Received 9 July 2023, accepted 15 August 2023, date of publication 21 August 2023, date of current version 29 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3307308

TOPICAL REVIEW

Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions

HUANG HUANG, ADELEH ASEMI ZAVAREH, (Member, IEEE), AND MUMTAZ BEGUM MUSTAFA, (Member, IEEE)
Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

Corresponding author: Mumtaz Begum Mustafa (mumtaz@um.edu.my)

This work was supported by the Ministry of Higher Education under the Fundamental Research Grant Scheme under Grant FRGS/1/2020/ICT09/UM/02/1.

ABSTRACT Sentiment analysis (SA), also referred to as opinion mining, has become a widely used real-world application of natural language processing in recent times. Its main goal is to identify the hidden emotions behind the plain text. SA is especially useful in e-commerce fields, where comments and reviews often contain a wealth of valuable business information that has great research value. The objective of this study is to examine the techniques used for SA in current e-commerce platforms as well as the future directions for SA in e-commerce. After examining the existing systematic review papers, it was found that there is a lack of a single comprehensive review paper that addresses research questions. The findings of this study can provide researchers in the field of SA with a comprehensive understanding of the current techniques and platforms utilized, as well as provide insights into the future directions. Through the utilization of specific keywords, we have identified 271 papers and have chosen 54 experimental papers for review. Among these, 26 papers (representing 48.%) have exclusively employed machine Learning techniques, while 24 (44.%) have looked into addressing SA through deep learning techniques, and 4 (7.%) have employed a hybrid approach using both machine learning and deep learning techniques. Additionally, our review revealed that Amazon and Twitter emerged as the two most favored data sources among researchers. Looking ahead, promising research avenues to include the development of more universal language models, aspect-based SA, implicit aspect recognition and extraction, sarcasm detection, and fine-grained sentiment analysis.

INDEX TERMS Sentiment analysis(SA), E-commerce, natural language processing, machine learning, deep learning, opinion mining.

I. INTRODUCTION

With the rapid development of smartphones, internet users are changing from traditional information receivers to information publishers. It has been revealed that e-commerce (EC) websites offer a significant volume of valuable information that surpasses the cognitive processing abilities of humans [3]. Zhang et al. mentioned that micro-blogs comprise complex and copious sentiments that depict the user's perspectives or viewpoints regarding a particular subject [10].

Sentiment analysis (SA) is becoming a necessity as it enables businesses to gain valuable insights from the customer's reviews about their products. Product reviews are a rich source of customer feedback and opinions, and analyzing the sentiments expressed in these reviews provides valuable information about how customers perceive and react to products.

Sentiment analysis is widely used for product reviews as it provides a comprehensive understanding of customer feedback. By analyzing sentiments expressed in reviews, businesses can assess product quality, gain insights into customer preferences, and identify areas for improvement. Sentiment analysis also enables competitive analysis, supporting

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

businesses in understanding their position in the market. Additionally, it aids in decision-making processes, guiding product development, marketing strategies, and customer service improvements. Effective SA further assists in brand reputation management by promptly addressing negative feedback and maintaining a positive brand image.

Existing SA techniques encompass rule-based, lexicon-based, machine learning, and deep learning techniques. Rule-based techniques utilize predefined rules or dictionaries, while lexicon-based techniques rely on sentiment lexicons or dictionaries with annotated scores. Machine learning algorithms like Support Vector Machines(SVM), Naive Bayes(NB), and Random Forest to learn sentiment patterns from labeled data. Deep learning techniques like Recurrent Neural Network(RNN) and Convolutional Neural Networks(CNN), excel at capturing complex patterns and context. These techniques provide valuable customer insights, support reputation management, aid in market research, enhance customer service, and optimize brand marketing efforts.

This study reviews papers published within the last five years (from 2018 to 2022), exclusively narrowing the discussion to machine learning and deep learning techniques. Despite the growing popularity of deep learning, existing research has yet to compare the efficiency of machine learning and deep learning techniques for e-commerce sentiment analysis. This study aims to address this gap by offering an overview of different techniques utilized in e-commerce sentiment analysis. Besides, this study aims to fill the gap in research by summarizing the most popular e-commerce platforms and identifying future directions of sentiment analysis in e-commerce, as no previous study has done so.

In this study, section II of the paper delves into a review of prior research on the application of sentiment analysis in e-commerce. Section III of this research focuses on the methodology employed to review the techniques utilized in sentiment analysis for e-commerce, while Section IV contains the results and pertinent discussions for each research question. Lastly, Section V provides a summary of the conclusions drawn from this review study.

II. RESEARCH BACKGROUND

A. EXISTING REVIEWS STUDIES ON SA IN E-COMMERCE

Recent studies primarily focus on machine learning techniques for multi-lingual SA [55], SA in Arabic [54], recent approaches of implicit aspect extraction for SA [52], SA in social media and its application [56], application of SA using machine learning techniques [57] and comparison of lexicon-based and Bidirectional Encoder Representations from Transformers (BERT) based SA in Italian [30]. Despite numerous studies on SA applied in e-commerce, there is currently no comprehensive work that summarizes the different techniques used, provides an overview of e-commerce data platforms, and suggests potential directions for future.

Some existing review papers focus on consolidating various machine-learning approaches to summarize SA techniques. Sagnika et al. [55] conducted a review study focusing on machine-learning techniques for multilingual SA. Mehta and Pandya [53] summarized various papers utilizing machine learning and lexicon analysis approaches. Shathik and Prasad [57] introduced the prevalent techniques used in SA from a machine learning perspective. Umar et al. [60] investigated the sentiment classification level or data source on which supervised machine learning techniques like SVM, NB, Maximum Entropy, and other technique such as lexicon-based which deliver optimal results in Shathik and Prasad [57] and Umar et al. [60] solely encompassed a review focusing on machine learning techniques. It is widely acknowledged that in addition to machine learning techniques, deep learning techniques remain effective for addressing SA problems. Thus, there is a need for a more comprehensive review that covers both machine learning and deep learning techniques.

Other works offer a more thorough and organized examination of machine learning techniques and deep learning approaches utilized in SA. Catelli et al. [30] compare lexicon-based approaches and machine/deep learning techniques in the context of SA, specifically focusing on the Italian market. Shayaa et al. [59] provided an overview of the publication trends in opinion mining and SA from 2000 to 2016. Given the swift progress of technological advancements, it is important to conduct a study that reviews papers published between 2018 and 2022, offering a more current viewpoint than earlier review studies.

Some of the review papers looks at the SA applications. Drus and Khalid [56] presented a comprehensive review report on SA in social media, which investigates the techniques employed, social media platforms utilized, and their applications. Baragash and Aldowah [61] conducted a systematic review to explore the recent application of SA in higher education. The review aimed to categorize and identify the SA techniques and techniques commonly and effectively employed within the higher education domains. Obiedat et al. [54] focused on an Arabic SA review of the application of SA in social media, higher education domains, and the Arabic texts approach. In addition, our study aims to investigate the utilization of SA, specifically in e-commerce.

The next group of review papers offers innovative ideas from a technical perspective. Yue et al. [62] focus on typical techniques in the social media field of SA from three distinct perspectives: task-oriented, granularity-oriented, and methodology-oriented. Aspect extraction, is a component of aspect-based SA, involves identifying explicit and implicit aspects of a text. While explicit aspects are directly mentioned, implicit aspects need to be inferred. Detecting implicit aspects poses challenges, yet it is crucial. Unfortunately, there is limited research dedicated to the extraction of implicit aspects. Ganganwar and Rajalakshmi [52] surveyed recently proposed techniques for detecting implicit aspects. Qazi et al. [65] examined the evidence presented in different reviews and outlined the challenges encountered by classification techniques in SA. They aim to comprehend how implementing enhanced techniques can address the conventional issues associated with classification techniques. These three related papers examine various aspects of SA techniques, including different orientations, aspect extraction techniques, and classification techniques, all from a technical standpoint. The primary limitation of these three papers is the absence of a comprehensive summary regarding data sources and future directions in SA.

Others have also conducted similar research papers exploring SA in e-commerce. Elzeheiry et al. [64] proposed an overview of the initial design for e-commerce is presented. Moreover, they discuss deep learning in e-commerce and SA. Then various versions of the commercial dataset are presented. Lastly, the challenges encountered by recommendation systems and directions for future research are explained. Elzeheiry highlighted the challenges faced and the future directions of recommendation systems. In contrast to their study, our research aims to enumerate specific future research directions in SA. Marong et al. [66] provided a comprehensive overview of SA and relevant techniques in the e-commerce sector, constantly seeking to under-stand consumers' opinions of their products and services. This paper primarily offers a technical perspective overview of SA in e-commerce. In contrast, our research aims to delve into the discussion of data sources and explore future directions in this context.

Despite numerous studies on SA applied in e-commerce, no comprehensive work currently summarizes the different techniques(including machine learning and deep learning techniques) used over the past five years, provides an overview of e-commerce data platforms, and suggests potential directions for future research.

B. IMPORTANCE OF THIS REVIEW STUDY

We cannot overstate the commercial value of SA in e-commerce. With the explosion of e-commerce in recent years, businesses need to be able to analyze and understand customer feedback on a massive scale. Through SA, businesses can identify patterns and trends in customer sentiment, such as which products or services are popular, which are not, and why. This information can help businesses make informed decisions about product development, marketing strategies, and customer service. Additionally, SA can help e-commerce businesses to stay ahead of their competition by identifying emerging trends and customer needs. The most prevalent form of feedback provided by customers on current e-commerce platforms is through comments. Extracting insights from text analysis can serve as a useful reference for other consumers and enable businesses on e-commerce platforms to enhance service quality and boost customer satisfaction [12].

E-commerce platforms primarily consist of natural language sentences in their text or comments, which means that SA is ubiquitous wherever text is present. From this study, we can get an overview of the most suitable technologies used in SA of e-commerce. Different approaches are available, like machine learning techniques (eg. SVM, Naive Bayes) require less data but more human intervention. Machine learning is currently the most convenient technique to train neural networks in the era of artificial intelligence. Deep learning techniques such as Long Short-Term Memory (LSTM), and BERT do not require feature engineering but rely on extensive data [16]. Deep learning techniques have demonstrated significant advancements by overcoming the sequential limitations of prior models and introducing operational parallelism, resulting in several advantages such as context analysis, making the constituent embeddings more dynamic.

This paper aims to provide an overview of the current SA techniques in e-commerce, which will facilitate researchers in obtaining a comprehensive understanding of the employed technologies [30].

In addition, this study summarizes the top e-commerce platforms that offer raw data, which can be analyzed to determine the most popular and promising industries. By identifying the primary sources of e-commerce data, such as Amazon, Amazon is among the largest e-commerce platforms, with many reviews available for viewing. These findings can also be applied to SA in similar platforms or industries, including social media platforms such as Twitter and Instagram, and even anti-terrorism work. Moreover, this research provides a valuable resource for future researchers to locate the appropriate data sources quickly.

In our review, we identified various research gaps and challenges that researchers would encounter in future work, such as applying aspect-wise SA in e-commerce. With the significant rise in user comments and reviews on social media and e-commerce websites, there has been a corresponding increase in the need to identify sentiment at the sentence and aspect levels.On review boards, it is common to encounter comments that are challenging to categorize as positive or negative because of the presence of multiple aspects [54]. For example, comments such as ''The restaurant has a nice environment, but the food is terrible'' has contrasting polarities for different aspects. The utilization of technology for aspect-wise SA in e-commerce remains a complex problem that requires further attention.

In practice, there are several promising research directions for the future, including more universal models in new domains and languages, Aspect-level SA models, implicit Aspect Recognition and Extraction, Sarcasm detection, Lexicon based models introducing external knowledge, fine-grained SA. These topics will be discussed in section IV later.

III. REVIEW METHOD

This study aims to thoroughly examine the literature on SA in the context of e-commerce. The primary aim of this review is to assess the present state of research in this area, focusing on multiple aspects such as research objectives, the online shopping platforms analyzed, state-of-the-art SA techniques employed, and diverse evaluation metrics employed to measure the effectiveness of the methodologies. In addition to highlighting the current issues and challenges in SA for e-commerce, this study also endeavors to suggest practical solutions and future directions for researchers in this field. By leveraging the vast array of resources and data available on online shopping platforms, this study aims to identify trends, patterns, and key insights that can inform future research and guide the development of more effective and accurate SA models. This study aims to contribute to the advancement of SA in e-commerce and help researchers and practitioners to better understand the challenges and opportunities ahead. Through a comprehensive review of existing works and a critical analysis of the current state of research, this study aims to provide a valuable resource for those seeking to make significant contributions to this exciting and rapidly evolving field.

There are three stages in the review process: Research questions raised in this review, Search methodology, Search results and analysis.

A. RESEARCH QUESTIONS RAISED IN THIS REVIEW STUDY

Here are the research questions to be addressed in this research:

Research Question 1: What are the current techniques used for sentiment analysis in e-commerce?

Research Question 2: What are the current e-commerce platforms that apply sentiment analysis?

Research Question 3: What are the future directions for sentiment analysis in e-commerce?

B. SEARCH METHODOLOGY

Here, we employed a comprehensive search methodology that encompassed various online databases and a meticulous scrutiny of the chosen literature. Our search spanned across a broad range of prominent databases such as:

- Science Direct;
- IEEE;
- Google Scholar.

Furthermore, we conducted a manual assessment in which we examined the title and abstract of the papers. Subsequently, we chose the relevant papers while disregarding any that were deemed irrelevant.

The key words are as follows:

- SA;
- e-commerce;
- opinion mining;
- machine learning;
- deep learning.

We have utilized the subsequent inclusion criteria in order to filter our initial search results and ensure that only pertinent studies are considered in our analysis.

- Search domain: Artificial Intelligence, Computer Science or Information Technology
- Publication type: Journals;
- Paper type: Full text;
- Language: English.

Here are the exclusion criteria were utilized to filter out irrelevant papers.

- Papers that do not specifically focus on SA and e-commerce.
- Papers that cover the topic of e-commerce with SA as a supplementary subject.
- Paper without detailed experiments

A collection of 54 papers published between 2018 to 2022 have been sourced from various databases. The databases that were manually searched to procure these papers include Science Direct, IEEE, and Google Scholar. The papers have been handpicked to ensure that they meet the relevance to the aims of the review. The selected papers provide the depth in SA topics, including cutting-edge technologies, innovative techniques, and groundbreaking research findings, making it an invaluable resource for researchers, scholars, and practitioners in the computer industry.

C. SEARCH RESULTS AND ANALYSIS

Here, we will provide the results of the search and selection process, along with an examination of the chosen experimental papers concerning the publication databases utilized and the explored techniques for conducting SA.

After conducting a keyword search, we limited the papers to only those published between 2018 and 2022, identifying 271 papers. We manually analyzed the titles and abstracts of the shortlisted papers to remove irrelevant ones, resulting in the removal of 139 papers and leaving us with 132 papers. We then read the full papers and applied the exclusion criteria, which resulted in the rejection of 78 papers and left us with 54 relevant papers. The final number of papers selected for the review study was 54.

As to the source databases, 16 articles (30.%), were obtained from the IEEE database. 34 papers (63.%) were retrieved from Google Scholar, while 2(4.%) others from Science Direct database.

From Figure 1 below, it is clear that out of the 54 selected papers, 26 (48.%) employed machine learning techniques, while 24 (44.%) exclusively utilized deep learning techniques. Additionally, 4 papers (7.%) use both machine and deep learning techniques. These findings indicate that machine learning and deep learning techniques hold equal weight in the realm of SA.

IV. RESULTS AND DISCUSSION

A. RESEARCH QUESTION: WHAT ARE THE CURRENT TECHNIQUES USED FOR SENTIMENT ANALYSIS IN E-COMMERCE?

SA, also referred to as opinion mining, is a practical application of natural language processing (NLP) aims to identify the hidden emotion behind the plain text. Chandio et al. [46] stated that SA falls under the umbrella of text classification,

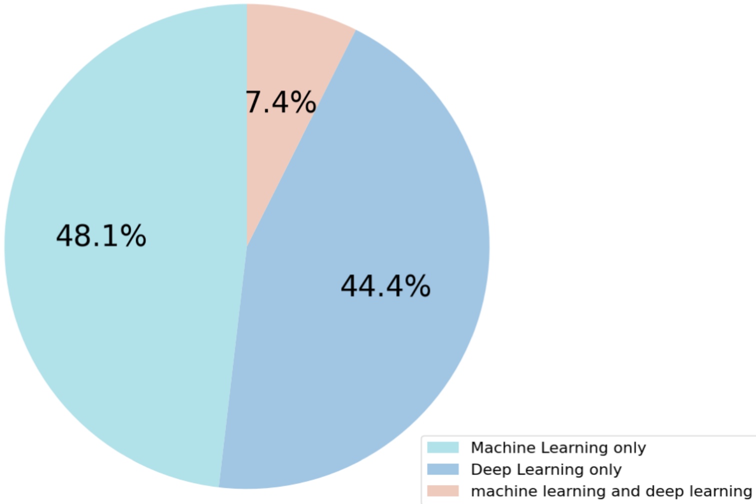

FIGURE 1. Proportion of machine learning and deep learning techniques.

specifically focusing on classifying user sentiments into positive, negative, or neutral categories.

In natural language processing tasks, SA is crucial, and there are various ways to approach it. In previous studies, deep learning techniques would be considered a subset of machine learning techniques. However, because of its differences with other machine learning techniques, we have separated deep learning techniques and other machine learning techniques in this study. In this study, all other machine learning techniques except deep learning techniques are called machine learning techniques.

Machine learning techniques rely on statistical techniques to learn from data and make predictions. These techeniques use features extracted from the text, such as word frequencies, sentiment lexicons, and grammatical structures, to train a model that can classify the sentiment of a given text. Jagdale et al. [4] indicated that NB and SVM are among the popular machine learning-based algorithms utilized in SA.

In contrast, deep learning based techniques leverage neural networks to automatically learn representations of the text and in recent time shown encouraging result in SA. These deep learning algorithms can capture complex relationships between words and can handle large volumes of data without the need for manual feature [13]. Deep learning techniques used in SA includes CNN, RNN, LSTM, Transformer, BERT and Generative Pre-Trained (GPT) models.

From Figure 1, the majority of research papers rely y on machine learning techniques, where 48% of the papers utilizes machine learning for SA. Furthermore, the trend towards deep learning methodologies is rising, with 44% of the papers employing this state-of-the-art technique. Notably, machine learning and deep learning techniques were applied in the remaining 8% of the papers. Notably, both machine and deep learning techniques are commonly used in SA in e-commerce fields.

1) MACHINE LEARNING BASED TECHNIQUES

Machine Learning techniques need human intervention to improve the model performance. Before training, if an inaccurate prediction occurs while training the model, an engineer has to step in and make adjustments. However, machine learning techniques require a smaller data size and less training time than deep learning models. Machine learning models have lower computing complexity than deep learning techniques.

From Figure 2 and Table 1, it is clear that SVM [1,2,4,6,8,10,14,16,19,24-25,31,38-39,46,50,73,75] and NB [2,4,6,8,10,14-16,24-25,34,39-40,45,48,50] are the most welcome machine learning techniques used in selected papers.

Some researchers employ SVM models directly for their studies. Huang et al. [1] use neural network (NN) model and SVM model as their prediction models. In Hantoro et al. [73], SA was performed on Shopee application using SVM classification. The analysis utilized 990 training data samples and 110 test data samples. Out of the total test data of 110 assessments, 28 evaluations were classified as negative, while the remaining 82 evaluations were classified as positive. The accuracy rate was 80.90%, indicating that 89 assessments were correctly classified in their respective sentiment classes. Pratama et al. [75] employed the SVM to classify reviews of e-commerce beauty products. Their objective is to develop a model capable of categorizing beauty product reviews and analyzing the accuracy of the classification. Elmurngi and Gherbi [8] conducted a comparison of four supervised machine learning techniques, namely NB, Decision Tree J48 (DT-J48), Logistic Regression (LR), and SVM, for detecting unfair reviews. In Haque al. [14], the performance of various classification models, including Linear SVM, Multinomial NB, Stochastic Gradient Descent, Random Forest, LR, and DT, was evaluated on Amazon product reviews.

TABLE 1. Techniques comparison used in selected papers.

Methods classes →	Machine Learning Techniques				Deep Learning Techniques							Others	
Methods / Article No.	SVC/SVM	LR	Naïve Bayes	DT	NN	CNN	RNN	GRU	LSTM	Transformer	BERT	Lexicon	Others
[1]	✓				✓							✓	
[2]	✓	✓	✓				✓						
[3]						✓	✓				✓		
[4]	✓		✓									✓	
[5]						✓							
[6]	✓		✓										
[7]									✓				
[8]	✓	✓	✓	✓									
[9]					✓				✓				
[10]	✓		✓									✓	
[11]							✓		✓				
[12]						✓		✓					
[13]							✓	✓					
[14]	✓	✓	✓	✓									
[15]			✓										
[16]	✓		✓										
[17]												✓	
[18]													✓
[19]	✓	✓				✓			✓				
[20]								✓					
[21]											✓		
[22]							✓	✓					
[23]						✓			✓				
[24]	✓	✓	✓										
[25]	✓		✓		✓		✓						
[26]											✓		
[27]							✓						
[28]		✓											
[29]										✓			
[30]											✓	✓	
[31]	✓												
[32]													✓
[33]													✓
[34]			✓	✓									
[35]						✓			✓				
[36]								✓			✓		
[37]													✓
[38]	✓												
[39]	✓		✓										
[40]			✓										
[41]								✓					
[42]													✓
[43]					✓								
[44]												✓	
[45]			✓										
[46]	✓												
[47]									✓				
[48]			✓										
[49]									✓				
[50]	✓		✓										
[73]	✓												
[74]						✓	✓					✓	
[75]	✓												
[76]									✓				
$\sum$	18	6	16	3	4	7	8	6	9	1	5	7	5

TABLE 2. E-commerce platforms used in selected papers.

Platforms / Article No.	Amazon	Ctrip	FLipkart	eBay	Walmart	Twitter	Alibaba	IMDB	Traveloka	TripAdvisor	dangdang	Others
[1]												✓
[2]	✓											
[3]	✓	✓	✓	✓	✓							
[4]	✓											
[5]							✓					✓
[6]	✓											
[7]									✓			
[8]	✓											
[9]						✓						
[10]											✓	
[11]												✓
[12]											✓	
[13]	✓											
[14]	✓											
[15]												✓
[16]	✓											
[17]	✓											
[18]	✓											
[19]												✓
[20]												✓
[21]	✓											
[22]	✓											
[23]								✓				
[24]												✓
[25]	✓											
[26]												✓
[27]												✓
[28]								✓				
[29]												✓
[30]	✓											
[31]	✓											
[32]	✓											
[33]						✓						
[34]						✓						✓
[35]										✓		
[36]								✓				
[37]				✓			✓					
[38]												✓
[39]												✓
[40]									✓			
[41]												✓
[42]						✓						
[43]												✓
[44]						✓						
[45]						✓						
[46]												✓
[47]		✓										
[48]												✓
[49]								✓				
[50]												✓
[73]												✓
[74]												✓
[75]	✓											
[76]												✓
$\sum$	17	2	1	2	1	6	2	4	2	1	2	21

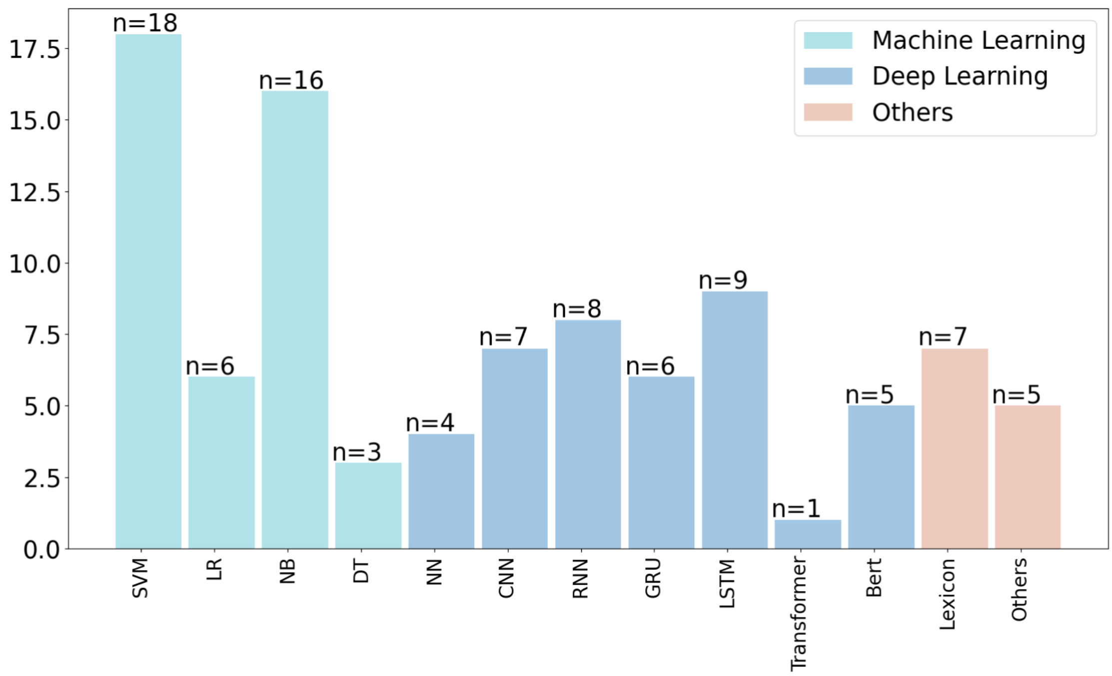

FIGURE 2. Techniques used in selected papers.

The findings revealed that SVM outperformed other techniques and achieved the most accurate classification results. Dey et al. [16] conducted a comparison between two machine learning techniques to analyze the sentiment of customer reviews on Amazon products. Experimental results have validated that SVM exhibits a higher accuracy rate in polarizing the feedback of Amazon products.

Linear Support Vector Classification (LinearSVC), which is an alternative implementation of SVM algorithms, also demonstrates superior performance compared to other machine learning algorithms. Ahmed et al. [2] employed applied linear SVC, Naive Bayes, and LR for SA. The results indicated that the linear SVM outperformed other classifiers. A medicine recommendation system was developed by Garg [24] utilizing patient reviews to predict sentiment. Various machine learning techniques were employed, and the results indicate that the Linear SVC classifier, combined with TF-IDF vectorization, outperforms all other models with an accuracy of 93%.

However, the performance could be better when comparing SVM with deep learning techniques like RNN. Dashtipour et al. [19] propose a novel hybrid framework for concept-level SA for Persian. The framework integrates linguistic rules and deep learning techniques to enhance polarity detection and optimize analysis. Their proposed framework performs superior to state-of-the-art approaches, including SVMs and LR. Mukherjee et al. [25] incorporated a tailored negation marking algorithm to detect explicit negation, and they conducted experiments using various machine learning algorithms such as NB, SVMs, ANN, and RNN for SA of Amazon, specifically focusing on smart phones. By assessing the impact of the negation algorithm on SA tasks, they found that the RNN, in combination with our negation marking processing, achieved the highest accuracy of 95.67%.

Many studies compared NB techniques and various other classification techniques. Jagdale [4] utilized different machine learning techniques to classify Amazon reviews, resulting in an accuracy of 98.17% for Naive Bayes and 93.54% for SVM, specifically in the case of Camera Reviews. Basani et al. [6] applied NB and SVM techniques to Amazon reviews datasets, and their accuracy and execution time were compared. The results demonstrate that the SVM outperforms the NB in terms of accuracy, and the SVM also exhibits shorter execution time than NB. In Bayhaqy'study [34], the performance of DT, K-Nearest Neighbor (KNN), and NB was compared on Tweets datasets. The highest result obtained was from the NB techniques, achieving an accuracy of 77%, precision of 88.50%, and recall of 64%. Lutfi et al. [39], performed SA on sales reviews of an Indonesian marketplace, utilizing SVM and NB. The result shows that SVM with a linear kernel has better than NB. Wongkar and Angdresey [45] carried out a comparative study using NB, SVM, and K-NN techniques in RapidMiner. The results revealed an accuracy value of 75.58% for NB, 63.99% for SVM, and 73.34% for K-NN.

SA was not limited to classifying sentiments as positive or negative, but also incorporated for assessing the quality of online shopping platforms. Sari et al. [15], analyzed data from Tokopedia, one of Indonesian largest e-commerce services, to evaluate the quality of its services. The NB classification technique was chosen due to its high accuracy and ability to handle large amounts of data. The findings indicated that the dimensions of personalization and reliability required more attention, as they many negative sentiments. On the other hand, the dimensions of trust and web design received high positive sentiments, indicating excellent service quality. The responsiveness dimension had a balanced mix of positive and negative sentiments. Aagte et al. [48], found that certain parties tried to post fake reviews on competitors' websites. They suggest the integration of NB technique to identify spam reviews. They cover the preprocessing of data prior to aspect identification, which involves classifying reviews as positive, negative, or neutral. Then aspect ranking is performed using NB.

To summarize, NB, DT, LR, and SVM are commonly used techniques in SA. Among them, SVM and NB techniques are popular. Previous studies have compared the performance of different techniques, and SVM and NB have consistently shown better results than others. NB can also be applied in assessing the quality of online shopping platforms and detecting spam. This highlights the versatility of SA. Additionally, some studies have compared SVM and NB with deep learning techniques like RNN, and the results have demonstrated that RNN outperforms other machine learning techniques. The subsequent sections will delve further into the discussion of deep learning techniques.

2) DEEP LEARNING BASED TECHNIQUES

In addition to machine learning techniques, deep learning are extensively employed in SA. A well-behaved deep learning technique can identify whether its prediction is precise or not based on its neural network (without human's help).

In recent studies, BERT [3,21,26,30,36], CNN [3,5,12,19, 23,35,74], RNN [2,3,11,13,22,25,27,74], LSTM [7,9,11,19, 23,35,47,49,76] classification or prediction are the most frequently used. As the model's development, more complicated models such as CNN-based bi-directional long short-term memory (BiLSTM) (CNN-BiLSTM), attentionbased Bidirectional Gated Recurrent Unit (BiGRU) models have also been developed to enhance the precision and efficiency of deep learning models.

Initially employed in computer vision domains, CNN exhibits exceptional performance by emulating the visual perception process of humans. Its application in SA represents a significant innovation, leveraging the strengths of CNN to capture and analyze textual information. Minaee et al. [23] introduces a model that combines LSTM and CNN. The LSTM component is utilized to capture the temporal information of the data, while the CNN component is employed to extract the local structure from the data. CNN can leverage the local correlations and patterns inherent in data through its learned feature maps. This distinctive characteristic of CNN makes it particularly powerful, and it has been successfully introduced to analyze text, enabling the extraction of meaningful features from textual data. Zuheros et al. [35] developed the ABSA model incorporated with both CNN and LSTM. LSTM is utilized for aspect term classification and aspect category classification tasks, while CNN is employed for a polarity classification task. Murfi et al. [74] build upon previous work on hybrid deep learning approaches by incorporating BERT representation for Indonesian SA. Their simulations demonstrate that utilizing BERT representation enhances the accuracies of all hybrid architectures. Particularly, the BERTLSTM-CNN model achieves slightly higher accuracies than other hybrid architectures based on BERT.

RNN imitates the process of human reading because of its ability to capture sequential information and model dependencies over time. Just like humans read and comprehend text by considering the context and previous information, RNNs can process input data sequentially, preserving information from previous steps and incorporating it into the current step.

In SA, various types of RNN and their variations have been widely applied. Muhammad et al. [7] has demonstrated the utilization of Word2Vec and LSTM for sentiment classification in hotel reviews. Agarap [11] utilized univariate and multivariate analyses on various dataset features, excluding review titles and texts. They implemented a bidirectional RNN with LSTM units for recommendation and sentiment classification tasks, and they observed that the bidirectional LSTM achieved an F1-score of 0.88 for recommendation classification and 0.93 for sentiment classification. A bidirectional LSTM is a type of RNN that processes input sequences in both forward and backward directions [11]. It uses two hidden states for each direction, allowing it to capture past and future context when making predictions or analyzing data sequences. This enables the bidirectional LSTM to have a more comprehensive understanding of the input sequence and potentially improve performance in tasks such as sequence.

RNN imitates the process of human reading because of its ability to capture sequential information and model dependencies over time. Just like humans read and comprehend text by considering the context and previous information, RNNs can process input data sequentially, preserving information from previous steps and incorporating it into the current step.

In SA, various types of recurrent neural networks (RNN) and their variations have been widely applied. Muhammad, [7] has demonstrated the utilization of Word2Vec and LSTM for sentiment classification in hotel reviews. Agarap [11] utilized univariate and multivariate analyses on various dataset features, excluding review titles and texts. They implemented a bidirectional recurrent neural network (RNN) with long-short-term memory (LSTM) units for recommendation and sentiment classification tasks, and they observed that the bidirectional LSTM achieved an F1-score of 0.88 for recommendation classification and 0.93 for sentiment classification. A bidirectional LSTM is a type of RNN that processes input sequences in both forward and backward directions. It uses two hidden states for each direction, allowing it to capture past and future context when making predictions or analyzing data sequences. This enables the bidirectional LSTM to have a more comprehensive understanding of the input sequence and potentially improve performance in tasks such as sequence classification or prediction.

Other deep learning algorithms or attention mechanisms are often combined with RNN or its variants for various applications. Meng et al. [9] presented a SA approach called Feature Enhanced Attention CNN-BiLSTM that operates at the aspect level. Their technique utilizes CNN to extract higher-level phrase representations from the embedding layer, enhancing subsequent coding tasks. They employ BiLSTM to capture local phrase features and global and temporal sentence semantics to improve context encoding quality and preserve semantic information. An attention mechanism is also incorporated to model the interaction between aspect words and sentences, enabling an adequate context representation. Xu et al. [47] introduced an enhanced word representation approach incorporating sentiment information into the traditional TF-IDF technique, resulting in weighted word vectors. These weighted word vectors are then fed into a BiLSTM model to effectively capture contextual information and improve the representation of comment vectors. The sentiment tendency of the comments is determined using a feedforward neural network classifier. Iqbal et al. [77] implemented deep learning-LSTM and RNN for sentiment classification and analysis. Three models were introduced, each utilizing different architectural variations and parameter tuning based on these deep-learning techniques. Chandio et al. [46] proposed RU-BiLSTM, a deep recurrent architecture for SA of Roman Urdu. The model combines BiLSTM with word embedding and an attention mechanism. By utilizing BiLSTM, the model captures context information in both directions, while the attention mechanism focuses on essential features. Finally, the last dense softmax output layer is employed to obtain binary and ternary classification results. Huang, Lin, et al. [76] proposed a ERNIE-BiLSTM-Att (EBLA). The technique incorporates the Enhanced Representation through Knowledge Integration (ERNIE) word embedding model to generate dynamic word vectors. These vectors are fed into a BiLSTM network to extract text features. The Attention Mechanism (Att) is applied to refine the weights of the hidden layer. Lastly, softmax is utilized as the output layer for sentiment classification.

Shrestha and Nasoz [13] organized and arranged review embeddings to create a product sequence to enhance the sentiment classification process. This sequence was inputted into a recurrent gated unit (GRU) to learn the product embedding. The review embeddings generated from paragraph vectors and product embeddings from GRU were used to train a SVM for sentiment classification. GRU [13] is another type of RNN architecture used for sequential data analysis, such as natural language processing tasks. GRU has gating mechanisms that enable it to selectively update and forget information from previous time steps allowing it to capture long-term dependencies in the data efficiently.

Figure 2 shows that the employment of various deep learning techniques is relatively evenly distributed among the groups. Among the selected papers, LSTM was the most commonly utilized technique, being employed in 10 out of the 54 papers. Both LSTM and GRU are derived from RNNs, and their recurrence mechanisms for resolving seq2seq issues have demonstrated exceptional performance.

In summary, the dominant deep learning techniques in NLP consist of RNN and its variations, such as LSTM and GRU. While CNN was originally developed for computer vision, it has also found applications in text analysis. RNN algorithms can take different forms, including Bidirectional RNN and Bidirectional LSTM, expanding their capabilities. Additionally, RNN can be effectively combined with other deep learning algorithms or integrated with attention mechanisms to concentrate on specific aspects of textual data.

B. RESEARCH QUESTION: WHAT ARE THE CURRENT E-COMMERCE PLATFORMS THAT APPLY SENTIMENT ANALYSIS?

The data source can significantly influence the accuracy and usefulness of collected data during the data collection. The SA task aims to understand and interpret human opinions on various objects, topics, and events. In this regard, comments and feedback from online shopping platforms have emerged as one of the most abundant and informative data source. By analyzing the data from various sources, we can gain an overview of the different fields where SA is applied. This study has divided the data sources into four categories: online shopping platforms, social media, travel agencies, and other sources. In other words, SA is commonly utilized in popular areas: online shopping, social media, travel industries, and other fields.

Based on Figure 3 and Table 2, it is evident that Amazon [24,6,8,13-14,16-18,21-22,25,30-32,75] is the primary platform from which researchers collect data, with 17 out of 50 papers utilizing it. Amazon, the largest online retailer in the world, is a treasure trove of customer feedback data. Through its platform, customers from all over the globe post their thoughts and opinions on various products and services, creating a massive repository of comments invaluable to data scientists and market researchers.

Social media platforms such as IMDB [23,28,36,49] and Twitter [9,33-34,42,44-45] are also crucial data sources, with 11 out of 50 papers using them. Given the widespread use of social media and other online platforms for purchasing goods and services, analyzing customer feedback and comments has become integral to SA. These comments provide valuable insights into the customer's perspective on a product or service and help businesses identify areas for improvement and enhance customer satisfaction. In addition, applying natural language processing techniques can aid in evaluating the sentiment conveyed in these comments, facilitating the automatic categorization of positive, negative, or neutral sentiments. This data can then be utilized to enhance marketing strategies, product development, and customer service.

Besides the above websites, there are other online platforms where users can express their opinions and share their experiences. Websites such as Ctrip [3,47], TripAdvisor [35] and Traveloka [7,40] are just a few examples of popular platforms where customers can leave reviews and ratings for hotels, restaurants, and tourist attractions. These customer sentiments are invaluable to other travelers looking for information to help plan their trips. By reading reviews, travelers can understand what to expect and make more informed decisions about where to stay, what to eat, and what to see during their travels In other industries, data comes from different platforms but all of them are in service industries platforms including Women's Clothing reviews, comments from dangdang.com, restaurant and laptop reviews from SemEval 2014 twitter dataset, The Alibaba Group and JD, Flipkart eBay, Walmart, BestBuy, gross box-office revenue data for the movies.

C. RESEARCH QUESTION: WHAT ARE THE FUTURE DIRECTIONS FOR SENTIMENT ANALYSIS IN E-COMMERCE?

According to the reviewed papers, many future direction suggestions are given by related authors.

1) MORE UNIVERSAL MODELS IN NEW DOMAINS AND LANGUAGES

According to researchers, a particular machine or deep learning model may only suit a single language. For example, if English was used to train a language model, its performance with Chinese or Arabic could be poor, suggesting that each language requires its unique language model.

In most implemented models, the data source is mainly in English. Therefore, those models can only perform better at predicting English inputs. Conducting cross-cultural studies in SA by obtaining sentiment reviews in languages other than English is still a promising direction.

Researchers have conducted SA in various languages other than English. Dashtipour et al. [19] presented a novel hybrid framework for concept-level SA for Persian language. Their framework combines linguistic rules and deep learning techniques to enhance the polarity detection and achieve better results. Yang et al. [12] proposed a model called SLCABG, which stands for Sentiment Lexicon-based CNN with attention-based Bidirectional Gated Recurrent Unit (BiGRU). It combines the sentiment lexicon with CNN and BiGRU to perform SA. The model is trained using Chinese comments data obtained from dangdang.com. Saleh et al. [67] proposed an optimized heterogeneous stacking ensemble model to improve the performance of Arabic SA. The model is designed to effectively combine multiple techniques to achieve better accuracy and results in analyzing sentiments in Arabic text. Noor et al. [38] focuses on analyzing Urdu Roman reviews obtained from Daraz.pk, one of the most popular and widely accessed e-commerce websites in Pakistan.

SA research focused on languages such as English, Chinese, Persian, Arabic, and Roman Urdu. However, there is a lack of extensive studies and training on other languages such as Japanese, Cantonese, and German text data using machine learning or deep learning techniques. Even in state-ofthe-art models like GPT-4, the support for minority languages remains limited. A universal model which can be applied across domains and languages still has a long way to go.

For one specific machine learning or deep learning model, in most situations, it can only perform well in one specific domain [3]. Hong [5] states that in the subsequent study, online feedback on various types of fresh agricultural products will be considered as the research subject to enhance the generalizability of the experimental outcomes. In Puspita Kencana Sari's study, they clarified a thorough investigation of the data collection process is necessary, particularly regarding the specific duration during which data is collected. To obtain more representative results, additional data source such as social media platforms like Twitter and Facebook should be utilized to evaluate service quality [15]. For instance, if a model is fed and trained by digital products review, it cannot perform well in restaurant service fields. Alhassan Mabrouk states that in the current system, they only concentrated on comparing a group of products belonging to the same category and summarizing their aspect-based opinions. However, they also aim to extend its application to other domains such as movies or restaurants [3]. In Wongkar and Angdresey they intend to examine the sentiment of public satisfaction with the performance of the elected president of the Republic of Indonesia by analyzing data from other social media platforms, such as Facebook and Instagram. A comprehensive language model that can analyze all languages in all domains should be proposed [45].

2) ASPECT-LEVEL SENTIMENT ANALYSIS TECHNIQUES

Previous studies have shown a shift from coarse-grained SA to fine-grained SA. Meanwhile, the research target of SA is changing from document-level SA to sentence-level SA. However, traditional coarse-grained SA cannot meet the requirements for business analysis, more fine-grained SA is a must for more advanced business analysis.

The emergence of aspect-based SA (ABSA) has garnered significant attention from researchers in machine learning and natural language processing. Most research studies in this area are centered around two main directions: neural networks and attention mechanisms. With the integration of attention mechanisms into ABSA, there has been a significant increase in the utilization of hybrid methods that combine neural networks and attention mechanisms. These hybrid approaches have gained widespread popularity as they capitalize on the respective strengths of both techniques.

Han et al. [20] proposed the Pretraining and Multi-task learning based on Double BiGRU (PM-DBiGRU). The PM-DBiGRU model leverages pre-trained weights obtained from a sentiment classification task on short text-level drug reviews to initialize the relevant weights in our model.It utilizes two Bidirectional Gated Recurrent Unit (BiGRU) networks to generate bidirectional semantic representations of the target and drug review. Additionally, an attention mechanism is employed to capture target-specific representations

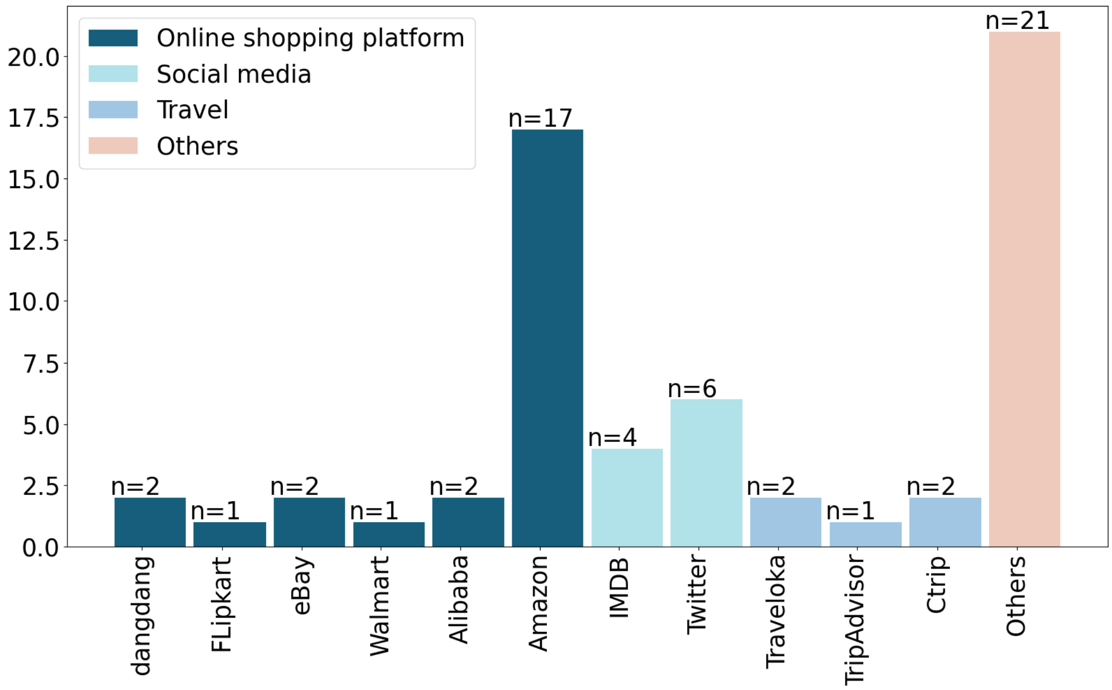

FIGURE 3. E-commerce platforms.

for aspect-level drug reviews. To further enhance the performance, they employ multi-task learning to transfer valuable domain knowledge from the short text-level drug review corpus. Meng et al. [9] presented a neural network model for aspect-level SA known as Feature Enhanced Attention CNN-BiLSTM. To enhance the context encoding quality and preserve semantic information, they utilize BiLSTM to capture not only the local features of phrases but also the global and temporal sentence semantics. Additionally, they incorporate an attention mechanism to model the interaction relationships between aspect words and sentences. This attention mechanism allows them to focus on the keywords related to the targets, enabling them to learn more effective context representations Feng et al. [77] included a customized masked attention mechanism designed for ABSA. They introduced two different approaches for generating the mask. The first technique involves setting an attention weight threshold based on the maximum weight value and retain only the attention scores above that threshold. The second approach was to select the top words with the highest weights. Both techniques aim to eliminate lower score components that are deemed less relevant to the targeted aspect, thereby focusing on the most important aspects.

Despite recent advancements, the current state-of-the-art models for aspect-level SA are limited in number. Aspect-based SA should be applicable across various types of datasets, ensuring its universality. Nguyen et al. [26] states that they plan to expand the proposed technique for aspect-based SA, and they plan to conduct more comprehensive experiments on various datasets. In addition to employing neural networks and attention mechanisms to address aspect-based SA, another novel approach is the utilization of generative models. Hosseini-Asl et al. [68] focuses on few-shot settings and aim to transform the extraction and prediction tasks into a sequence generation task. To achieve this, they employ a generative language model with unidirectional attention, primarily using GPT2 unless specified otherwise. By adopting this approach, the model learns to perform the tasks through language generation, eliminating the requirement for training task-specific layers. Up to this point, the performance of GPT-4 in this research has been exceptionally impressive, indicating a promising research direction, especially for developing large language models.

3) IMPLICIT ASPECT RECOGNITION AND EXTRACTION IN ASPECT-BASED SENTIMENT ANALYSIS

Over the past years, aspect extraction has emerged as a critical stage in SA for performing condensed sentiment classification. Nonetheless, prior research on SA has primarily concentrated on extracting explicit aspects, with relatively little attention given to implicit aspects.

Implicit aspects can be inferred from the sentence's context and the world's knowledge. For instance, if someone says, "This phone is running fast," we understand that they are referring to the phone's performance even though "performance" is not explicitly mentioned. Similarly, when someone says, "I cannot see anything in this room," we know they are referring to the lack of light in the room, even though the word "lightness" is not explicitly mentioned.

Maitama et al. [69] have identified techniques for extracting implicit aspects, explicit aspects, or both. Wang et al. [70] have developed a model called "Hierarchical Knowledge

Enhancement and Multi-pooling'' (HKEM) that effectively integrates knowledge information from various levels within the text. This integration is achieved through hierarchical knowledge enhancement, addressing the issue of ''weak features'' to enhance the overall performance. While the study takes into account the overlapping of features in two distinct segments using the domain's feature hierarchy, it has limited effectiveness in aspect classification. Wei et al. [71] proposed an implicit SA model named BiLSTM with multi-polarity orthogonal attention. By utilizing multi-polarity attention instead of the traditional single attention model, they can distinguish the variance between words and sentiment orientation. This disparity serves as a crucial feature in implicit SA. To improve the performance of their model, they incorporates external knowledge bases, such as a common-sense knowledge base, into the model. Becasue external knowledge is commonly implicit, it is not explicitly mentioned in sentences or directly indicated by the context. Therefore, it is still essential for them to determine an appropriate technique to represent this external knowledge.

Co-occurrence, semantic-based approaches, ontology, CRF, SVM, LSTM, Hierarchy, CNN, lexicon-based techniques, matrix factorization, and topic modeling are commonly used in solving implicit aspect extraction problems. Implicit aspect extraction is recognized as a relatively new and inherently ambiguous area, relying more on semantic understanding rather than explicit indications. Consequently, a considerable number of studies suggest that implicit aspect extraction holds promising potential as a future direction.

4) SARCASM DETECTION

Using positive words to express negative sentiments is a common feature of sarcastic text, which poses a challenge for SA models that are not explicitly designed to detect it. In real-life situations, individuals may express their opinions sarcastically, which can be challenging to discern. Therefore, in future work, it is necessary to enhance the technique to classify sentiments in such cases [29].

Sarcasm is frequently found in user-generated content, such as Facebook comments and tweets, and detecting it accurately in SA requires a thorough understanding of the context, topic, and environment.

Sarcasm can be a challenge for SA because it involves using language that is intended to convey the opposite meaning of what is being said. This can be difficult for machines to detect because the words used may appear positive or negative in isolation, but their intended meaning is the opposite.

For example, ''Great, now I have to walk home in the rain'' may appear positive because ''great'' is usually associated with positive sentiment. However, the intended meaning is negative because the person is expressing frustration about walking home in the rain. To address the sarcasm problem in SA, researchers have explored various techniques such as using contextual information, analyzing the speaker's tone of voice or facial expressions, and incorporating knowledge graphs or other sources of external information. Sait et al. [72] developed a sarcasm classification technique called Deep Learning with Natural Language Processing (DLNLP-SA). The DLNLP-SA technique utilizes deep learning and natural language processing to detect and classify instances of sarcasm in the input data.

However, despite these efforts, sarcasm detection remains a challenging problem in SA, and further research is needed to improve the accuracy of SA in identifying and understanding sarcastic language.

A different form of sarcasm known as ''numerical sarcasm'' was discussed by Kumar et al. [51], which is commonly observed on social media platforms. This type of sarcasm is based on alterations in numerical values that ultimately impact the polarity of the text. An example of numerical sarcasm would be:

''They drive so slowly—only 20 km/h.'' (Non-sarcastic)

''They drive so slowly—only 160 km/h.'' (Sarcastic)

It can be challenging to identify sarcasm when it is concealed within numerical data or the context of a statement during SA. Detecting covert sarcasm in text requires contextual understanding and can be addressed using algorithms such as RNN, LSTM, GRU, Transformers, or Bert models. These models are either descended from RNN algorithms or rely on attention mechanisms. The efficiency of sarcasm detection has greatly improved due to the contextual understanding capabilities of RNN algorithms and the attention mechanism. The future of sarcasm detection models is expected to be dominated by large language models.

5) FINE-GRAINED SENTIMENT ANALYSIS

In traditional SA classes, researchers tend to classify sentiments into only two categories: Positive or negative, alternatively, the polarity can be classified exclusively as positive, negative, or neutral [49]. However, sentiment should not be limited to two or three groups. In the future, there will be a greater emphasis on performing More fine-grained SA.

In Chandio's [49] they plan to create a benchmark dataset in Roman Urdu to facilitate fine-grained SA and question classification. This fine-grained SA will encompass five distinct categories, namely strongly negative, weakly negative, neutral, weakly positive, and strongly positive. In Yang's study, the approach presented is limited to classifying sentiment into only positive and negative categories, which may not be adequate in domains that demand more fine-grain SA. Hence, the subsequent step involves exploring the classification of sentiment granularity in text [12].

V. CONCLUSION

This paper concluded that recent papers applied SA in e-commerce. It categorized the models used based on the approaches and the platform from which the source data was obtained. This paper also discussed the limitations and future works as suggested by the researchers. Despite the challenges faced in applying SA in e-commerce, its importance cannot be overlooked, and further research in this area is warranted.

SA applied in e-commerce relies on machine learning and deep learning techniques, each with their own advantages and disadvantages. The primary data sources for SA, such as Amazon, Twitter, and IMDB, remain crucial. In the future, researchers can explore applying SA to other e-commerce platforms, such as stock exchange platforms. Future studies could focus on developing more universal models for new domains and languages, aspect-level SA models, implicit aspect recognition and extraction, sarcasm detection, and fine-grained SA to increase the usage of SA in e-commerce. It is hoped that more attention will be given to these areas.

REFERENCES

[1] L. Huang, Z. Dou, Y. Hu, and R. Huang, "Textual analysis for online reviews: A polymerization topic sentiment model," *IEEE Access*, vol. 7, pp. 91940–91945, 2019.

[2] H. M. Ahmed, M. J. Awan, N. S. Khan, A. Yasin, and H. M. F. Shehzad, "Sentiment analysis of online food reviews using big data analytics," *Elementary Educ. Online*, vol. 20, no. 2, pp. 827–836, 2021.

[3] A. Mabrouk, R. P. D. Redondo, and M. Kayed, "SEOpinion: Summarization and exploration of opinion from E-commerce websites," *Sensors*, vol. 21, no. 2, p. 636, Jan. 2021.

[4] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," in *Proc. Cogn. Inform. Soft Comput. (CISC)*. Singapore: Springer, 2019, pp. 639–647.

[5] W. Hong, C. Zheng, L. Wu, and X. Pu, "Analyzing the relationship between consumer satisfaction and fresh E-commerce logistics service using text mining techniques," *Sustainability*, vol. 11, no. 13, p. 3570, Jun. 2019.

[6] Y. Basani, H. V. Sibuea, S. I. P. Sianipar, and J. P. Samosir, "Application of sentiment analysis on product review E-commerce," *J. Phys., Conf. Ser.*, vol. 1175, no. 1, Mar. 2019, Art. no. 012103.

[7] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment analysis using word2vec and long short-term memory (LSTM) for Indonesian hotel reviews," *Proc. Comput. Sci.*, vol. 179, pp. 728–735, Jan. 2021.

[8] E. I. Elmurngi and A. Gherbi, "Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques," *J. Comput. Sci.*, vol. 14, no. 5, pp. 714–726, May 2018.

[9] W. Meng, Y. Wei, P. Liu, Z. Zhu, and H. Yin, "Aspect based sentiment analysis with feature enhanced attention CNN-BiLSTM," *IEEE Access*, vol. 7, pp. 167240–167249, 2019.

[10] S. Zhang, Z. Wei, Y. Wang, and T. Liao, "Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary," *Future Gener. Comput. Syst.*, vol. 81, pp. 395–403, Apr. 2018.

[11] A. F. Agarap, "Statistical analysis on E-commerce reviews, with sentiment classification using bidirectional recurrent neural network (RNN)," 2018, *arXiv:1805.03687*.

[12] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020.

[13] N. Shrestha and F. Nasoz, "Deep learning sentiment analysis of Amazon.com reviews and ratings," 2019, *arXiv:12704.04096*.

[14] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," in *Proc. IEEE Int. Conf. Innov. Res. Develop. (ICIRD)*, May 2018, pp. 1–6.

[15] P. K. Sari, A. Alamsyah, and S. Wibowo, "Measuring E-commerce service quality from online customer review using sentiment analysis," *J. Phys., Conf. Ser.*, vol. 971, Mar. 2018, Art. no. 012053.

[16] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A comparative study of support vector machine and naive Bayes classifier for sentiment analysis on Amazon product reviews," in *Proc. Int. Conf. Contemp. Comput. Appl. (ICA)*, Feb. 2020, pp. 217–220.

[17] S. Zhang, D. Zhang, H. Zhong, and G. Wang, "A multiclassification model of sentiment for E-commerce reviews," *IEEE Access*, vol. 8, pp. 189513–189526, 2020.

[18] S. Wassan, X. Chen, T. Shen, M. Waqar, and N. Z. Jhanjhi, "Amazon product sentiment analysis using machine learning techniques," *Revista Argentina de Clínica Psicolgica*, vol. 30, no. 1, p. 695, 2021.

[19] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, "A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks," *Neurocomputing*, vol. 380, pp. 1–10, Mar. 2020.

[20] Y. Han, M. Liu, and W. Jing, "Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer," *IEEE Access*, vol. 8, pp. 21314–21325, 2020.

[21] H. Xu, B. Liu, L. Shu, and P. S. Yu, "BERT post-training for review reading comprehension and aspect-based sentiment analysis," 2019, *arXiv:1904.02232*.

[22] H. T. Ismet, T. Mustaqim, and D. Purwitasari, "Aspect based sentiment analysis of product review using memory network," *Sci. J. Informat.*, vol. 9, no. 1, pp. 73–83, May 2022.

[23] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-sentiment: Sentiment analysis using ensemble of CNN and bi-LSTM models," 2019, *arXiv:1904.04206*.

[24] S. Garg, "Drug recommendation system based on sentiment analysis of drug reviews using machine learning," in *Proc. 11th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2021, pp. 175–181.

[25] P. Mukherjee, Y. Badr, S. Doppalapudi, S. M. Srinivasan, R. S. Sangwan, and R. Sharma, "Effect of negation in sentences on sentiment analysis and polarity detection," *Proc. Comput. Sci.*, vol. 185, pp. 370–379, Jan. 2021.

[26] Q. T. Nguyen, T. L. Nguyen, N. H. Luong, and Q. H. Ngo, "Fine-tuning BERT for sentiment analysis of Vietnamese reviews," in *Proc. 7th NAFOSTED Conf. Inf. Comput. Sci. (NICS)*, Nov. 2020, pp. 302–307.

[27] N. Majumder, S. Poria, A. Gelbukh, M. S. Akhtar, E. Cambria, and A. Ekbal, "IARM: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 3402–3411.

[28] A. Dubey, A. Gupta, N. Raturi, and P. Saxena, "Item-based collaborative filtering using sentiment analysis of user reviews," in *Proc. 1st Int. Conf. Appl. Comput. Commun. Technol. (ICACCT)*, Delhi, India. Singapore: Springer, Mar. 2018, pp. 77–87.

[29] H. Nguyen, T. Huynh, S. Hoang, V. Pham, and I. Zelinka, "Language-oriented sentiment analysis based on the grammar structure and improved self-attention network," in *Proc. 15th Int. Conf. Eval. Novel Approaches Softw. Eng. (ENASE)*, 2020, pp. 339–346.

[30] R. Catelli, S. Pelosi, and M. Esposito, "Lexicon-based vs. BERT-based sentiment analysis: A comparative study in Italian," *Electronics*, vol. 11, no. 3, p. 374, Jan. 2022.

[31] N. Nandal, R. Tanwar, and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products," *Spatial Inf. Res.*, vol. 28, no. 5, pp. 601–607, Oct. 2020.

[32] S. Zhang and H. Zhong, "Mining users trust from E-commerce reviews based on sentiment similarity analysis," *IEEE Access*, vol. 7, pp. 13523–13535, 2019.

[33] P.-F. Pai and C.-H. Liu, "Predicting vehicle sales by sentiment analysis of Twitter data and stock market values," *IEEE Access*, vol. 6, pp. 57655–57662, 2018.

[34] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment analysis about E-commerce from tweets using decision tree, K-nearest neighbor, and Naïve Bayes," in *Proc. Int. Conf. Orange Technol. (ICOT)*, Oct. 2018, pp. 1–6.

[35] C. Zuheros, E. Martínez-Cámara, E. Herrera-Viedma, and F. Herrera, "Sentiment analysis based multi-person multi-criteria decision making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews," *Inf. Fusion*, vol. 68, pp. 22–36, Apr. 2021.

[36] Y. Liu, J. Lu, J. Yang, and F. Mao, "Sentiment analysis for e-commerce product reviews by deep learning model of BERT-BiGRU-softmax," *Math. Biosci. Eng.*, vol. 17, no. 6, pp. 7819–7837, 2020.

[37] S. Riaz, M. Fatima, M. Kamran, and M. W. Nisar, "Opinion mining on large scale data using sentiment analysis and k-means clustering," *Cluster Comput.*, vol. 22, no. S3, pp. 7149–7164, May 2019.

[38] F. Noor, M. Bakhtyar, and J. Baber, "Sentiment analysis in E-commerce using SVM on Roman Urdu text," in *Proc. 2nd Int. Conf. Emerg. Technol. Comput. (iCETiC)*, London, U.K. Cham, Switzerland: Springer, Aug. 2019, pp. 213–222.

[39] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, "Sentiment analysis in the sales review of Indonesian marketplace by utilizing support vector machine," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 1, pp. 57–64, 2018.

[40] R. Watrianthos, S. Suryadi, D. Irmayani, M. Nasution, and E. F. Simanjorang, ''Sentiment analysis of traveloka app using nave Bayes classifier method,'' *Int. J. Sci. Technol. Res.*, vol. 8, no. 7, pp. 786–788, 2019.

[41] M. Umer, I. Ashraf, A. Mehmood, S. Kumari, S. Ullah, and G. Sang Choi, ''Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model,'' *Comput. Intell.*, vol. 37, no. 1, pp. 409–434, Feb. 2021.

[42] E. Prabhakar, M. Santhosh, A. H. Krishnan, T. Kumar, and R. Sudhakar, ''Sentiment analysis of U.S. airline Twitter data using new Adaboost approach,'' *Int. J. Eng. Res. Technol.*, vol. 7, no. 1, pp. 1–6, 2019.

[43] M. H. Munna, M. R. I. Rifat, and A. S. M. Badrudduza, ''Sentiment analysis and product review classification in E-commerce platform,'' in *Proc. 23rd Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2020, pp. 1–6.

[44] I. Awajan, M. Mohamad, and A. Al-Quran, ''Sentiment analysis technique and neutrosophic set theory for mining and ranking big data from online reviews,'' *IEEE Access*, vol. 9, pp. 47338–47353, 2021.

[45] M. Wongkar and A. Angdresey, ''Sentiment analysis using naive Bayes algorithm of the data crawler: Twitter,'' in *Proc. 4th Int. Conf. Informat. Comput. (ICIC)*, Oct. 2019, pp. 1–5.

[46] B. Chandio, A. Shaikh, M. Bakhtyar, M. Alrizq, J. Baber, A. Sulaiman, A. Rajab, and W. Noor, ''Sentiment analysis of Roman Urdu on E-commerce reviews using machine learning,'' *Comput. Model. Eng. Sci.*, vol. 131, no. 3, pp. 1263–1287, 2022.

[47] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, ''Sentiment analysis of comment texts based on BiLSTM,'' *IEEE Access*, vol. 7, pp. 51522–51532, 2019.

[48] A. A. Aagte, W. Vrushali, P. Vishwakarma, and S. Kamble, ''Spam detection using sentiment analysis,'' in *Proc. 10th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, 2019, pp. 209–214.

[49] B. A. Chandio, A. S. Imran, M. Bakhtyar, S. M. Daudpota, and J. Baber, ''Attention-based RU-BiLSTM sentiment analysis model for Roman Urdu,'' *Appl. Sci.*, vol. 12, no. 7, p. 3641, Apr. 2022.

[50] F. A. Ramadhan, R. R. P. Ruslan, and A. Zahra, ''Sentiment analysis of E-commerce product reviews for content interaction using machine learning,'' *Cakrawala Repositori IMWI*, vol. 6, no. 1, pp. 207–220, Feb. 2023.

[51] L. Kumar, A. Somani, and P. Bhattacharyya, '''Having 2 hours to write a paper is fun!': Detecting sarcasm in numerical portions of text,'' 2017, *arXiv:1709.01950*.

[52] V. Ganganwar and R. Rajalakshmi, ''Implicit aspect extraction for sentiment analysis: A survey of recent approaches,'' *Proc. Comput. Sci.*, vol. 165, pp. 485–491, Jan. 2019.

[53] P. Mehta and S. Pandya, ''A review on sentiment analysis methodologies, practices and applications,'' *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.

[54] R. Obiedat, D. Al-Darras, E. Alzaghoul, and O. Harfoushi, ''Arabic aspect-based sentiment analysis: A systematic literature review,'' *IEEE Access*, vol. 9, pp. 152628–152645, 2021.

[55] S. Sagnika, A. Pattanaik, B. S. P. Mishra, and S. K. Meher, ''A review on multi-lingual sentiment analysis by machine learning methods,'' *J. Eng. Sci. Technol. Rev.*, vol. 13, no. 2, pp. 154–166, Apr. 2020.

[56] Z. Drus and H. Khalid, ''Sentiment analysis in social media and its application: Systematic literature review,'' *Proc. Comput. Sci.*, vol. 161, pp. 707–714, Jan. 2019.

[57] J. A. Shathik and K. K. Prasad, ''A literature review on application of sentiment analysis using machine learning techniques,'' *Int. J. Appl. Eng. Manag. Lett.*, vol. 4, no. 2, pp. 41–67, 2020.

[58] A. Ahmet and T. Abdullah, ''Recent trends and advances in deep learning-based sentiment analysis,'' in *Deep Learning-Based Approaches for Sentiment Analysis* (Algorithms for Intelligent Systems), B. Agarwal, R. Nayak, N. Mittal, and S. Patnaik, Eds. Singapore: Springer, 2020, doi: 10.1007/978-981-15-1216-2_2.

[59] S. Shayaa, N. I. Jaafar, S. Bahri, A. Sulaiman, P. S. Wai, Y. W. Chung, A. Z. Piprani, and M. A. Al-Garadi, ''Sentiment analysis of big data: Methods, applications, and open challenges,'' *IEEE Access*, vol. 6, pp. 37807–37827, 2018.

[60] M. Umar, A. A. A. Bena, and B. Wadata, ''Sentiment analysis techniques and application-survey and taxonomy,'' *J. Eng. Technol.*, vol. 4, no. 1, pp. 781–789, 2021.

[61] R. Baragash and H. Aldowah, ''Sentiment analysis in higher education: A systematic mapping review,'' *J. Phys., Conf. Ser.*, vol. 1860, no. 1, Mar. 2021, Art. no. 012002.

[62] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, ''A survey of sentiment analysis in social media,'' *Knowl. Inf. Syst.*, vol. 60, pp. 617–663, Jul. 2019.

[63] M. Hilario, D. Esenarro, I. Petrlik, and C. Rodriguez, ''Systematic literature review of sentiment analysis techniques,'' *J. Contemp. Issues Bus. Government*, vol. 27, no. 1, pp. 506–517, 2021.

[64] S. Elzeheiry, W. A. Gab-Allah, N. Mekky, and M. Elmogy, ''Sentiment analysis for e-commerce product reviews: Current trends and future directions,'' *Preprints*, May 2023, Art. no. 2023051649. [Online]. Available: https://www.preprints.org/manuscript/202305.1649/v1, doi: 10.20944/preprints202305.1649.v1.

[65] A. Qazi, R. G. Raj, G. Hardaker, and C. Standing, ''A systematic literature review on opinion types and sentiment analysis techniques,'' *Internet Res.*, vol. 27, no. 3, pp. 608–630, 2017.

[66] M. Marong, N. K. Batcha, and R. Mafas, ''Sentiment analysis in E-commerce: A review on the techniques and algorithms,'' *J. Appl. Technol. Innov.*, vol. 4, no. 1, p. 6, 2020.

[67] H. Saleh, S. Mostafa, A. Alharbi, S. El-Sappagh, and T. Alkhalifah, ''Heterogeneous ensemble deep learning model for enhanced Arabic sentiment analysis,'' *Sensors*, vol. 22, no. 10, p. 3707, May 2022.

[68] E. Hosseini-Asl, W. Liu, and C. Xiong, ''A generative language model for few-shot aspect-based sentiment analysis,'' 2022, *arXiv:2204.05356*.

[69] 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.

[70] H. Wang, M. Hou, F. Li, and Y. Zhang, ''Chinese implicit sentiment analysis based on hierarchical knowledge enhancement and multi-pooling,'' *IEEE Access*, vol. 8, pp. 126051–126065, 2020.

[71] J. Wei, J. Liao, Z. Yang, S. Wang, and Q. Zhao, ''BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis,'' *Neurocomputing*, vol. 383, pp. 165–173, Mar. 2020.

[72] A. R. W. Sait and M. K. Ishak, ''Deep learning with natural language processing enabled sentimental analysis on sarcasm classification,'' *Comput. Syst. Sci. Eng.*, vol. 44, no. 3, pp. 2553–2567, 2023.

[73] K. K. Hantoro, D. Handayani, and S. Setiawati, ''A implementation of text mining in sentiment analysis of Shopee Indonesia using SVM,'' *Bull. Inf. Technol.*, vol. 3, no. 2, pp. 115–120, 2022.

[74] H. Murfi, Syamsyuriani, T. Gowandi, G. Ardaneswari, and S. Nurrohmah, ''BERT-based combination of convolutional and recurrent neural network for Indonesian sentiment analysis,'' 2022, *arXiv:2211.05273*.

[75] M. R. Pratama, F. A. G. Soerawinata, R. R. Zhafari, Rendy, and H. N. Imanda, ''Sentiment analysis of beauty product E-commerce using support vector machine method,'' *Jurnal RESTI, Rekayasa Sistem dan Teknologi Informasi*, vol. 6, no. 2, pp. 269–274, Apr. 2022.

[76] W. Huang, M. Lin, and Y. Wang, ''Sentiment analysis of Chinese E-commerce product reviews using ERNIE word embedding and attention mechanism,'' *Appl. Sci.*, vol. 12, no. 14, p. 7182, Jul. 2022.

[77] A. Feng, X. Zhang, and X. Song, ''Unrestricted attention may not be all you need–masked attention mechanism focuses better on relevant parts in aspect-based sentiment analysis,'' *IEEE Access*, vol. 10, pp. 8518–8528, 2022.

HUANG HUANG received the B.S. degree from Wuhan Textile University and the M.S. degree in data science from the University of Malaya, Kuala Lumpur, in 2021.

In college, his GPA is 3.24. From 2019 to 2021, he was majoring in data science with the University of Malaya. Since 2021, he has been a Software Engineer with HSBC. He has been contributing to the advancement of state-of-the-art models within the field of sentiment analysis, since 2021. His research interests include natural language processing, with a particular emphasis on sentiment analysis. He won the ''Excellent Graduate'' Prize. He has participated in ''Internet+'' competition for two times. His project ''Online Monitoring System Based on Image Recognition and Multi-Sensors'' won the Bronze Award from the third ''Internet+'' College Student Innovation and Entrepreneurship Competition, from to October 2016 to July 2017. Another project named ''The system of Raspberry pi Sensor Real-Time Motion Detection based on Cloud Computing'' is a ''National-Level Project,'' from December 2017 to September 2018, on college student innovation and entrepreneurship platform.

ADELEH ASEMI ZAVAREH (Member, IEEE) received the B.Sc. degree from Ashrafi Esfahani University, Iran, in 2006, the M.Sc. degree from Pune University, India, in 2008, and the Ph.D. degree from the Faculty of Computer Science and Information Technology, University of Malaya, under the Prestigious MTCP Scholarship, in 2014.

She was conducting research in HCI and Human Decision Making Simulation area, as a Postdoctoral Fellow with the Department of Software Engineering, University of Malaya, in 2015. She has been a Lecturer and an Assistant Professor with Esfahan University and the Safahan Institute of Higher Education, Esfahan, Iran, respectively, since 2011. She received training in modeling and simulation and attended workshops on big data and R-programming with the University of Malaya. She supervised more than 30 postgraduate students and taught both undergraduate and postgraduate computer science subjects. She joined the Faculty of Computer Science and Information Technology, University of Malaya, as a Visiting Research Fellow, in 2019, and has been an Affiliated Senior Lecturer, since 2020. She is currently a Senior Lecturer with the Department of Software Engineering, University of Malaya. Her research interests include the intersection of decision analysis including simulation/prediction/evaluation and artificial intelligence. She has worked to improve the decision analysis by using theories from disparate areas and formulate the evaluations of different alternatives, such as engineering methods and systems to multi-criteria decision analysis. The breadth of her research agenda spans theoretical analysis and systems development, and her research style is not only to develop decision-making techniques and analyze their formal guarantees but also to implement them for industries in the physical world. In addition, she has forged interdisciplinary alliances with experts in agriculture, software engineering, law, environmental science, and speech recognition. Since 2019, she has been extended her area of research to the multi-criteria evaluation of big data and the IoT techniques.

MUMTAZ BEGUM MUSTAFA (Member, IEEE) received the B.Sc. degree in software engineering from Universiti Putra Malaysia (UPM), in 2002, and the M.Sc. degree in software engineering and the Ph.D. degree in computer science from the University of Malaya (UM), in 2006 and 2012, respectively. She is currently an Associate Professor with UM. She has undertaken several Speech Synthesis research and holds grants from the Ministry of Higher Education. She has established network with a number of the International Speech Research Laboratory in Japan and Singapore. She supervises a group of Ph.D. and master's students, working on speech synthesis, speech recognition, and speech signal processing. She has published several papers in prestigious speech conferences and journals. She has published her work in many of prestigious international journals. Her research interests include emotional speech synthesis and speech assistive tools for disabled individuals. Her research and development of the HMM-based Malay speech synthesis system has won the Most Prestigious Award (MPA) for Excellent Research 2012 from the MIMOS Berhad the National Research and Development Centre in ICT and has won gold medals for several national level competitions.