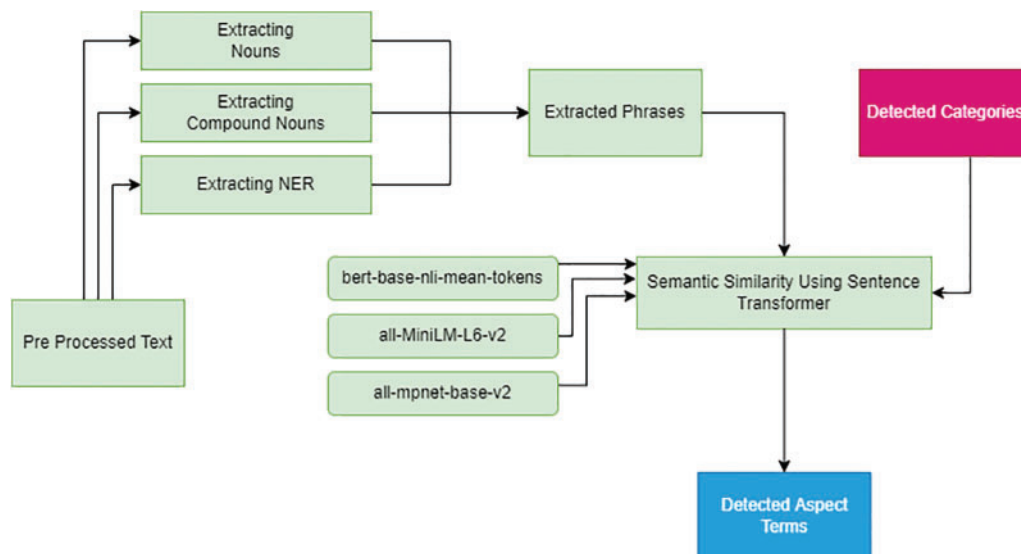
```

1: assigned ← selected categories
2: category ← Predicted Categories from ZSL
3: if category.scores > 0.93 then
4:   Append categories to assigned
5: else
6:   if assigned is empty then
7:     Get the MAX Score and round it by 1 Decimal Place
8:     if MAX Score equals to other scores with 1 Decimal Place then
9:       Append categories to assigned
10:    end if
11:   end if
12: end if
  
```

---

### 3.4 Aspect Term Extraction

Aspect Term Extraction aims to extract aspect terms in the opinionated text. This study helps extract a set of meaningful phrases with several techniques from each text. The similarity of the phrases is then measured against the categories predicted by ZSL with semantic similarity using the transformer model. Fig. 6 shows the complete process of ATE.



**Figure 6:** Aspect term extraction

#### 3.4.1 Extracted Phrases

This study extracts the phrases by employing several techniques. Initially, COMPOUND NOUNS, NOUNS, and PRONOUNS are extracted with POS tagging. Secondly, NER annotations are extracted with applicable entities (e.g., organizations, locations, etc.) using spaCy. Finally, the

phrases which contain ADJECTIVE are removed from the set of phrases, as shown in Table 3. The steps for phrase extraction are shown in Algorithm 2.

**Table 3:** Extracted phrases for each text ranging from L1 to L3

Label	Text	Aspect terms	Extracted phrases
L1	The decor is not special at all but their food and amazing prices make up for it.	‘prices’, ‘food’, ‘décor’	‘prices’, ‘décor’, ‘food’
L2	When tables opened up, the manager sat another party before us.	‘manager’, ‘tables’	‘party’, ‘tables’, ‘manager’
L3	Though the menu includes some unorthodox offerings (a peanut butter roll, for instance), the classics are pure and great we’ve never had better sushi anywhere, including Japan.	‘sushi’, ‘classics’, ‘peanut butter roll’, ‘menu’	‘butter roll’, ‘instance’, ‘japan’, ‘menu’, ‘peanut butter’, ‘offerings’, ‘sushi’

**Algorithm 2:** Process of Extracting the Phrases

```

1: tokens ← pre processed texts
2: Set filtered tokens
3: for each token ∈ tokens do
4: if token is NOUN OR COMPOUND NOUN OR PRONOUNS OR belongs to ANY NER Entity
   AND NOT ADJECTIVE then Append token to filtered tokens
5:   end if
6: end for

```

### 3.4.2 Semantic Similarity

The study compares the similarity of phrases against the categories predicted by ZSL by converting them into embedded vectors in which the semantically similar phrases corresponding to each sentence are selected. This study has tested the approach with three pre-trained Sentence-BERT NLI models [65], including **all-mpnet-base-v2 (B1)**, **bert-base-nli-mean-tokens (B2)**, **all-MiniLM-L6-v2 (B3)**. The models yield different results based on the threshold. This study has adapted in Algorithm 3 for the phrases as per the categories for L1, as shown in Table 4.

**Table 4:** Semantic similarity for L1 text represented with the pre-trained models

Corresponding label from extracted phrases	Pre-trained BERT model	Predicted aspects
L1	bert-base-nli-mean-tokens (B2)	‘food’, ‘decor’, ‘prices’
	all-mpnet-base-v2 (B1)	‘food’, ‘prices’
	all-MiniLM-L6-v2 (B3)	‘food’, ‘prices’

---

**Algorithm 3:** Selecting the Aspect Terms

---

```

1: Selected Aspects Terms
2: phrases ← phrases with semantic similarity
3: for phrase in phrases do
4:   if phrase.score > max(phrase.score) - 1.0 then
5:     Append phrase to Selected Aspects Terms
6:   end if
7: end for

```

---

## 4 Results & Discussion

The study mainly focuses on two main subtasks in ABSA, including ACD (Aspect Category Detection) and ATE (Aspect Term Extraction). ACD serves as the input to ATE to extract the aspect terms, as shown in Fig. 2. The study begins with detecting aspect categories; further, the categories are served as the input to detect the aspect terms.

### 4.1 Dataset and Their Sub Tasks

The study has chosen two ABSA datasets to evaluate the extraction of including ACD (Aspect Category Detection) and ATE (Aspect Term Extraction), whereas Multi-Aspect Multi-Sentiment (MAMS) dataset holds two separate files for each subtask hence, we performed ACD separately for each XML file within the datasets. To the best of our knowledge, the MAMS challenge dataset does not contain any reported accuracy for ACD or ATE subtasks. Hence, this study establishes baseline results for both sub-tasks.

The SemEval-2014 Task 4 subtask 1 restaurant dataset mainly addresses ATE and does not include ACD. Therefore, we tested topic modeling under the MAMS dataset for ACD. Moreover, for ATE we evaluated the performance of the SemEval 2014 restaurant dataset against other studies along with the MAMS ATE subtask. As per our knowledge, the MAMS challenge dataset does not contain any reported accuracy for ACD or ATE subtasks hence, this study sets a baseline for both the tasks.

### 4.2 Performance Metrics

The study uses the three most commonly used performance matrices in NLP and text mining, precision, recall, and F1 score, to evaluate the performance of ACD and ATE subtasks. The aspect terms or aspect categories are marked as correct only if they match the original values. The performance metrics for precision (P), recall (R), and F1 score (F1) are calculated as follows:

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2PR}{P + R} \quad (5)$$

TP, TN, FP, and FN indicate the total number of true positives, true negatives, false positives, and false negatives, respectively.

### 4.3 Aspect Category Detection

The Aspect Category Detection relies on topic modeling to classify the filtered topics as prompts with Zero-Shot Learning (ZSL) along with augmented texts.

#### 4.3.1 Topic Modeling

The study has seen notable differences in two of the topic models employed. We have chosen MAMS ACSA as the benchmark for ACD to select the topics, whereas we have compared the original categories generated within the dataset (e.g., *staff*, *menu*, *food*, *price*, *service*, *ambiance*, *place*, *miscellaneous*). The BERTopic model has extracted more meaningful topics, which serve as the filtered categories. The topics identified with the BERTopic model have seven categories within the pool of topics.

The study tested two linguistic features for LDA, including Trigram and TF-IDF. We observed that 39, 37 topics for LDA with Trigram and TF-IDF were selected via choosing the highest probability emitted by the topics, and further, the topics are filtered with NOUNS and PROPER NOUNS from the total number of topics through LDA, whereas the same steps are followed in BERTopic. Both settings failed to extract the topic *ambiance* as shown in Table 1. Moreover, the topics extracted by the BERTopic model are more meaningful in choosing the categories. The topics identified with blue are the ones that are given by the oracle hence these selected topics are then served as the categories.

#### 4.3.2 Classifying the Categories

We use ZSL along with two pre-trained models, including **'Fb-improved-zeroshot' (X1)**, **'bart-large-mnli' (X2)** to evaluate the performance, and then the filtered categories are turned into prompts along with contextual augmentation, as it includes more significant words within the text to predict the categories accurately. However, ZSL struggles to detect the category *"miscellaneous"* because it linguistically did not make sense in many texts. We have set the threshold value of 0.93 with several settings to filter the detected categories for each text.

We further we have employed different contextual augmentation models, including **roberta-base (P1)**, **bert-base-uncased (P2)**, **distilbert-base-uncased (P3)**. We found that X1 along with P2 under substitution performs better compared to other settings though a somewhat similar performance is seen on X1 along with P1 under substitution. Whereas X1 with P1 to P3 under insertion, substitution performs better than X2 along with other contextual models. When comparing X1 and X2 with ZSL, we found that X1 has performed well, as shown in Figs. 7 and 8. Similar F1 scores are found in X1, and X2 with contextual augmentation models ranging from 57% to 61%, whereas X1 with P2 under F1 score has a slight improvement with a score of 61.23%. P3 under X1 or X2 has the lowest performance compared to other contextual augmentations. Insertion and Substitution under X1 perform comparatively similar to X2, whereas we noticed X2 has a slight difference in F1 scores comparing insertion and substitution, as shown in Table 5.

### 4.4 Aspect Term Extraction

For the Aspect Term Extraction, the pre-processed text is used to extract COMPOUND NOUN, NOUN, and PROPER NOUN via part-of-speech (POS) tagger along with the specified entities in a body or bodies of texts via several Named Entity Recognition (NER) models excluding adjectives. The phrases are compared with semantic similarity models, including **all-mpnet-base-v2 (B1)**, **bert-base-nli-mean-tokens (B2)**, **all-MiniLM-L6-v2 (B3)** to select the most similar phrases. Hence, based on the performance measures emitted with the different models under Table 5, we have selected X1 along

with P2 as our base model settings to perform Aspect Term Extraction under ATSA in MAMS and SemEval-2014 Task 4 subtask 1.

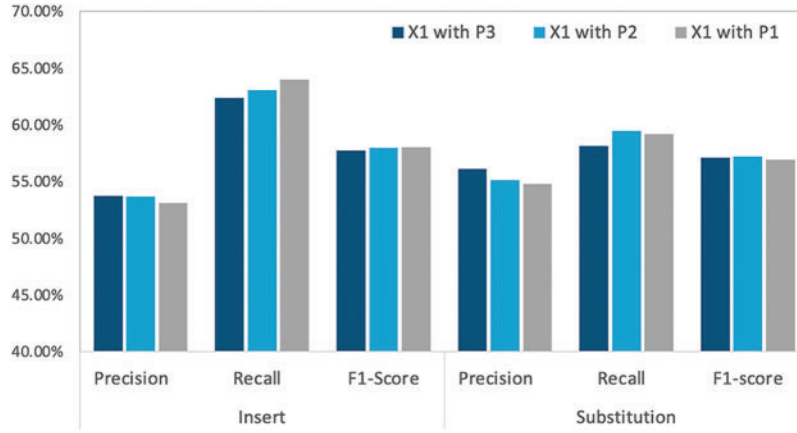


Figure 7: ZSL with Fb-improved-zeroshot

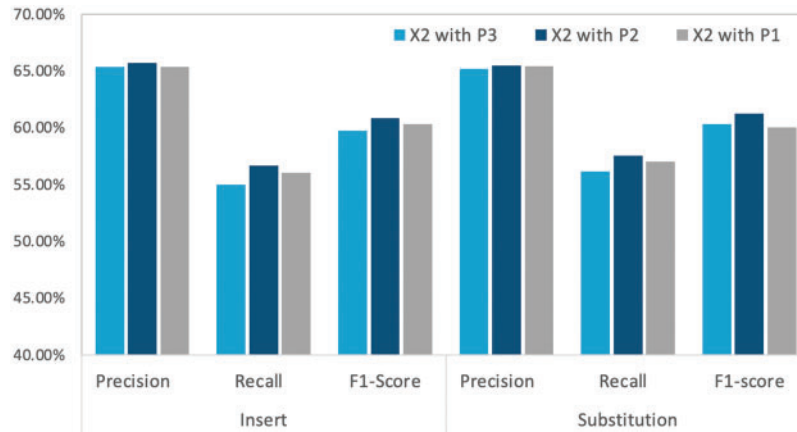


Figure 8: ZSL with bart-large-mnli

Table 5: Reported performance scores for acd with different contextual augmentation models

Label	Contextual augmentation	Insert			Substitution		
		P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
X1	P1	65.33	56.03	60.33	65.40	57.40	61.13
	P2	65.69	56.67	60.85	<b>65.44</b>	<b>57.53</b>	<b>61.23</b>
	P3	65.35	54.96	59.71	65.07	56.46	60.46
X2	P1	53.07	64.01	58.03	54.79	59.20	56.91
	P2	53.65	63.04	57.97	55.13	59.47	57.22
	P3	53.76	62.38	57.75	56.13	58.14	57.12

#### 4.4.1 ATE in MAMS

The study has used topics that are curated with the aid of topic modeling using BERTopic. It performs comparatively better than LDA with trigram or TF-IDF settings under MAMS for ACSA. We have found the pool of topics as shown in Fig. 9, and we chose eight categories/topics which describe the domain, including ‘staff’, ‘menu’, ‘food’, ‘price’, ‘service’, ‘ambiance’ and then the topics were served as the prompts to ZSL using the baseline settings of X1 along with P2. The extracted phrases in each text were compared with B1 to B3 pre-trained sentence transformer models, and we have found the following results in Table 6.



Figure 9: Pool of topics

Table 6: Reported performance scores for MAMS-ATE

MAMS-ATE			
Label	P (%)	R (%)	F1 (%)
B1	65.18	70.44	67.71
B2	<b>64.60</b>	<b>73.24</b>	<b>68.65</b>
B3	65.14	70.70	67.81

The study has found that B2 performs better compared to other pre-trained sentence transformer models. We have noticed that B1 and B2 have similar F1 scores though B2 has the highest recall compared to other models. Hence, the study sets a baseline for both MAMS ATE and ACD subtasks, as shown in Table 7.

#### 4.4.2 ATE in SemEval-2014 Task 4 Subtask 1

The study has also applied the same experiment settings with SemEval-2014 Task 4 subtask 1, addressing the ATE. The dataset focuses on the laptop domain, whereas the MAMS dataset is an extended version of the SemEval-2014 Task 4 Restaurant domain. The study extracted the most useful words, including ‘camera’, ‘display’, ‘keyboard’, ‘memory’, ‘battery’, ‘price’, ‘computer’, ‘software’ with the aid of the BERTopic model.



**Table 7:** Reported performance scores for ATE, ACD

MAMS-ACD (Baseline)			
Approach taken	P (%)	R (%)	F1 (%)
ZSL + X1 + P2 (Substitution)	<b>65.44</b>	<b>57.53</b>	<b>61.23</b>
MAMS-ATE (Baseline)			
<b>B2</b>	<b>64.60</b>	<b>73.24</b>	<b>68.65</b>

The selected categories are then served as the prompt with default settings applied in MAMS dataset under the same pre-trained sentence transformer models, yielding the following results in [Table 8](#). In contrast, the study has found that both MAMS ATSA and SemEval-2014 Task 4 subtask 1 on ATE has achieved significant performance with B2 pre-trained sentence transformer model compared to other models.

**Table 8:** Reported performance scores for SemEval-2014-ATE

MAMS-ATE			
Label	P (%)	R (%)	F1 (%)
B1	79.93	62.48	70.14
<b>B2</b>	<b>79.87</b>	<b>66.03</b>	<b>72.30</b>
B3	79.86	72.87	70.35

#### 4.5 Comparison of ATE in SemEval-2014 Task 4 Subtask 1

We have compared the performance of SemEval-2014 Task 4 subtask 1 against the other baseline models, as shown in [Table 9](#), whereas the corresponding state-of-the-art unsupervised/semi, supervised models and their performance metrics are shown as per the systems along with the methods.

**Table 9:** Results of ATE in SemEval-2014 Task 4 subtask 1

No.	Systems	Method	P (%)	R (%)	F1 (%)
01	Wei et al. 2017	Supervised	–	–	83.65
02	Wu et al. 2018	Semi/unsupervised	55.91	66.51	60.75
03	Luo et al. 2019	Supervised	–	–	80.57
04	Aggerri et al. 2019	Supervised	81.15	87.30	84.11
05	Akhtar et al. 2020	Supervised	84.11	82.62	83.36
06	Manju et al. 2022	Semi/unsupervised	75.32	86.61	80.57
07	<b>Our approach</b>	Semi/unsupervised	79.87	66.03	72.30

#### 4.5.1 Supervised Models and Their Baselines

- Xue et al. [66] used a supervised approach to extract aspects with Bi-LSTM network in which word embedding helps to transform the input words into vectors.
- Luo et al. [67] proposed a Bi-directional dependency tree with embedded representation and Bi-LSTM with CRF for tree structure and sequential features for ATE.
- Agerri et al. [68] used a perceptron algorithm with three groups of features, including orthographic features, word shape, N-gram features, and context to cluster based on unigram matching.
- Akhtar et al. [69] used CNN and Bi-LSTM to extract aspects and sentiments. The Bi-LSTM model is deployed to understand the sequential pattern of the sentence, whereas the CNN evaluates the features of aspect terms and sentiment, working along with Bi-LSTM to gain correct predictions.

#### 4.5.2 Semi/Unsupervised Models and Their Baselines

- Wu et al. [70] proposed a hybrid approach with a rule, machine learning settings focusing on three main modules. The study begins with extracting the opinion targets and aspects and then uses the domain co-relation method to filter the opinion targets. The opinion targets and aspects are then fed into a deep-gated recurrent unit (GRU) network along with a rule-based approach for prediction.
- The study by Venugopalan et al. [46] took seed words for each category while focusing on the subtask ATE using Guided LDA along with a BERT-based semantic filter on the restaurant domain

We have found that several studies have not included all the matrices rather than only reporting the F1 scores, whereas we have noticed that the F1 ranges from 60% to 85%. This study performs better compared to other supervised and unsupervised approaches. Further, we have noticed a recent study by Venugopalan et al. [46] achieved an 80.57% F1 score which primarily focused on extracting aspect terms. The study takes categories along with some seed words to extract the aspect terms, whereas this study is centered on a domain-independent approach and minimal human effort considering only filtered categories suggested with topic modeling.

## 5 Conclusion

ABSA is an essential part of NLP, extracting aspects from the opinionated text. This research formulates an approach to extract aspects based on terms via aspect categories using a semi-supervised approach which helps to combine human supervision with an unsupervised approach to extract the aspects. The study focuses on two main subtasks in ABSA, including ACD and ATE, from the opinionated text. We use ZSL to classify the categories with the prompt words gathered from topic modeling and contextual augmentation. Furthermore, the categories are compared with the extracted phrases via several techniques, which results in the proper extraction of aspect terms. The study sets a baseline accuracy for the MAMS dataset under ACD and ATE subtasks. For ACD with the P (Precision), R (Recall), F1 (F1 Score) as 65.44%, 57.53%, 61.23% along with the matrices for ATE 64.60%, 73.24%, 68.65%, respectively. Moreover, the study has achieved a comparable accuracy for SemEval-2014 Task 4 subtask 1.

This study will open new ways to address aspect extraction in different domains with the help of aspect extraction. As a future direction, the study can be enhanced with a broader linguistic pattern

to extract more phrases to support ATE, and performing post-training on models to increase the accuracies could enhance and contribute to the body of knowledge.

This work can also be extended to other local languages. Customers usually prefer to express their sentiments in their first language which may not necessarily be English. So, the proposed approach can further be extended to extract aspects from opinionated text written in Urdu, Hindi, Spanish, French, and others [71,72]. Another possible extension of the work is assessing the impact of text generation replacing augmentation. The latest text generation techniques based on LSTM and Transformer, such as [10,73] can be employed to generate synthetic text, which might produce comparable results as achieved through text augmentation in this study.

**Acknowledgement:** None.

**Funding Statement:** The authors received no specific funding for this study.

**Author Contributions:** The authors confirm contribution to the paper as follows: Experiments, analysis and interpretation of results and design: Bishrul Haq, Sher Muhammad Daudpota; Supervision of experiments and conceived the idea of the study: Sher Muhammad Daudpota; Draft manuscript preparation and contributed to writing: Zenun Kastrati, Ali Shariq Imran; Providing visualization to the paper: Zenun Kastrati; Improving the overall language of the paper: Waheed Noor. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The authors confirm the dataset used in the study is publicly available and is accessible within the article.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] W. Zhang, X. Li, Y. Deng, L. Bing and W. Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," arXiv Prepr. arXiv2203.01054, 2022.
- [2] B. Fakieh, A. S. Al-Malaise AL-Ghamdi, F. Saleem and M. Ragab, "Optimal machine learning driven sentiment analysis on COVID-19 twitter data," *Computers, Materials & Continua*, vol. 75, no. 1, pp. 81–97, 2023.
- [3] Australian Computer Society, ACS code of ethics, 2018. [Online]. Available: <https://www.acs.org.au/content/dam/acs/acs-documents/Code-of-Ethics.pdf>
- [4] H. H. Do, P. W. C. Prasad, A. Maag and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: A comparative review," *Expert System with Applications*, vol. 118, pp. 272–299, 2019.
- [5] D. Gu, J. Wang, S. Cai, C. Yang, Z. Song *et al.*, "Targeted aspect based multimodal sentiment analysis: An attention capsule extraction and multi-head fusion network," arXiv Prepr. arXiv2103.07659, 2021.
- [6] L. Yue, W. Chen, X. Li, W. Zuo and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems*, vol. 60, no. 1, pp. 617–663, 2019.
- [7] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, vol. 3, no. 2, pp. 143–157, 2009.
- [8] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci and M. A. Wani, "Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study," *Applied Sciences*, vol. 11, no. 9, pp. 3986, 2021.

- [9] C. Shorten, T. M. Khoshgoftaar and B. Furht, "Text data augmentation for deep learning," *Journal of Big Data*, vol. 8, no. 1, pp. 101, 2021.
- [10] A. S. Imran, R. Yang, Z. Kastrati, S. M. Daudpota and S. Shaikh, "The impact of synthetic text generation for sentiment analysis using GAN based models," *Egyptian Informatics Journal*, vol. 23, no. 3, pp. 547–557, 2022.
- [11] S. M. Jiménez-Zafra, M. T. Mart'in-Valdivia, E. Mart'inez-Cámara and L. A. Ureña-López, "Combining resources to improve unsupervised sentiment analysis at aspect-level," *Journal of Information Science*, vol. 42, no. 2, pp. 213–229, 2016.
- [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, pp. e1044, 2022.
- [13] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14–46, 2015.
- [14] H. Han, J. Zhang, J. Yang, Y. Shen and Y. Zhang, "Generate domain-specific sentiment lexicon for review sentiment analysis," *Multimedia Tools and Applications*, vol. 77, no. 16, pp. 21265–21280, 2018.
- [15] M. S. Akhtar, D. Gupta, A. Ekbal and P. Bhattacharyya, "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis," *Knowledge-Based Systems*, vol. 125, pp. 116–135, 2017.
- [16] A. S. Shafie, N. M. Sharef, M. A. A. Murad and A. Azman, "Aspect extraction performance with POS tag pattern of dependency relation in aspect-based sentiment analysis," in *2018 Fourth Int. Conf. on Information Retrieval and Knowledge Management (CAMP)*, Kota Kinabalu, Malaysia, pp. 1–6, 2018.
- [17] K. Srividya and A. M. Sowjanya, "Aspect based sentiment analysis using POS tagging and TFIDF," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6, pp. 1960–1963, 2019.
- [18] A. R. Martinez, "Part-of-speech tagging," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 1, pp. 107–113, 2012.
- [19] M. Banko and R. C. Moore, "Part-of-speech tagging in context," in *COLING 2004: Proc. of the 20th Int. Conf. on Computational Linguistics*, Geneva, Switzerland, pp. 556–561, 2004.
- [20] J. Singh, N. Joshi and I. Mathur, "Development of Marathi part of speech tagger using statistical approach," in *2013 Int. Conf. on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, India, pp. 1554–1559, 2013.
- [21] M. A. Faruque, S. Rahman, P. Chakraborty, T. Choudhury, J. S. Um *et al.*, "Ascertaining polarity of public opinions on Bangladesh cricket using machine learning techniques," *Spatial Information Research*, vol. 30, no. 1, pp. 1–8, 2022.
- [22] S. Haque, T. Rahman, A. K. Shakir, M. S. Arman, K. B. B. Biplob *et al.*, "Aspect based sentiment analysis in Bangla dataset based on aspect term extraction," in *Int. Conf. on Cyber Security and Computer Science*, Dhaka, Bangladesh, pp. 403–413, 2020.
- [23] A. S. Mohammad, O. Qwasmeh, B. Talafha, M. Al-Ayyoub, Y. Jararweh *et al.*, "An enhanced framework for aspect-based sentiment analysis of hotels' reviews: Arabic reviews case study," in *2016 11th Int. Conf. for Internet Technology and Secured Transactions (ICITST)*, Barcelona, Spain, pp. 98–103, 2016.
- [24] A. Ozcan, B. Emiral and A. B. Cetin, "Deep hotel recommender system using aspect-based sentiment analysis of users' reviews," in *26th Int. Conf. on Pattern Recognition (ICPR)*, Montreal, QC, Canada, pp. 3090–3096, 2022.
- [25] 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.
- [26] X. Wang, F. Li, Z. Zhang, G. Xu, J. Zhang *et al.*, "A unified position-aware convolutional neural network for aspect based sentiment analysis," *Neurocomputing*, vol. 450, pp. 91–103, 2021.
- [27] R. K. Yadav, L. Jiao, M. Goodwin and O. C. Granmo, "Positionless aspect based sentiment analysis using attention mechanism," *Knowledge-Based Systems*, vol. 226, pp. 107136, 2021.
- [28] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

- [29] R. Masadeh and B. H. Sa'ad Al-Azzam, "A hybrid approach of lexicon-based and corpus-based techniques for arabic book aspect and review polarity detection," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 4, pp. 4336–4340, 2020.
- [30] 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.
- [31] S. R. A. R. Sarno and C. Fatichah, "Aspect-based sentiment analysis for sentence types with implicit aspect and explicit opinion in restaurant review using grammatical rules, hybrid approach, and SentiCircle," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 5, pp. 294–305, 2021.
- [32] A. Kumar and V. H. C. Albuquerque, "Sentiment analysis using XLM-R transformer and zero-shot transfer learning on resource-poor Indian language," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, pp. 1–13, 2021.
- [33] K. T. K. Phan, D. N. Hao, D. van Thin and N. L. T. Nguyen, "Exploring zero-shot cross-lingual aspect-based sentiment analysis using pre-trained multilingual language models," in *2021 Int. Conf. on Multimedia Analysis and Pattern Recognition (MAPR)*, Hanoi, Vietnam, pp. 1–6, 2021.
- [34] N. C. Lê, N. T. Lam, S. H. Nguyen and D. T. Nguyen, "On Vietnamese sentiment analysis: A transfer learning method," in *2020 RIVF Int. Conf. on Computing and Communication Technologies (RIVF)*, Ho Chi Minh City, Vietnam, pp. 1–5, 2020.
- [35] P. A. Utama, N. S. Moosavi, V. Sanh and I. Gurevych, "Avoiding inference heuristics in few-shot prompt-based finetuning," arXiv Prepr. arXiv2109.04144, 2021.
- [36] B. Wang, L. Ding, Q. Zhong, X. Li and D. Tao, "A contrastive cross-channel data augmentation framework for aspect-based sentiment analysis," arXiv Prepr. arXiv2204.07832, 2022.
- [37] R. Egger and E. Gokce, "Natural language processing (NLP): An introduction: Making sense of textual data," *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications*, Springer, vol. 2022, pp. 307–334, 2022.
- [38] M. Edalati, A. S. Imran, Z. Kastrati and S. M. Daudpota, "The potential of machine learning algorithms for sentiment classification of students' feedback on MOOC," in *Proc. of SAI Intelligent Systems Conf.*, London, UK, pp. 11–22, 2022.
- [39] B. Liu, "Sentiment analysis and subjectivity," *Handbook of Natural Language Processing*, vol. 2, no. 2010, pp. 627–666, 2010.
- [40] Z. Kastrati, A. S. Imran and A. Kurti, "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106799–106810, 2020.
- [41] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos *et al.*, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. of the 8th Int. Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, pp. 27–35, 2014.
- [42] D. Tang, B. Qin and T. Liu, "Aspect level sentiment classification with deep memory network," arXiv Prepr. arXiv1605.08900, 2016.
- [43] C. Zhang, Q. Li and D. Song, "Aspect-based sentiment classification with aspect-specific graph convolutional networks," arXiv Prepr. arXiv1909.03477, 2019.
- [44] B. Verma and R. S. Thakur, "Sentiment analysis using lexicon and machine learning-based approaches: A survey," in *Proc. of Int. Conf. on Recent Advancement on Computer and Communication*, Singapore, pp. 441–447, 2018.
- [45] A. García-Pablos, M. Cuadros and G. Rigau, "W2VLDA: Almost unsupervised system for aspect based sentiment analysis," *Expert Systems with Applications*, vol. 91, pp. 127–137, 2018.
- [46] M. Venugopalan and D. Gupta, "An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis," *Knowledge-Based Systems*, vol. 246, pp. 108668, 2022.
- [47] A. Purpura, C. Masiero and G. A. Susto, "WS4ABSA: An NMF-based weakly-supervised approach for aspect-based sentiment analysis with application to online reviews," in *Int. Conf. on Discovery Science*, Halifax, NS, Canada, pp. 386–401, 2018.

- [48] W. Wang, V. W. Zheng, H. Yu and C. Miao, “A survey of zero-shot learning: Settings, methods, and applications,” *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–37, 2019.
- [49] L. Shu, J. Chen, B. Liu and H. Xu, “Zero-shot aspect-based sentiment analysis,” arXiv Prepr. arXiv2202.01924, 2022.
- [50] R. Seoh, I. Birle, M. Tak, H. S. Chang, B. Pinette *et al.*, “Open aspect target sentiment classification with natural language prompts,” arXiv Prepr. arXiv2109.03685, 2021.
- [51] S. Kobayashi, “Contextual augmentation: Data augmentation by words with paradigmatic relations,” arXiv Prepr. arXiv1805.06201, 2018.
- [52] T. Liesting, F. Frasincar and M. M. Trușcă, “Data augmentation in a hybrid approach for aspect-based sentiment analysis,” in *Proc. of the 36th Annual ACM Symp. on Applied Computing*, New York, NY, USA, pp. 828–835, 2021.
- [53] Q. Jiang, L. Chen, R. Xu, X. Ao and M. Yang, “A challenge dataset and effective models for aspect-based sentiment analysis,” in *Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, pp. 6280–6285, 2019.
- [54] T. Hercig, T. Brychcín, L. Svoboda and M. Konkol, “UWB at semeval-2016 task 5: Aspect based sentiment analysis,” in *Proc. of the 10th Int. Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, California, pp. 342–349, 2016.
- [55] I. Vayansky and S. A. P. Kumar, “A review of topic modeling methods,” *Information Systems*, vol. 94, pp. 101582, 2020.
- [56] P. Kherwa and P. Bansal, “Topic modeling: A comprehensive review,” *ICST Transactions on Scalable Information Systems*, vol. 7, no. 24, pp. 2–10, 2019.
- [57] D. M. Blei, A. Y. Ng and M. I. Jordan, “Latent dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [58] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang *et al.*, “Latent dirichlet allocation (LDA) and topic modeling: Models, applications, a survey,” *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169–15211, 2019.
- [59] M. Grootendorst, “BERTopic: Neural topic modeling with a class-based TF-IDF procedure,” arXiv preprint arXiv:2203.05794, 2022.
- [60] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, pp. 60, 2019.
- [61] S. Y. Feng, V. Gangal, J. Wei, S. Chandar, S. Vosoughi *et al.*, “A survey of data augmentation approaches for NLP,” arXiv Prepr. arXiv2105.03075, 2021.
- [62] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi *et al.*, “RoBERTa: A robustly optimized bert pretraining approach,” arXiv Prepr. arXiv1907.11692, 2019.
- [63] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” arXiv Prepr. arXiv1810.04805, 2018.
- [64] V. Sanh, L. Debut, J. Chaumond and T. Wolf, “DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter,” arXiv Prepr. arXiv1910.01108, 2019.
- [65] N. Reimers and I. Gurevych, “Sentence-bert: Sentence embeddings using siamese bert-networks,” arXiv Prepr. arXiv1908.10084, 2019.
- [66] W. Xue, W. Zhou, T. Li and Q. Wang, “MTNA: A neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews,” in *Proc. of the Eighth Int. Joint Conf. on Natural Language Processing*, Taipei, Taiwan, pp. 151–156, 2017.
- [67] H. Luo, T. Li, B. Liu, B. Wang and H. Unger, “Improving aspect term extraction with bidirectional dependency tree representation,” *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 27, no. 7, pp. 1201–1212, 2019.
- [68] R. Agerri and G. Rigau, “Language independent sequence labelling for opinion target extraction,” *Artificial Intelligence*, vol. 268, pp. 85–95, 2019.

- [69] M. S. Akhtar, T. Garg and A. Ekbal, "Multi-task learning for aspect term extraction and aspect sentiment classification," *Neurocomputing*, vol. 398, pp. 247–256, 2020.
- [70] C. Wu, F. Wu, S. Wu, Z. Yuan and Y. Huang, "A hybrid unsupervised method for aspect term and opinion target extraction," *Knowledge-Based Systems*, vol. 148, pp. 66–73, 2018.
- [71] L. Khan, A. Amjad, N. Ashraf, H. Chang and A. Gelbukh, "Urdu sentiment analysis with deep learning methods," *IEEE Access*, vol. 9, pp. 97803–97812, 2021.
- [72] A. Ghafoor, A. S. Imran, S. M. Daudpota, Z. Kastrati and S. Abdullah, "The impact of translating resource-rich datasets to low-resource languages through multi-lingual text processing," *IEEE Access*, vol. 9, pp. 124478–124490, 2021.
- [73] Y. Qu, P. Liu, W. Song, L. Liu and M. Cheng, "A text generation and prediction system: Pre-training on new corpora using BERT and GPT-2," in *Proc. of the IEEE 10th Int. Conf. on Electronics Information and Emergency Communication (ICEIEC)*, Beijing, China, pp. 323–326, 2020.