IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 10 July 2024, accepted 4 August 2024, date of publication 8 August 2024, date of current version 19 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3440357

RESEARCH ARTICLE

Aspect-Based Sentiment Analysis for Service Industry

AFSHEEN MAROOF[1], SHAUKAT WASI[1], SYED IMRAN JAMI[1], AND MUHAMMAD SHOAIB SIDDIQUI[2], (Member, IEEE)

[1]Department of Computer Science, Muhammad Ali Jinnah University, Karachi 75400, Pakistan
[2]Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia

Corresponding author: Afsheen Maroof (sp20phcs0005@maju.edu.pk)

ABSTRACT In today's digital age, customer feedback, particularly gathered from various sources like mobile application reviews, has emerged as a critical resource for service-providing organizations to gain valuable insights into their customers' experiences. As the key objective of service-providing organizations is to facilitate their customers with better services, customer feedback or opinion is a vital resource for such organizations to improve and enhance their services for the betterment of their customers. Explicitly mentioned opinions have been widely studied in research, while a significant gap exists in addressing implicitly described views. Furthermore, most existing research focuses on product-oriented corpora, emphasizing specific product aspects and features. This article presents a novel study on performing end-to-end aspect-based sentiment analysis (ABSA) by extracting implicit opinion terms, categorizing them, and assigning polarity to each term from mobile app reviews in English. Through this study, we developed a domain-specific, service-oriented, and aspect-based annotated dataset and introduced a novel two-step hybrid approach. The first step involves extracting multiple opinion terms using a rule-based approach. The second step employs machine learning and deep learning algorithms to classify the extracted opinion terms into general aspect categories. This two-step approach effectively addresses the double-implicit problem commonly encountered in the previous work on implicit aspects and opinion mining. In addition to traditional machine learning and deep learning models, we fine-tuned BERT to carry out the ABSA task. This approach utilized a pipeline method, where each task's output serves as the subsequent task's input, ensuring a seamless flow of information and improved performance. This multi-step pipeline begins with the extracted opinion terms classification into aspect categories and ends with the assignment of sentiment polarity. Experiments with a hold-out test set for the first step (opinion term extraction using a rule-based approach) achieved an accuracy and precision score of 81.4% and an F1 score of 0.99 %, outperforming several baselines. Further experiments with a range of machine learning and deep learning algorithms for classifying extracted opinion terms into general aspect categories yielded accuracy scores ranging from 0.68% to 0.74 % and F1 scores ranging from 0.23% to 0.28%. Experiments on sentiment classification using various machine learning algorithms showed accuracy ranges from 0.58% to 0.68% and F1 scores from 0.47% to 0.49%. This two-step approach for implicit opinion term, aspect extraction, and classification outperforms many baseline systems. By leveraging BERT's contextual understanding and fine-tuning it for our specific domain, we significantly improved the accuracy and robustness of our aspect-based sentiment analysis. This approach effectively captured both explicit and implicit opinions from mobile app reviews. Specifically, our method achieved an accuracy of 0.80% and an F1 score of 0.78% for aspect categorization, and an accuracy of 0.79% and an F1 score of 0.70% for sentiment classification, demonstrating substantial improvements over traditional methods.

INDEX TERMS Implicit aspects, aspect categories, aspect sentiment classification, pattern creation, lexicon and rules-based, machine learning, mobile app review, service industry, feedback, decision making, consumer protection, aspect-based sentiment analysis (ABSA).

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

I. INTRODUCTION

With the expansion of interactive mobile apps, people exchange an enormous quantity of data daily, expressing their thoughts about products, services, issues, etc. [1]. It can impact a customer's outcomes in purchasing a product or selecting a specific service and support organizations to observe the experiences of their users [2]. In service-providing companies, customer opinion and feedback about their experiences are vital in offering valuable insights as these companies keep customers' needs at the forefront and aim to deliver efficient and cost-effective yet quality services. The key objective of service-providing companies is to deal with customers' or clients' feedback, criticism, or reviews [3]. Similarly, these opinions and feedback also help other customers in decision-making, who often ask about the feedback of the customers who have already experienced the service. Also, they can use this feedback to protect their consumer rights if the promised services are not provided.

Reference [4] defines the term **opinion** as *"a concept covering sentiment, evaluation, appraisal, or attitude held by a person"*. Aspects and opinions are the more likely topics in text classification. Text analysis at this level was coined as *feature-based or aspect-based analysis* by [5]. Aspect-based Sentiment Analysis (ABSA) is a fine-grained type of sentiment analysis that entails customer feedback or opinion about a specific characteristic of a product or service. Aspects are different features relating to a review item (product or service) [6]. Aspect-based analysis determines the target of the opinion. It is categorized into three tasks: Aspect Term Extraction, Aspect Categorization, and Aspect Sentiment Classification [7]. Aspect Term Extraction is a dynamic field of research as it is a crucial step in Aspect-Based recognition [8]. The goal of Aspect Term Extraction is to extract all the opinion targets and relevant aspects. Opinion targets are the entities on which the opinion is expressed. Extracted terms can be explicit or implicit [9]. Explicit terms generally consist of opinion terms that indicate the general aspect category. For example, the review *"worst payment experience"* contains an explicit term *"payment"* with an adjective worst that shows negative sentiment towards the aspect category *"payment"*. However, in the text there is no direct aspect mentioned in the implicit aspect. Rather, we need to identify those from the words or expressions expressed in the customer reviews [10]. For example the review *"very good service"* contains an opinion word *"good service"* which implies the aspect category *"Company Service"*.

The next task of aspect-based sentiment analysis is Aspect Sentiment Classification, where we determine sentiment polarities (e.g. positive, negative, neutral) for specific aspects in a review text [11]. This article defines around ten general aspect categories of domain-specific service-oriented mobile app datasets after carefully analyzing the domain data. Our goal is to extract explicit opinion terms from a given sentence, map them to predefined categories, and classify them according to sentiment polarities. To achieve this, we developed a model that aids service-oriented organizations in gaining deeper insights into customers' needs and likes and dislikes through a two-step hybrid method.

In the first step, opinion terms are extracted from sentences using NLP techniques, including part-of-speech (POS) tagging and dependency parsing. POS tagging plays a crucial role in detecting aspect and sentiment words by identifying the grammatical structure of sentences [12], while dependency parsing helps understand the relationships between words. Extracted terms are then validated by comparing them with an integrated lexicon consisting of opinion terms.

In the second step, various machine learning and deep learning models map the predicted opinion terms to predefined categories. These predefined categories include *service quality, company service, service schedule, customer support, service person attitude, app UI, app functionality, app utility, cost, and payment*.

The final part of the proposed approach involves Aspect Sentiment Classification for each identified aspect. Here, we employ sophisticated machine learning and deep learning models to determine whether the sentiments expressed towards each predicted aspect category are positive or negative. In addition to traditional machine learning and deep learning models, we fine-tuned BERT specifically for the ABSA task. Our approach leveraged a pipeline method, where the output of each task is seamlessly fed into the next, ensuring a cohesive flow of information and enhancing overall performance. This multi-step pipeline begins with the classification of extracted opinion terms into aspect categories, followed by the assignment of sentiment polarity.

To achieve our goal, we created a new dataset,[1] which, to the best of our knowledge, is the first to carry out all three tasks of Aspect-Based Sentiment Analysis (ABSA) together for a service-oriented mobile app. This is also the first attempt to provide directions that help service-oriented companies identify efficient areas and the areas that need improvement. Companies can address the challenges identified through our model and make data-driven decisions to enhance their offerings, improve customer experiences, and drive business growth. Our specific contributions are listed below.

1) Construction of a new domain-specific service-oriented dataset.
2) Proposed a hybrid approach using Linguistic features, machine learning, and deep learning models to resolve the double implicit problem of ABSA.
3) Extraction of single and multiple Multiple Opinion terms.
4) Fine-tuning BERT on the newly constructed dataset to perform ABSA tasks using the pipeline method.

A. ORGANIZATION OF THE ARTICLE

The article is structured as follows. In Section II, the relevant work in the domain is reviewed. Section III illustrates the

[1] https://data.mendeley.com/datasets/hr6j6yr3gd/3

research methodology and related experiments to extract aspects and classify the sentiments of the app reviews. Section IV discusses the experiments and performance results. Section V presents the conclusion.

II. BACKGROUND STUDY

A. ASPECT EXTRACTION

This section includes a brief review of current and cutting-edge techniques that were used for the Aspect Term Extraction phase. The conducted work can be broadly grouped into three categories. Supervised methods appear to have labeled data to decide the sense of a sentence [13]. All unsupervised models are trained on the unlabeled dataset [14], while semi-supervised methods handle a small portion of labeled corpora to label [15].

Reference [4] discussed four ways for Explicit Aspect Extraction: (1) Frequent nouns and noun phrases extraction, (2) Exploiting opinion and target relations extraction, (3) Supervised learning, and (4) Topic modeling.

The authors in [16] used an opinion lexicon along with POS tagging and rules to discover aspects and opinion terms from review, as well as, arrange them into groups based on similarity. POS tagging is the method of parsing each part of the sentence based on detecting linguistic tags. The authors in [17] proposed a rule-based approach to extract aspect terms from reviews using a sequence of patterns that are created based on the dependency relations between opinion and its nearby words on a Hindi dataset.

The authors in [10] suggested two methods for the aspect category prediction work. First, a multivariate method of selecting the features is proposed that demonstrates that if the redundancy among selected features is minimized then the same features boost the performance of classification. Then they chose dependency relations-based features for Aspect Category detection and demonstrated that the grammatical rules used for feature selection give better results in classification. The authors in [18] also used Dependency Parsing, and part-of-speech tagging of natural texts to obtain the syntactic structure of sentences by employing a dependency relation rule. They employed Stanford dependency relations and Natural Language Processing as linguistic features and presented an aspect-based opinion-mining extraction. They evaluated their approach to the product review dataset of Amazon.

The authors in [19] aimed a supervised Aspect Extraction algorithm for the extraction of explicit aspects from reviews. For this purpose, they used pattern extraction and dependency rules. They also enhanced the Whale Optimization Algorithm to address the rule selection issue with an improved algorithm called *improved WOA*. Explicit Aspects Extraction has been studied broadly and earned attention; however, there is still room left for implicit aspects as it is more challenging to identify them. Limited studies have been done on Implicit Aspect Extraction and their mapping with Explicit Aspects.

In [20], the authors gave an explicit and implicit aspect opinion mining structure and algorithm that first decoded the explicit aspects with the help of frequent nouns and then extracted the implicit aspects using the implicit aspect recognizer by supervised machine-learning technique i.e. CRF. They have used the dataset from the Tourism Domain for their work.

The authors in [21] provided a detailed survey on suggested methods for determining implicit aspects. They grouped their studies according to the applied algorithms. They also advocated for the issues of implicit aspects which can open new directions for future work.

Reference [22] proposed a model that is a combination of lexicon and rules for aspect extraction in smart government apps and organizing all corresponding sentiments.

Authors in [23] suggested a way to use descriptive logic to identify explicit aspects as well as their sentiment polarities. They evaluated the results using a public dataset in Arabic based on the SemEval-2014 workshop and baseline experiments conducted by HAAD providers [24] proposed a new approach to extract all aspect category-opinion-sentiment quadruples. They also extended several datasets of the Laptop and Restaurant domains for their task. The new dataset consists of aspect categories, opinions, and sentiments, along with Implicit Aspects and opinions. Also, they focused the task on four new baselines; however, the proposed systems are relatively simple, and further improvements are needed.

Implicit aspect work was also explored by [25], where the authors attempted to boost the information retrieval techniques by mining both explicit and implicit aspects. They extracted sentence dependency using heuristic patterns with linguistic features and different implicit features were detected using a cuckoo search optimization algorithm. The suggested approach was evaluated in both English and Arabic language.

The authors in [26] utilized dependency tree-based rules for the extraction of the concepts and aspects and identification of the attached sentiments.

Similarly, [27] tried to cater to the issue of extracting aspects from product reviews. They proposed a new rule-based model that utilizes a sentence dependency tree and common-sense knowledge to find both explicit and implicit aspects and performed experiments on two product review datasets.

Authors in [28] performed a series of experiments in the Aspect Extraction domain. They used micro-blog messages from the stock investment domain. Predefined aspects were identified via taxonomy. Distributional semantic models and machine learning methods like Boost, Random Forest, SVC, and CRF were used. Evaluation results show that it performs well for Explicit identification with an accuracy of 0.82%, but Implicit Extraction with an accuracy of 0.35% is still an issue to tackle with a larger dataset and improved feature engineering. In their research efforts, [29] used a new two-way method that initially enhances pre-trained embeddings for sequential with SGD conditional random fields (CRF) and then used machine and ensemble methods to categorize the implied aspects. They used a domain-specific airline dataset to perform their experiments.

The authors in [30] aimed to extract Implicit opinions with their corresponding Implicit Aspects. They used a Chinese hotel dataset as a case study and clustered them as Positive, Negative, or Neutral. They constructed an opinion-based dictionary. If a word's POS is NN, it occurs at least 100 times and is accompanied by an opinionated word, which is considered an Aspect. They found it challenging to predict the Implicit Aspects for short-length sentences.

The authors in [31] tried to enhance the feedback system of the educational institution. They composed student responses from Twitter API and analyzed them by calculating semantic relatedness between aspect words and students' views using different clustering and classification techniques. Similarly, [32] also performed ABSA on student survey corpus in the Serbian language using an integrated approach of dictionary-based and machine learning models. The authors in [19] provided a different perspective to review Implicit Aspects. They suggested that the first perspective can be, to make an evaluation of available techniques for extraction of implicit aspect. They have discussed a summary of each technique from all three supervised, unsupervised, and semi-supervised learning. The second way they suggested is to categorize and measure the performance, datasets, language, and flaws of the available techniques. In their findings, they emphasized that most researchers followed unsupervised methods for implicit extraction. Additionally, they found that datasets utilized for Implicit Extraction were made by the researchers themselves and are not accessible openly.

The authors in [33] discussed the sentiment analysis at the aspect level for a guest house. They scrapped the reviews from *trip advisor* and *booking.com* and conducted the experiments on 5 aspects using two machine learning classifiers; SVM and Naive base.

To predict restaurant survival factors, [34] investigated customer reviews of two U.S. restaurants using a machine learning-based conditional survival forest model. This analysis highlighted various aspects of restaurant management and identified the most impactful factors on overall survival prediction.

B. ASPECT SENTIMENT CLASSIFICATION

In previous studies, two leading approaches discussed for Aspect Sentiment Classification (ASC) are the lexicon-based method and the supervised learning approach. In the Lexicon-based approach, ASC was handled manually by using a sentiment lexicon, opinions rules, and the sentence parsing tree to find out the sentiment polarity of every aspect of a sentence. But, by the headway of deep-learning mechanisms, ASC methods have been anticipating automatic learning of aspects and their opinions, giving the best answers to many issues in ABSA [8].

Reference [35] presented an attention over-attention neural network that jointly represents aspect and sentence, but their model cannot handle complex sentiments.

C. BERT FOR ABSA

The authors in [36] analyzed the pre-training of BERT. They investigated the pretext of the masked language model. They explored that BERT uses limited self-attention to encode opinion words from aspect. They used an Amazon review dataset along with the Yelp dataset. Bidirectional Encoder Representation from the Transformer used the Attention mechanism to learn the context in the text.

The authors in [37] fine-tuned the BERT to extract the implicit aspect from the Chinese clothing review dataset for six different aspects. Reference [38] analyzed the modeling capabilities of contextual embeddings from the BERT model. By incorporating BERT embeddings with a simple classification layer, they found improved results compared to traditional methods. They tested this approach on three Arabic datasets.

III. PROPOSED METHODOLOGY

The service-providing companies must determine if the quality of their service meets the needs of their customers [39]. In a service review, people generally give feedback on several aspects and have different opinions on various aspects of that service. Particularly, in a review, it might consist of three issues:

1) **Extraction of Implicit Opinion Terms:** We came across the fact that users of the service industry do not always use explicit terms to express their thoughts instead they use some implicit words or terms to convey their opinions. For example, in **"too much expensive"**, the user is talking about cost but he used the implicit term **"expensive"**. All these implicit terms should be under some general aspect category.
2) **Extraction of Multi-word Opinion Terms:** Opinion terms do not need to be always a single term; rather, user may use phrases to share their thoughts. For example, in **"worst OTP service"** term, **"OTP service"** conveys the actual meaning these phrases need to be extracted.
3) In a given review, there can be multiple opinion terms, like in example **"excellent app. excellent service."**, the user expresses his feelings in two opinion terms. Here, all the terms should be extracted.

A rule-based approach incorporating an integrated lexicon addressed these issues comprehensively. The sentences were split to ensure each section contained at least one aspect term, ensuring the extraction of multiple terms from single sentences. Explicit terms, including multi-word expressions, were extracted using a proposed lexicon and pattern-based technique. Following the aspect extraction phase, supervised machine learning and deep learning techniques were employed to assign a general category to each extracted term.

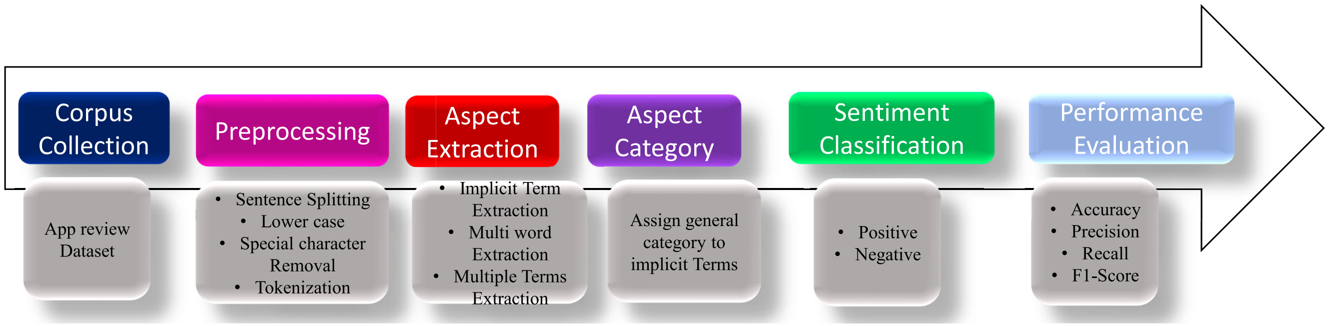


FIGURE 1. Research Methodology.

TABLE 1. Dataset statistics.

Total Number of Sentences	4164
Sentence with tokens Less than three	2045
Sentence with tokens More than three	2118
Distinct Opinion Terms in dataset	977
Commonly used distinct term	464

TABLE 2. Top 10 distinct terms.

good service	464
excellent service	272
good app	125
nice service	107
great service	83
good experience	80
worst service	73
nice app	68
good services	67
bad service	66

Subsequently, the terms were classified based on sentiment using both machine learning and deep learning models. Figure 1 illustrates the proposed methodology. The following subsections illustrate the methodology in detail.

A. CORPUS DETAILS

To achieve the goal of this study, which is to perform end-to-end aspect-based sentiment analysis (ABSA) and extract implicit aspects, we constructed a novel dataset from mobile app reviews of service-providing companies. To the best of our knowledge, this is the first dataset[2] created for this specific domain. The dataset was manually annotated and subsequently verified using a lexicon tailored for this domain.

After the curation process, we selected 4,165 random sentences, with the maximum token length per sentence being 10. Each review sentence was annotated with the following labels: *opinionterm1, aspectcategory1, aspectpolarity1, opinionterm2, aspectcategory2*, and *aspectpolarity2*. The annotation process adhered to the guidelines established by SemEval.[3] An inter-annotator agreement [40] analysis was conducted over the same data of 200 reviews and Cohen Kappa's [41] coefficient was used to determine the quality of the annotation. Table 1 summarizes detailed statistics analysis, while Table 2 depicts the top ten distinct terms.

1) PRE-PROCESSING

In real-world customer reviews, the sentences are often very filthy; grammar and punctuation rules are not followed; however, the basic principles of the approach remain the same [42]. The first and foremost step in the preprocessing was to split the sentence so that we could get the Opinion Terms from each part of the sentence. Because users have the leverage to write anything in the realistic dataset, we use three handicrafts rules to split.

1) **Split the sentences by punctuation:** The first rule states that "if there is a punctuation such as a comma or period in a sentence, split the sentence at that point".
2) **Split the sentences by Conjunction:** The second rule states that "if in a sentence there occurs a conjunction, such as (and, for, etc.), split the sentence at that point".
3) **Split the sentences by Noun Term:** The third rule states that "if there occurs a noun term, split the sentence there".

We set priorities on these rules. Our proposed approach is illustrated in Function 1. The function will first search for punctuation. If it finds any, it will apply rule 1 and the sentence will be split there. If no punctuation is found in the review, the proposed algorithm will search for conjunction, and rule 2 will be applied. If such a sentence occurs where there is neither punctuation nor a conjunction then the proposed algorithm will look for the occurrence of a noun term and will split the sentence as soon as it gets the first noun term. Frequent noun terms are commonly used for opinion extraction, but here we use them for sentence splitting.

Examples of all three rules are discussed in Table 3, while Table 4 depicts the split sentence statistics. 1998 sentences were split using these rules while the rest have a token length of less than four so they remain the same.

[2]https://data.mendeley.com/datasets/hr6j6yr3gd/3
[3]https://alt.qcri.org/semeval2016/task5/data/uploads/absa2016_annotationguidelines.pdf

TABLE 3. Sentence splitting examples.

Sentences	Broken Sentences	Split Rule# Number
excellent service....very professional electrician they have	['excellent service', 'very professional electrician they have']	Rule #A (1.1)
Perfect App. Good Services Offered.	['Perfect App', 'Good Services Offered']	Rule #A (1.2)
nice facilities and rates are too much high	['nice facilities', 'rates are too much high']	Rule #2
came on time and worked professionally complete	['came on time', 'worked professionally complete']	Rule #3

Algorithm 1 Sentence Splitting Algorithm

```
function SplitSentence(sentence,tokens,conjunction_preposition)
    if token_length > 4 then
        if sentence contains any punctuation then
            split_sentences ←Split sentence using regex pattern
            return split_sentences, 'Rule #1'
        end if
        for each word in conjunction_preposition do
            if word is in tokens then
                split_sentences ← Split sentence using regex
                    pattern '{}' with max 1 split
                if length of split_sentences is 2 then
                    return split_sentences, 'Rule #2'
                end if
            end if
        end for
        POSTAG ← get_pos_tags(tokens)
        for i ← 1 to length(POSTAG) - 1 do
            (, pos) ← POSTAG[i]
            if pos starts with 'NN' then
                split_sentences ← [' '.join(tokens[:i+1]), '
                    '.join(tokens[i+1:])]
                return split_sentences, 'Rule #3'
            end if
        end for
    end if
    return sentence
end function = 0
```

TABLE 4. Split sentence statistics.

Splitting Rule	Number of Sentences Split
Rule 1.1	69
Rule 1.2	515
Rule 2	865
Rule 3	549

The rest of the preprocessing phase includes lower casing, removal of any special characters, tokenization, POS tagging, and dependency parsing. For POS tagging, we used the Natural Language Tool Kit (NLTK)[4] tagger, while for dependency parsing, a spaCy dependency parser was used. Sample output is shown in Figure 3. The output of the dependency parser is stored in a suitable data structure and is illustrated in Table 5. Here, sentence ID represents the place of sentences, word ID represents the position of the word in a sentence, Word denotes the original word in a sentence, Root describes the root word for that original word, POS tag represents the part of speech of the word, that could be JJ(adjective), PROPN(proper Noun), CCONJ(coordinating conjunction), etc. and Label is denoting dependency label of each word.

B. OPINION TERM EXTRACTION

To carry out the Opinion Term / Aspect Term extraction task, we used a lexicon[5] and a Pattern-based approach to extract single and multiple terms. The main reason to opt for this kind of approach is the selected unstructured nature of the dataset. Figure 2 gives a schematic overview of the proposed system starting from preprocessing to POS tagging followed by a Pattern.

1) PATTERN EXTRACTION AND RULES CREATION

The important task of Opinion Term Extraction is the generation of patterns. Patterns are generated based on relationships between POS tags and the dependency labels. We deeply observed the association between nouns with adjectives, nouns with verbs, nouns with nouns, etc. A total of 20 patterns are created along with the patterns of the sequence under each POS tag. Table 6 describes the details of sequence patterns, while Table7 shows the examples of extracted terms.

Class1: adjective association In the review sentence if a noun/noun phrase, adverb, or verb is directly associated with an adjective then the noun/noun phrase, verb, or adverb followed by the adjective will be picked as an opinion term.

Class2: noun association In the review sentence if a noun/noun phrase is followed by an adjective, adverb, verb, or coordinating conjunction then both terms will be picked as opinion terms.

Class3: proper noun association In the review sentence if there is an association between a proper noun and other words. Both the terms will be extracted as opinion terms.

Class 4: adverb association In the review sentence if there is an association between an adverb and another word.

[4]NLTK is a popular and widely used library of Python that offers a wide range of resources, including lexical resources, and trained models for POS tagging in several languages.

[5]In NLP, a lexicon is a component that contains a collection of words or phrases along with its linguistic information.

TABLE 5. Output detail of parser.

Sentence ID	Word ID	Word	Root	POS tag	Label
0	1	excellent	electrician	ADJ	amod
0	1	service	electrician	ADJ	amod
0	2	professional	electrician	ADJ	amod
0	3	electrician	Root	NOUN	Root
0	4	they	have	PRON	Label
0	5	have	electrician	VERB	nsubj
1	0	nice	app	ADJ	amod
1	1	and	nice	CCONJ	cc
1	2	good	nice	ADJ	conj
1	3	app	Root	NOUN	Root
2	0	came	Root	VERB	Root
2	1	on	came	ADP	prep
2	2	time	work	NOUN	compound
2	3	work	professionally	NOUN	compound
2	4	professionally	came	PROPN	npadvmod
2	5	complete	professionally	ADJ	amod

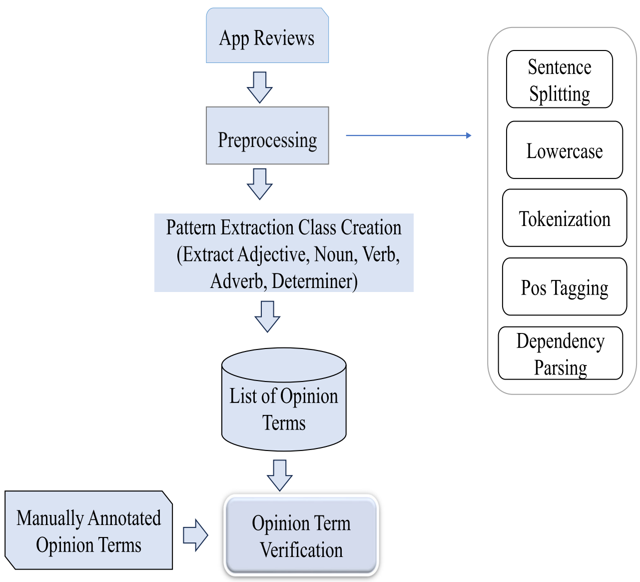


FIGURE 2. Schematic diagram of the proposed approach for Opinion term extraction.

Both the terms will be extracted as opinion terms. the whole process is summarized in Function 2.

The extracted terms are validated against a manually constructed opinion lexicon using the same dataset. The validation process ensures the relevance and correctness of the terms. Subsequently, the method efficiency is evaluated using a range of metrics, including accuracy, precision, recall, and F1-score, to assess its performance, comprehensively.

Table 8 presents the statistics for Opinion Term 1, while Table 9 provides the statistics for Opinion Term 2.

C. ASPECT CATEGORIZATION

The second step of the proposed approach is aspect categorization. Aspect categorization is the process in which terms with similar meanings are mapped to a single general category [43]. Categories are crucial to handle the double implicit problem. For a better understanding of the concept of aspect categorization, consider the example *"too much expensive"*. In this sentence term **expensive** can be placed in the general category **Cost**.

To fulfill the app review domain requirements, we have defined about **10** general aspect categories. Some categories are according to the written standards of Android, while others are defined by deeply examining the domain of reviews in the dataset.

The resulting categories were **App UI, APP functionality, APP utility, company's service, service schedule, service quality, service person attitude, customer support,** and **cost payment**.

We employed various machine learning and deep learning techniques to carry out aspect categorization process, enabling more precise and efficient identification and classification of different aspects within textual data.

1) ASPECT CATEGORIZATION WITH MACHINE LEARNING CLASSIFIERS

Figure 4 depicts the process of aspect categorization using machine learning models. The process starts with a concatenated input consisting of the original sentence and the extracted opinion term, followed by additional preprocessing steps like removing stop words and punctuation. Subsequently, this input data is transformed into features, which serve as the distinguishing characteristics utilized in machine learning approaches for classification [44]. For feature engineering, we used TFIDF.[6] TF-IDF identifies the most important and insignificant words from

[6]TFIDF is a widely used statistical method to measure how relevant a word of the document is in a collection of words. It is obtained by multiplying two matrices, the number of times words appear in a document and the inverse document frequency that indicates the proportion of documents in the dataset that contains the term.

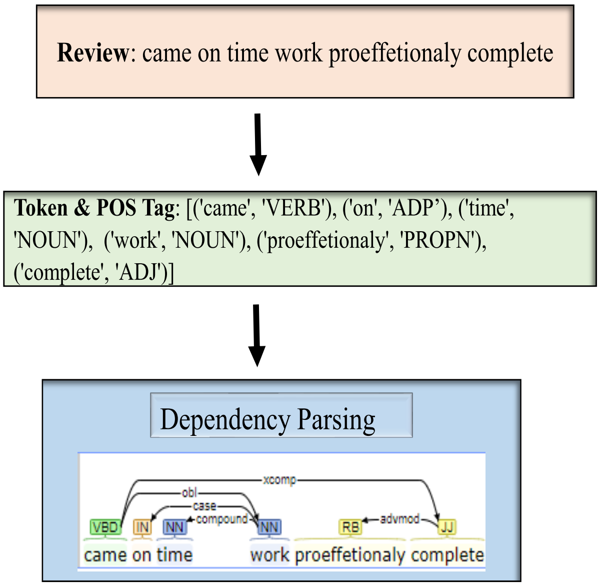


FIGURE 3. Output of POS Tagging and Dependency Parser.

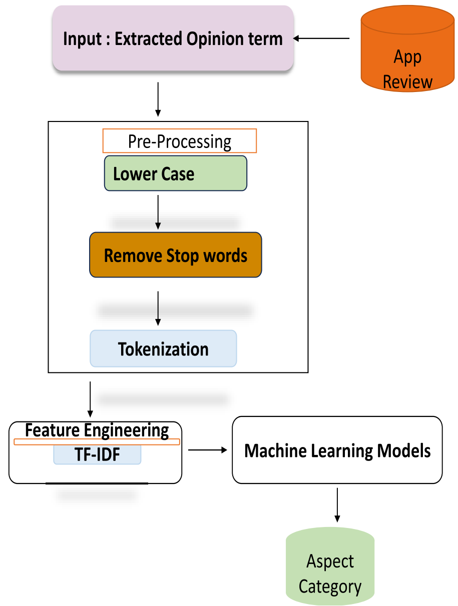


FIGURE 4. Aspect Category of Extracted Terms.

Algorithm 2 Pattern Extraction Algorithm

function Extract Patterns(text)
  patterns ← list of tuples {List of pattern descriptions}
  matches ← SplitTextWithRegularExpression(text)

  patterns&instances ← []
 {List to store patterns and instances}
  for pattern, instances, priority **in** patterns **do**
   instance ← [] {List to store instances of the current pattern}
   for i ← 1 **to** length(matches) - 1 **do**
    match ← regex.Match(matches[i - 1] **CONCATENATED WITH** matches[i])
    if match **then**
     instance.Add(matches[i - 1] **CONCATENATED WITH** matches[i])
    else if instance **then**
      {If an instance exists}
     patterns_and_instances.Add((pattern, instance, priority))
    end if
   end for
  end for
  return patterns_and_instances
 {Return the list of patterns and instances}
end function

the dataset [45]. The dataset was split into two subsets: training and testing. The training dataset comprises of 2914 (70%) of the reviews and the testing set comprises of 1249 (30%). Supervised machine learning models like SVM, KNN, and Naive Bayes were used to extract the aspect categories.

TABLE 6. Details of sequence pattern.

S.NO	First Word	Second word	Third word
1	JJ	NN	-
2	NNP	NN	-
3	JJ	NNS	-
4	'NN	NN'	-
5	NNP	NNP	
6	RB	JJ	
7	RB	NN	
8	NN	VBZ	JJ
9	NN	NNS	
10	NN	JJ	
11	CD	NN	
12	VBN	NNS	
13	JJR	NNS	
14	JJ	NN	NN
15	JJS	NN	
16	NNP	NN	NN
17	RB	VBN	
18	NN		
19	NNS		
20	JJ		

'JJ": Adjective, "NN": Noun (Singular), "NNP": Proper Noun (Singular), "NNS": Noun (Plural), "RB": Adverb, "VBZ": Verb (3rd person singular present), "CD": Cardinal Number, "VBN": Verb (Past Participle), "JJR": Adjective (Comparative), "JJS": Adjective (Superlative)

2) ASPECT CATEGORIZATION USING LSTM

Deep learning models represent an advanced type of machine learning, distinguished by their ability to autonomously extract features from data—a capability not typically found in traditional machine learning models. For the aspect categorization task, we utilized Long Short-Term Memory (LSTM) networks to capture contextual information from both the input sentence and the extracted opinion terms. By concatenating the LSTM inputs, the model effectively

TABLE 7. Examples of sequence patterns and extracted terms.

Sentence	Split Sentences	Sequence Pattern & Extracted Term 1	Sequence Pattern & Extracted Term 2
good professionals and good service excellent	['good professionals', 'good service excellent']	[JJ, NNS, good professionals]	[JJ, NN, good service]
Awesome service, prompt and helpful technicians	['Awesome service', 'prompt and helpful technicians']	[NNP, NN, Awesome service]	[JJ, NN, helpful technicians]
came on time work professionally completed	['came on time', 'work professionally complete']	[IN, NN, on time]	[JJ, NN, work professionally]
Wonderful experience.	-	[JJ, NN]	-

TABLE 8. Stats of opinion term 1.

True cases	3391
False cases	773

TABLE 9. Stats of opinion term 2.

True cases	3158
False cases	1006

integrates information from both sources to predict the aspect category.

D. ASPECT SENTIMENT CLASSIFICATION

After extracting and categorizing all opinion terms from a review sentence, our final step involves assigning sentiment polarity. Sentences were classified as positive or negative using various machine learning and deep learning models, including SVM, KNN, Naive Bayes, XGBoost, and LSTM. Our experiments demonstrated that SVM achieved the highest accuracy of 0.91%.

E. ABSA WITH BERT

Bidirectional Encoder Representations from Transformers outperform on a variety of NLP tasks [46], therefore we also fine-tuned the BERT for the ABSA tasks and extracted the aspect categories from the given sentence. A pipeline method is employed to conduct ABSA tasks (Aspect Categorization and Aspect Sentiment Classification), where the output of the first task serves as the input to the next task.

1) FOUR PARTS OF FINE-TUNING BERT

Fine-tuning BERT consists of four parts that are, the input layer, the embedding layer, the pooling layer, and the output layer (softmax layer).

In the input layer, each sentence is padded to ensure uniform length because BERT requires fixed-length inputs. In our case, the maximum input length for BERT is 128 tokens.

The embedding layer stores contextual embeddings derived from the input by the BERT model. These embeddings capture the semantic meaning and contextual information of each token within the input sequences.

In the pooling layer, the contextual information from the BERT embeddings is condensed into a fixed-sized vector. Finally, a dense layer with softmax activation is used as the output layer to predict the aspect category.

2) FINE-TUNING PROCESS

There are two steps in the training process *pretraining* and *Fine Tuning*. The first three stages of fine-tuning BERT during the pre-training step involve training on a large unlabelled dataset [37]. There are various pre-trained BERT. We have utilized BERT-BASED[7] for our work.

IV. EVALUATION MATRICES

A standard evaluation matrix is used to measure the performance. Opinion terms were evaluated against the manually-built lexicon of unique terms. For aspect categorization and classification, performance was measured on 1249 reviews(30%) of the overall 4164 sentences. This can assist in calculating advanced metrics, such as precision-recall accuracy. Performance evaluation measures are calculated according to the following formulas:

$$Accuracy = \frac{(TN) + (TP)}{(FP) + (TN) + (TP) + (FN)} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 \times \frac{(Precsion) \times (Recall)}{(Precision) + (Recall)} \quad (4)$$

TABLE 10. Opinion term extraction Confusion Matrix.

4*Actual (Annotated Dataset)	Predicted (Proposed opinion term extraction Approach)		
		Retrieved	Not Retrieved
	Relevant	TP (True Positive) Number of opinion terms extracted correctly	FP (False Positive) Number of opinion terms that are annotated but not extracted
	Irrelevant	FN (False negative) Number of opinion terms that are not annotated, but extracted by the algorithm	TN (True negative) Number of opinion terms that are not annotated, and not extracted by the algorithm

TABLE 11. Aspect Category extraction Confusion Matrix.

4*Actual (Annotated Dataset)	Predicted (Proposed Aspect Category Extraction Approach)		
		Retrieved	Not Retrieved
	Relevant	TP (True Positive) Number of Aspect categories extracted correctly	FP (False Positive) Number of Aspect Categories that are annotated but not extracted
	Irrelevant	FN (False negative) Number of Aspect Categories that are not annotated, but extracted by the algorithm	TN (True negative) Number of Aspect Categories that are not annotated, and not extracted by the algorithm

The confusion matrix for opinion term extraction, categorization, and sentiment classification is defined as illustrated in Table 10, Table 11, and Table 12. Patterns categorized under 3 classes of section II have been analyzed by comparing the predicted terms with term manually annotated lexicon and standard evaluation matrices were used to measure the performance.

[7](Layers=12, Hiddensize=768, Attention Head =12, Total Parameters=110M)

TABLE 12. Aspect Sentiment Classification Confusion Matrix.

4*Actual (Annotated Dataset)	Predicted (Aspect Sentiment Classification Approach)		
		Retrieved	Not Retrieved
	Relevant	TP (True Positive) Number of Polarities extracted correctly	FP (False Positive) Number of Polarities that are annotated but not extracted
	Irrelevant	FN (False negative) Number of Polarities that are not annotated, but extracted by the algorithm	TN (True negative) Number of Polarities that are not annotated, and not extracted by the algorithm

TABLE 13. Result Analysis for Opinion Term1 Extraction.

Baseline	Accuracy	Precision	Recall	F1-Score
RuleBased	0.4021	0.40	0.59	0.48
SVM	0.589	0.229157	0.239999	0.226934
KNN	0.549	0.186983	0.226842	0.198111
Naive Bayse	0.334	0.035079	0.039958	0.034244
Proposed Method	0.814	0.8145	0.999	0.9987

"KNN": K nearest neighbor, "SVM": Support Vector Machine

TABLE 14. Result Analysis for Opinion Term2 Extraction.

Baseline	Accuracy	Precision	Recall	F1-Score
RuleBased	0.4021	0.40	0.59	0.48
SVM	0.058	0.0175	0.0140	0.0122
KNN	0.0348	0.0168	0.0149	0.01169
Naive Bayse	0.044	0.0013	0.0040	0.000945
Proposed Method	0.752	0.893	0.8336	0.8336

"KNN": K nearest neighbor, "SVM": Support Vector Machine

V. RESULTS AND DISCUSSIONS

A. RULE BASED APPROACH

The effectiveness of our proposed rule-based lexicon integrated approach for opinion term extraction was rigorously evaluated by comparing it against the method proposed by [26] as well as other state-of-the-art machine learning algorithms. Detailed performance metrics, including precision, recall, and F1-score, are presented in Table 13 and Table 14, providing a comprehensive overview of the results. This comparative analysis highlights the superior performance of our proposed method, demonstrating its relative strengths and improved outcomes across various evaluation benchmarks.

B. MACHINE LEARNING AND DEEP LEARNING MODELS

To evaluate aspect categorization and sentiment classification, we employ a variety of supervised machine learning and deep learning algorithms, including Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Naive Bayes, XGBoost, and Long Short-Term Memory (LSTM) networks.

For aspect categorization, our machine learning models achieved an accuracy of 0.75% using SVM, while our deep learning models achieved an accuracy of 0.76%. For sentiment classification, the machine learning models, with SVM, again leading, attained an accuracy of 0.91%, whereas the deep learning models reached an accuracy of 0.92%. These results demonstrate the effectiveness of both traditional machine learning and modern deep learning approaches in handling these natural language processing tasks. Table15 and Table 16 present a detailed analysis of the results for aspect categorization and sentiment classification, respectively.

TABLE 15. Result Analysis for Aspect Categorization.

Methods	Accuracy	Precision	Recall	F1-Score
SVM	0.749100	0.369932	0.254867	0.282069
KNN	0.701080	0.306901	0.282432	0.291495
Naive Bayse	0.650660	0.154322	0.090392	0.090216
XGBoost	0.687875	0.276058	0.039958	0.231925
LSTM	0.76	0.75	0.76	0.75

TABLE 16. Result Analysis for Sentiment Classification.

Methods	Accuracy	Precision	Recall	F1-Score
SVM	0.91	0.495	0.495	0.495
KNN	0.91	0.467	0.467	0.467
Naive Bayse	0.88	0.486	0.4867	0.4867
XGBoost	0.75	0.477	0.477	0.477
LSTM	0.914	0.912	0.914	0.913

TABLE 17. Result Analysis for BERT Model.

Methods	Aspect catgerorization	Aspect sentiment classifcation
Accuracy	0.80	0.79
Precision	0.79	0.78
Recall	0.80	0.79
F1-Score	0.78	0.70

C. BERT MODEL RESULTS

In the BERT fine-tuning process, we configured the hyperparameters as a maximum sequence length of 150 tokens, 5 training epochs, a batch size of 32, and a learning rate of 1e-5. These settings were chosen to optimize the model's performance on our specific tasks. Detailed performance metrics, including precision, recall, and F1-score, are presented in Table 17, providing a comprehensive evaluation of the model's effectiveness.

VI. CONCLUSION AND FUTURE WORK

Aspect-based sentiment analysis (ABSA) is considered a formidable challenge within sentiment analysis research. It is crucial to ensure that all feedback is accurately extracted and categorized so that service-oriented companies can effectively respond to their customers' needs.

To address this, we employed a lexicon and pattern-based approach to extract implicit aspects. Patterns were designed using Part-of-Speech (POS) tagging and dependency parsing, and our proposed method significantly outperformed state-of-the-art machine learning models. Once extracted, these aspects were categorized and the corresponding sentiment was assigned using both machine learning and deep learning models.

In addition to traditional machine learning and deep learning models, we fine-tuned BERT in a pipeline fashion to carry out end-to-end ABSA tasks. Among the machine learning models, Support Vector Machine (SVM) yielded the best results. However, the performance was significantly enhanced by fine-tuning BERT. We tested our proposed approach exclusively on mobile app reviews of service-oriented organizations.

In the future, this approach can be applied to datasets from various service-oriented domains to evaluate its generalizability and effectiveness across different contexts. Moreover, we plan to increase the size of the dataset to achieve better results by having more diverse and representative data for training and evaluation.

Another limitation of our current study is that we focused solely on English-language reviews. Future research could extend this approach to reviews in different languages, enabling a broader application and enhancing its utility in a multilingual context. This expansion would help capture insights and sentiments expressed in various languages, thus making the analysis more inclusive and applicable across global markets.

REFERENCES

[1] R. Bensoltane and T. Zaki, "Aspect-based sentiment analysis: An overview in the use of Arabic language," *Artif. Intell. Rev.*, vol. 56, no. 3, pp. 2325–2363, Mar. 2023.

[2] F. P. Tajudeen, N. I. Jaafar, and S. Ainin, "Understanding the impact of social media usage among organizations," *Inf. Manage.*, vol. 55, no. 3, pp. 308–321, Apr. 2018.

[3] A. J. Salem Al Mukhaiti, S. Siddiqui, and K. Shaalan, "Dataset built for Arabic sentiment analysis," in *Proc. Int. Conf. Adv. Intell. Syst. Inform.* Cham, Switzerland: Springer, 2017, pp. 406–416.

[4] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Human Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[5] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2004, pp. 168–177.

[6] K. W. Trisna and H. J. Jie, "Deep learning approach for aspect-based sentiment classification: A comparative review," *Appl. Artif. Intell.*, vol. 36, no. 1, Dec. 2022, Art. no. 2014186, doi: 10.1080/08839514.2021.2014186.

[7] O. Alqaryouti, N. Siyam, and K. Shaalan, "A sentiment analysis lexical resource and dataset for government smart apps domain," in *Proc. Int. Conf. Adv. Intell. Syst. Inform.*, 2019, pp. 230–240.

[8] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 845–863, Apr. 2022.

[9] J. Z. Maitama, N. Idris, A. Abdi, L. Shuib, and R. Fauzi, "A systematic review on implicit and explicit aspect extraction in sentiment analysis," *IEEE Access*, vol. 8, pp. 194166–194191, 2020.

[10] B. R. Bhamare and J. Prabhu, "A supervised scheme for aspect extraction in sentiment analysis using the hybrid feature set of word dependency relations and lemmas," *PeerJ Comput. Sci.*, vol. 7, p. e347, Feb. 2021.

[11] L. You, F. Han, J. Peng, H. Jin, and C. Claramunt, "ASK-RoBERTa: A pretraining model for aspect-based sentiment classification via sentiment knowledge mining," *Knowl.-Based Syst.*, vol. 253, Oct. 2022, Art. no. 109511. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705122007584

[12] A. S. Shafie, N. M. Sharef, M. A. Azmi Murad, and A. Azman, "Aspect extraction performance with POS tag pattern of dependency relation in aspect-based sentiment analysis," in *Proc. 4th Int. Conf. Inf. Retr. Knowl. Manage. (CAMP)*, Mar. 2018, pp. 1–6.

[13] S. Singh, H. Kaur, R. Kanozia, and G. Kaur, "Empirical analysis of supervised and unsupervised machine learning algorithms with aspect-based sentiment analysis," *Appl. Comput. Syst.*, vol. 28, no. 1, pp. 125–136, Jun. 2023, doi: 10.2478/acss-2023-0012.

[14] T. Hercig, T. Brychcín, L. Svoboda, M. Konkol, and J. Steinberger, "Unsupervised methods to improve aspect-based sentiment analysis in Czech," *Computación y Sistemas*, vol. 20, no. 3, pp. 365–375, Sep. 2016.

[15] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Boston, MA, USA: Springer, 2012, pp. 415–463. [Online]. Available: https://api.semanticscholar.org/CorpusID:15257275

[16] A. K. Samha, Y. Li, and J. Zhang, "Aspect-based opinion extraction from customer reviews," 2014, *arXiv:1404.1982*.

[17] C. Ojha, M. Venugopalan, and D. Gupta, "A rule based approach for aspect extraction in Hindi reviews," *J. Intell. Fuzzy Syst.*, vol. 41, pp. 1–9, Mar. 2021.

[18] A. K. Samha. (2016). *Aspect-based Opinion Mining Using Dependency Relations*. [Online]. Available: https://api.semanticscholar.org/CorpusID

[19] M. Tubishat, N. Idris, and M. A. M. Abushariah, "Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges," *Inf. Process. Manage.*, vol. 54, no. 4, pp. 545–563, Jul. 2018.

[20] V. Bhatnagar, M. Goyal, and M. A. Hussain, "A proposed framework for improved identification of implicit aspects in tourism domain using supervised learning technique," in *Proc. Int. Conf. Adv. Inf. Commun. Technol. Comput. (AICTC)*, 2016, pp. 1–4.

[21] P. K. Soni and R. Rambola, "A survey on implicit aspect detection for sentiment analysis: Terminology, issues, and scope," *IEEE Access*, vol. 10, pp. 63932–63957, 2022.

[22] S. Areed, O. Alqaryouti, B. Siyam, and K. Shaalan, "Aspect-based sentiment analysis for Arabic government reviews," in *Recent Advances in NLP: The Case of Arabic Language*. Cham, Switzerland: Springer, Jan. 2020, pp. 143–162.

[23] S. Behdenna, G. Belalem, and F. Barigou, "An ontology-based approach to enhance explicit aspect extraction in standard Arabic reviews," *Int. J. Comput. Digit. Syst.*, vol. 11, no. 1, pp. 277–287, Jan. 2022.

[24] H. Cai, R. Xia, and J. Yu, "Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, 2021, pp. 340–350.

[25] S. S. Eldin, A. Mohammed, A. S. Eldin, and H. Hefny, "An enhanced opinion retrieval approach via implicit feature identification," *J. Intell. Inf. Syst.*, vol. 57, no. 1, pp. 101–126, Aug. 2021.

[26] S. Poria, N. Ofek, A. Gelbukh, A. Hussain, and L. Rokach, "Dependency tree-based rules for concept-level aspect-based sentiment analysis," in *Proc. Semantic Web Eval. Challenge (SemWebEval)*, vol. 475, May 2014, pp. 41–47.

[27] S. Poria, E. Cambria, L.-W. Ku, C. Gui, and A. Gelbukh, "A rule-based approach to aspect extraction from product reviews," in *Proc. 2nd Workshop Natural Lang. Process. Social Media (SocialNLP)*, Jan. 2014, pp. 28–37.

[28] T. Gaillat, B. Stearns, G. Sridhar, R. McDermott, M. Zarrouk, and B. Davis, "Implicit and explicit aspect extraction in financial microblogs," in *Proc. 1st Workshop Econ. Natural Lang. Process.*, 2018, pp. 55–61.

[29] K. Verma and B. Davis, "Implicit aspect-based opinion mining and analysis of airline industry based on user-generated reviews," *Social Netw. Comput. Sci.*, vol. 2, no. 4, p. 286, Jul. 2021.

[30] H.-Y. Chen and H.-H. Chen, "Implicit polarity and implicit aspect recognition in opinion mining," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 20–25.

[31] M. Sivakumar and U. S. Reddy, "Aspect based sentiment analysis of students opinion using machine learning techniques," in *Proc. Int. Conf. Inventive Comput. Informat. (ICICI)*, Nov. 2017, pp. 726–731.

[32] N. Nikolić, O. Grljević, and A. Kovačević, "Aspect-based sentiment analysis of reviews in the domain of higher education," *Electron. Library*, vol. 38, no. 1, pp. 44–64, Mar. 2020.

[33] F. Bachtiar, W. Paulina, and A. N. Rusydi, "Text mining for aspect based sentiment analysis on customer review: A case study in the hotel industry," Mar. 2020.

[34] H. Li, B. X. B. Yu, G. Li, and H. Gao, "Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews," *Tourism Manage.*, vol. 96, Jun. 2023, Art. no. 104707.

[35] B. Huang, Y. Ou, and K. M. Carley, "Aspect level sentiment classification with attention-over-attention neural networks," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Model. Predict. Behav. Represent. Modeling Simulation*. Cham, Switzerland: Springer, 2018, pp. 197–206.

[36] H. Xu, L. Shu, P. Yu, and B. Liu, "Understanding pre-trained BERT for aspect-based sentiment analysis," in *Proc. 28th Int. Conf. Comput. Linguistics*, D. Scott, N. Bel, and C. Zong, Eds. Barcelona, Spain: International Committee on Computational Linguistics, Dec. 2020, pp. 244–250. [Online]. Available: https://aclanthology.org/2020.coling-main.21

[37] L. Yu and X. Bai, "Implicit aspect extraction from online clothing reviews with fine-tuning BERT algorithm," *J. Phys., Conf. Ser.*, vol. 1995, no. 1, Aug. 2021, Art. no. 012040.

[38] M. M. Abdelgwad, T. H. A. Soliman, and A. I. Taloba, "Arabic aspect sentiment polarity classification using BERT," *J. Big Data*, vol. 9, no. 1, p. 115, Dec. 2022, doi: 10.1186/S40537-022-00656-6.

[39] E. Brunova and Y. Bidulya, "Aspect extraction and sentiment analysis in user reviews in Russian about bank service quality," in *Proc. IEEE 11th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Sep. 2017, pp. 1–4.

[40] N. Ide, *Introduction: The Handbook of Linguistic Annotation*. Dordrecht, The Netherlands: Springer, Jun. 2017, pp. 1–18.

[41] A. Rosenberg and E. Binkowski, "Augmenting the Kappa statistic to determine interannotator reliability for multiply labeled data points," in *Proc. Short Papers XX - HLT-NAACL*. Boston, MA, USA: Association for Computational Linguistics, 2004, pp. 77–80.

[42] B. Ozyurt and M. A. Akcayol, "A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114231. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417420309519

[43] D. Kumar, A. Gupta, V. K. Gupta, and A. Gupta, "Aspect-based sentiment analysis using machine learning and deep learning approaches," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 5s, pp. 118–138, May 2023.

[44] V. Yadav, P. Verma, and V. Katiyar, "E-commerce product reviews using aspect based Hindi sentiment analysis," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2021, pp. 1–8.

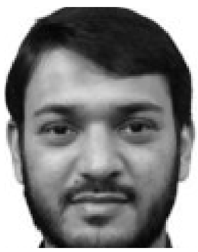
[45] F. A. Naim, "Bangla aspect-based sentiment analysis based on corresponding term extraction," in *Proc. Int. Conf. Inf. Commun. Technol. Sustain. Develop. (ICICT4SD)*, Feb. 2021, pp. 65–69.

[46] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

AFSHEEN MAROOF received the B.S. degree in computer science from The University of Azad Jammu and Kashmir, Pakistan, in 2012, and the M.S. degree in computer science from Muhammad Ali Jinnah University, Karachi, Pakistan, in 2016, where she is currently pursuing the Ph.D. degree. She is also a Lecturer of computer science with DHA Suffa University, Karachi, Pakistan. Her research interests include natural language processing, text classification, machine learning, and deep learning.

SHAUKAT WASI received the degree in computer science from the University of Karachi, and the master's and Ph.D. degrees in computer science from the FAST-National University of Computer and Emerging Sciences (NUCES). He started his professional career at FAST. He was one of the founding faculty members of the Computer Science Department with DHA Suffa University, Karachi. Currently, he is a Professor and the Director of ORIC, Muhammad Ali Jinnah University (MAJU), Karachi. He did his intermediate from Cadet College Petaro, in 1998. He has expertise in text classification and mining, information retrieval and extraction, and human–computer interaction. He is heading the Interactive and Intelligent Natural Language Processing (IINLP) Research Group, FOC, MAJU. He has published 19 publications in local and international conferences and journals. He is honored to be a Program Evaluator for the National Computing Education Accreditation Council (NCEAC), Pakistan. He is with FOC, MAJU, Karachi has planned an international computing conference in collaboration with the IEEE Karachi Section, in 2021.

SYED IMRAN JAMI received the B.S. degree in computer science from the University of Karachi in 2000, the M.S. degree in computer science from Lahore University of Management Sciences in 2004, and the Ph.D. degree in computer science from the National University of Computer Emerging Sciences in 2011. He is currently a Professor with the Department of Computer Science, Muhammad Ali Jinnah University Karachi. Also, he is one of the founding members of the Centre for Research in Ubiquitous Computing and has been associated with it since 2006. He also worked with the Haptics Research Laboratory and the Pervasive and Networked Systems Research Group at Deakin University, Australia. He has authored 16 journal papers and ten conference papers. He worked on several funded research projects and supervised 12 graduate and Ph.D. students.



MUHAMMAD SHOAIB SIDDIQUI (Member, IEEE) received the B.S. degree from the Department of Computer Sciences, University of Karachi, in 2004, and the M.S. and Ph.D. degrees in computer engineering from Kyung Hee University, South Korea, in 2008 and 2012, respectively. Currently, he is an Associate Professor with the Islamic University of Madinah, Saudi Arabia. His research interests include routing, security, management in wireless networks, sensor networks, IP traceback, secure provenance, blockchain technologies, and remote monitoring using the IoT. He is a member of ACM.