0009-0000-8189-8446 ©2023 IEEE Sentiment Analysis: Methods and Application Using Machine Learning in Different Fields

Article · December 2023

2 authors, including:

Archana Merline Lawrence
Bournemouth University

Sentiment Analysis: Methods and Application Using Machine Learning in Different Fields

Archana Merline Lawrence
Department of Computing and Informatics
Bournemouth University
Bournemouth, UK
https://orcid.org/0009-0000-8189-8446

Anchal Pradhan Adhikari
Department of Computing and Informatics
Bournemouth University
Bournemouth, UK
https://orcid.org/0009-0003-9051-7221

***Abstract*— This research discusses the application of Machine Learning (ML) models for Sentiment analysis (SA) across diverse industries and professional domains, such as social media, consumer reviews, healthcare, and banking, in response to the exuberant volume of textual data on the internet. Exploring various ML models used by researchers which include preprocessing techniques and data collection methods, this study unveils the detailed methodologies employed in SA. The integration of ML emerges as a catalyst for positive outcomes in domains including social media, natural disaster crisis monitoring, healthcare, education, tourism, hospitality, finance, and e-commerce. Positive outcomes in a variety of fields, such as social media, healthcare, education, tourism, hospitality, finance, and e-commerce, are being spurred by the incorporation of ML. These results not only demonstrate the versatility of sentiment analysis but also point to new uses that have not yet been fully explored, offering researchers a possible path forward for additional investigation and the creation of customized techniques. These difficulties, which include linguistic heterogeneity, moral dilemmas, and domain-specific obstacles, call for additional study to improve algorithms and handle these complexities. By doing this, it will be possible to promote sentiment analysis in a variety of professional domains and make it more sophisticated and domain-centric.**

Keywords*—*Sentiment Analysis (SA), Machine Learning (ML), Supervised Learning (SL), Unsupervised Learning (UL), Natural Language Processing (NLP), Healthcare, Social Media, Finance, E-commerce.

I. Introduction

Before the Internet, people relied on word-of-mouth and businesses analyzed sales data to understand customer behaviour[1]. Social media changed this by allowing users to express their opinions on social media be it personal or business posts[2], leading to the development of advanced algorithms for SA that transformed how businesses gain insights[3]. Industries like finance, healthcare, politics, and customer service began using internet data for sentiment analysis to better understand customer behaviour[4]. With a majority of the global population using the internet, a large amount of textual data is generated for various business purposes[5]. SA goes beyond simple classification, and it aims to understand the emotions conveyed in the text data. SA is crucial for unravelling the intended meaning in different types of textual data[6]. It involves identifying and extracting significance from sentiments expressed in text documents, offering a deeper understanding of the conveyed emotions, whether they are positive, negative, neutral, subjective, or objective[3]. For SA research there are different levels:

Document Level: Focuses on understanding the general view of the document, it is categorised as good, bad and ok instead of providing the insights or actual sense or purpose of the document [7], [8].

Sentence Level: Sentences are treated as separate parts and given labels denoting positive, negative, or neutral sentiment. Sentence-based analysis is a useful technique to detect changes in a document's opinion[9]. It provides a more complex interpretation of the message by highlighting variations and fluctuating emotions throughout the text[10].

Aspect/Feature Level: This finds and assesses the sentiment connected to certain topics or sections of a document rather than examining the general emotions, it divides the document according to various features discussed in the text[10], [11]. For example: quality, colour, capacity, and waterproofing are some of the aspects of a bag.

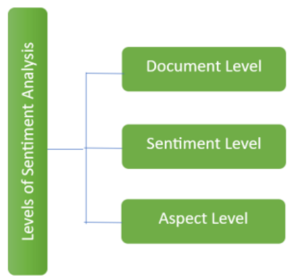

Fig.1 Levels of SA [12], [13, 14]

SA has advanced in recent years for its capability to use NLP techniques driven by ML models applied to various domains[15]. Studies have shown promising results in the fields of marketing, customer feedback, finance, etc. [16]. All of them use different methods of ML models to achieve a set goal curated only to a specific field and even if the same methods are used in different areas the results vary in accuracy depending on the dataset[5]. Examining SA applications in different areas not only aids researchers in understanding the diverse approaches but also emphasizes the potential for applying successful models from one domain to benefit others by sharing knowledge and the transfer of successful practices from one field to another for broader applicability[17]. Acknowledging the replicability of successful models across diverse professional fields enhances collaboration and adaptability. Therefore, we aim to build a list of SA approaches used in various industries and understand the applicability purpose in the respective domains. The objectives of the paper are: (i) List different methods researchers used to apply SA in their respective fields. (ii) To summarize the purpose and challenges of each field applying SA for insights.

II. Related Work

A. Evolution of Sentiment Analysis

SA has emerged as one of the most active fields of NLP research since the year 2000[18]. In this period of rapid digital transformation, a specific feature that establishes the current information environment is the increasing popularity of textual data on Internet platforms. Websites, user reviews, social networking platforms, and other internet outlets generate a high volume of random information[19]. The increasing amount of textual information demonstrates the dynamic nature of human opinion, which consists of a vast array of topics, emotions, and linguistic expressions. With this data, high-tech companies have created ChatGPT which is a chat-based platform, among applying many ML models SA is one of the methods. This platform has gained much attention for its versatile knowledge in various subjects and English language skills[20]. SA sometimes referred to as opinion mining, is a crucial part of NLP that offers a systematic and automated way to extract attitudes, views, and emotions from text[21].

B. Machine Learning Techniques in Sentiment Analysis

Pre-processing techniques for text data, including clustering and SVM classification, were examined in a study of SA product reviews. The authors used a range of datasets to train and test the SVM method, and they found that it provided superior accuracy (89.98%)[22].

Another study by Yh Ko *el at.* to create a model that accurately categorizes the customer experience levels by analysing tone in voices. By extracting features from data and constructing a model, the proposed SVM model with Auto Encoder achieved a high accuracy of 79.3%, outperforming the LSTM-RNN model[23]. With reconstructed feature vectors using SVM models, the paper highlights a drastic improvement in average accuracy, from 53.79% to 73.97%. These accuracy metrics demonstrate the effectiveness of the developed model in identifying customer satisfaction levels based on the tone of their voices.

Further in research text mining is used for tweets and significant data management along with deep neural networks to increase the accuracy of SA[24]. The accuracy was 75.03% for the deep learning approach and 52.60% for the multilayer perceptron method.

Hammi's proposed CBRS architecture is to efficiently categorize review's feelings about aspect phrases as positive or negative. By integrating the SVM model with deep learning models (CNN and bi-RNN), it can extract contextual information from the input text and identify local features. On the smartphone dataset, the CBRS demonstrated remarkable accuracy and resilience in aspect-based sentiment analysis with a 94.05% F-measure[25]. Competitive precision, recall, and F-measure values were obtained from the evaluation of the SemEval-2016 restaurant dataset, confirming the architecture's ability to reliably categorize feelings towards features.

III. Sentiment Analysis Approaches

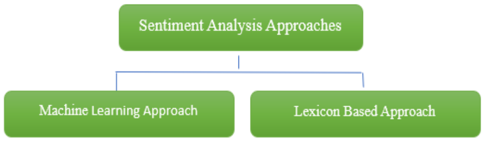

Fig.1 SA Approaches [26], [27]

Sentiment analysis is a technique or strategy that determines whether a text expresses either favourable or unfavourable emotions[17]. Consider using it to train a model to recognize if a written word is amusing, depressing, or something in between[28]. These methods might be as simple as focusing on particular words or as complex as employing modern techniques like machine learning or deep learning[29]. It essentially comes down to figuring out how to efficiently assist a machine in understanding and decoding human emotions from written words.

1) Machine Learning Approach

Artificial intelligence (AI) includes ML, to enhance decision-making and prediction, ML algorithms uncover essential patterns in data[30]. SA is more effective when using the ML approach because it is entirely computerised and can handle an high volume of data from the internet, compared to the other methods[31]. Algorithms used in machine learning techniques aim to identify patterns in data and link those patterns to specific groups of samples within the data. ML is categorised as Supervised learning (SL) and unsupervised learning (UL).

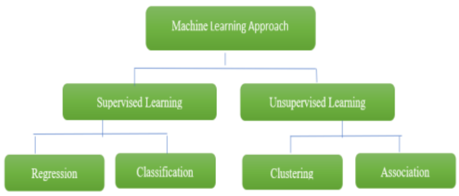

Fig.2 ML Approach for SA [32]

a) Supervised learning: The SL has two methods, classification, and regression. With two or more classes in classification, the output variable is categorical. Regression is a relationship between two or more variables in which a change in one variable is linked to a change in another[33].

b) Unsupervised Learning: The term "learning" in the context of UL refers to the machine attempting to recognize patterns and respond accordingly. UL is further divided into Clustering and Association[34]. The ML creates a group based on how the data behaves during clustering. Rule-based ML is used in association to find significant relationships between variables in big datasets.

2) lexicon-Based Approach

It is a text-based technique for emotion mining that develops assumptions about sentiments based on a word or group of words[35]. This strategy is frequently referred to as a keyword-based technique. A lexicon, which either gathers information about emotions or, at the very least, conveys words or phrases, can be used to identify the substance of a message as well as the feelings and emotions contained in these words[36], [37].

a) Dictionary-Based Approach: It is also known as a rule-based approach, is a sentiment analysis technique that uses a pre-established dictionary or lexicon of words that have been assigned sentiment scores[38]. Based on the matching sentiment term in the lexicon, each word in a phrase is given a sentiment score. The total, or average, of the sentiment scores of the words that make up a sentence, is then used to get the sentiment score of the phrase[39]. Because it enables the rapid and automated analysis of vast volumes of text data, this method is helpful for sentiment analysis.

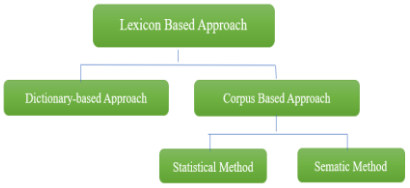

Fig.3 Lexicon Based Approach for SA [40], [41]

b) Corpus-Based Approach: It is a methodology that searches for patterns and regularities in language used by analyzing vast collections of texts, or corpora. This methodology examines language use without depending on emotions or prior conceptions about how language functions, but rather in a logical and data-driven manner[42]. It may discover frequently used words and phrases as well as more complex language patterns like collocations, formulaic sequences, and grammatical structures by examining corpora. This method has been widely applied in a variety of domains, including NLP, linguistics, and language instruction.

IV. Sentiment Analysis Methodology

To automatically identify the available pattern in the provided collection of data, ML algorithms can be approached as a combination of techniques.

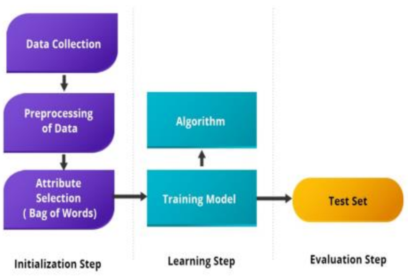

Fig.4 Steps Involved in Training a Classifier for Sentiment Analysis[16],

A. Data Collection

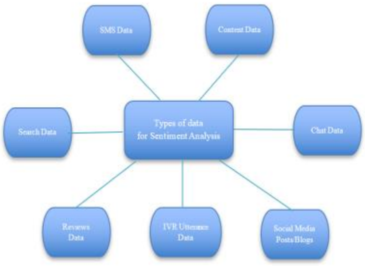

Fig.5 Types of data for SA

Online shopping websites, social media, surveys, customer reviews, and form datasets that are freely accessible on the internet are some of the methods used to get raw data. Data collection [43] is used for further data analysis and processing. With increasing popularity in deriving profitable insights critical data is difficult to extract unless we pay for API[44]. Collecting textual live data is an expensive affair. Valid and high-quality data collection is a crucial part of SA and ML performance.

B. Preprocessing of Data

In this process the raw data is transformed to workable or sequential or organized data for deriving useful information relevant to the problem at hand and achieving set expectations. In SA grammatical symbols are removed from the complete data set, and each paragraph is shortened into a single word[45]. Such that it is categorized into the following steps - tokenization, N-grams, stemming, stop words removal, part-of-speech tagging, etc.

The quality of data preprocessing has a major impact on the SA model's efficiency, as it ensures its capacity to reliably interpret and extend perspectives across a variety of textual inputs.

C. Feature Selection

SA feature selection is challenging because processed text data is converted into a format that machine learning algorithms can understand. Several techniques [46] are employed, including Term Frequency-Inverse Document Frequency (TF-IDF) and Bag-of-Words (BoW). BoW is a fundamental method for turning text into numbers. An array of the frequency of each dictionary word in the sentence represents a tokenized text[31]. Another method, known as TF-IDF, considers both word frequency and the relevance of a word based on how often it appears in documents. GloVe and Word2Vec [47] other methods that are gaining popularity in attribute selection.

D. Model Training

A labelled dataset containing both text and samples is necessary for model training; only the prediction model can extract sentiment from text. The model is trained using ML techniques like support vector machines or Naïve Bayes. Data are divided into training and validation sets to ensure model is not over or under trained. Based on the trained model performance we can tune the parameters accordingly to improve performance[48].

E. Test Set

After training the model in the previous step, we apply the trained model on raw unseen data. The percentage of accuracy will determine the success of the trained SA model[48]. The trained can be tuned further to improve performance if there are limitations found post testing.

V. APPLICATION OF SENTIMENT ANALYSIS IN DIFFERENT INDUSTRIES

A. Social Media

In social media, there are many platforms available that allow people to share their opinions and thoughts, which are used by researchers and business owners to do SA by retrieving data[49]. Various methods have been explored in the field of sentiment analysis on social media to understand the feelings and thoughts of users. One interesting study by MA Shafin et al. employed customer feedback and NLP to analyse online shopper reviews in Bangla language. The investigation comprised testing and training sets with different data splits (30%, 40%, 50%, 60%, and 70%). The results showed that SVM had better accuracy at 88.81%, whereas KNN performed lower at 80.14% in 30% of test data usage[50]. Classifying emotions within textual data was the goal of another study that focused on tweets from events such as the UNGA conference, the Houston Howdy Modi gathering, the release of the movie The Sky is Pink, and the Haryana Assembly Polls[51]. The LSTM system obtained 95.97% accuracy for the Howdy Modi gathering in Houston, while the Naïve Bayes method showed an astounding 98.4% accuracy for the Haryana Assembly Polls.

Additionally, a novel study approach that merged ontology-based feature-level sentiment analysis, with review fraud detection produced better results compared to existing methods in terms of accuracy, precision, recall, and F1-score estimates. Although this approach shows potential uses for consumers making decisions about what to buy and businesses reviewing consumer reviews[52]. These studies highlight the various uses and develop techniques of SA in the social media context.

B. Natural Disaster

By improving our awareness of the emotional component of emergencies, SA in the context of natural disasters facilitates more knowledgeable and generous decision-making for successful disaster response, recovery, and long-term resilience development[53]. It is clear from recent research projects that sentiment analysis in the context of natural disasters is changing in which the Kashmir floods dataset was used in on crisis-related social media, where a Naïve Bayes classifier trained on features including Part-of-Speech, tagging and, unigrams was evaluated and yielded an overall accuracy of 66.88%[54]. Both governmental and non-governmental organisations can benefit from the insightful information provided by this method to improve their ability to handle situations of crisis.

Another modern strategy uses SA and geographic information derived from artificial intelligence (AI) to automate disaster monitoring via social media posts. This technique used a deep learning model based on CNN, showed remarkable recall (88%), precision (93%), and F1-Score (90%), resulting in an overall accuracy of 97% in identifying disaster zones[55]. These research highlights how sophisticated sentiment analysis methods have grown to become and how they can provide effective means of monitoring and responding to emergency situations by using insights from social media data.

C. Healthcare

SA is essential in the healthcare industry to address the challenging work of extracting senses from data about beneficiaries, such as patients, and the complexity of biological tests[56]. Modern studies have shown that SA is becoming an increasingly useful tool in the field of healthcare. Edara el at. work is remarkable for its creative application of text mining and machine learning techniques. It focuses on tweets from online communities supporting cancer patients. By utilising a dispersed architecture, the research presented a LSTM neural network model and demonstrated its efficacy instead of traditional sentiment analysis techniques. The accuracy percentages of the LSTM model ranged from 58.16% to 84.10%, indicating its adaptability to various datasets and feature extraction techniques on single-node and multi-node machines[57]. A more comprehensive scoping assessment in the healthcare field places a particular focus on clinical narrative sentiment analysis. The evaluation covers a wide range of approaches, including machine learning-based domain-specific sentiment analysis techniques as well as general domain tools like TextBlob and sentiment lexicons like SentiWordNet[58]. However, these studies with a healthcare focus indicate accuracy scores ranging from 71.5% to 88.2%. These results emphasise the expanding importance of sentiment analysis in medical research, by providing insights into patient's emotional states, sentiment analysis may have a critical role in advancing patient care and healthcare communication.

D. Education

The Valence-Aware Dictionary for Sentiment Reasoning (VADER) served to examine qualitative information obtained from research concerning the use of technology by young children[59]. The M. R. Yaakub *et al.* shows an inconsistency between beliefs and actual practices, despite positive sentiments indicated in both qualitative and quantitative data. This demonstrates the potential of sentiment analysis when combined with learning analytics and knowledge research, as well as the difficulty of interpreting sentiment in educational environments. It also emphasises Support Vector Machine (SVM) as a sentiment analysis model that works well, especially when processing unstructured text data.

Further investigation presents a hybrid method for sentiment analysis of student feedback that combines lexicon-based approaches and machine learning. With an excellent accuracy of 93% and an F-measure of 92%, the proposed model exceeds prior approaches by utilising TF-IDF and domain-specific lexical characteristics[60]. It emphasises how important it is to create a thorough vocabulary specifically for the academic setting. In addition, it highlights how useful TF-IDF and unigram features are for sentiment analysis, offering insightful information about the mindsets of students.

E. *Tourism and Hospitality*

This study makes a major contribution to the field of tourism research by applying the SVM algorithm to multidimensional sentiment analysis at important Indonesian tourist destinations. With the goal of expanding the scope of sentiment analysis techniques in the context of tourism, the study covers a wide range of locations, including DKI Jakarta, Banten, East Java, Central Java, and West Java.

The research clarifies the relative efficacy of two different data labelling techniques, Textblob and Transformer for classifying visitor comments and reviews[61]. Although the study offers insightful information, it properly notes its limitations, especially its dependence on just two labelling methodologies. Other than that, an additional study conducted in the SA of Bandung Tourist Destination examines how well the SVM and NB algorithms perform when analyzing sentiments about tourist destinations. To rectify the data imbalance, the artificial minority oversampling technique (SMOTE) is used[62].

TABLE1 - METHODS USED IN SENTIMENT ANALYSIS

Studies	**Machine Learning Approach**								**Lexicon Based Approach**
	SVM	***KNN***	***LR***	***NB***	***CNN***	***DT***	***LSTM***	***RF***	
[50]	✓	✓	✓	-	-	✓	-	✓	-
[51]	✓	-	-	✓	-	-	✓	✓	-
[54]	-	-	-	✓	-	-	-	-	-
[57]	✓	-	✓	✓	✓	✓	✓	-	-
[58]	-	-	-	-	-	-	-	-	✓
[59]	✓	-	-	-	-	-	-	-	-
[60]	-	-	-	-	-	-	-	-	✓
[61]	✓	-	-	-	-	-	-	-	-
[62]	✓	-	-	✓	-	-	-	-	-
[63]	-	-	-	-	-	-	-	-	✓
[64]	-	-	-	-	-	-	✓	-	-
[65]	-	-	-	✓	-	-	-	-	-
[66]	✓	-	-	-	-	-	-	-	-
[67]	✓	-	-	-	-	-	-	-	-

With an astounding accuracy score of 88.51%, the results demonstrate the supremacy of SVM with SMOTE and provide important context for understanding how effectively machine learning algorithms work for SA in various tourism scenarios.

A hotel document polarity classification using a supervised machine learning technique for SA is applied, in which a unigram feature containing two distinct forms of information—frequency and TF-IDF. The results show that in this specific application, TF-IDF information works better than frequency[68]. Experimental results successfully classify hotel ratings, supporting the usefulness of the suggested strategy. The SVM approach is the most successful sentiment analysis technique, emphasising its superiority in identifying feelings in documents pertaining to hotels. By highlighting the effectiveness of using TF-IDF information and the superiority of SVM in the context of hotel reservations, this research offers insightful information about sentiment analysis techniques for the hospitality industry.

F. *Finance*

The banking sector can greatly improve its overall performance by putting more emphasis on providing high-quality services and customer satisfaction to attract new consumers and retain current ones[69]. Text mining techniques are utilised in the "Lexicon-based Sentiment Analysis of the Indian Union Budget 2016–17" to do sentiment analysis on the yearly budget. The authors prefer to provide further insight into the perspective conveyed in the official document by taking data out of the budget and analysing word associations. This method provides sophisticated knowledge of the sentiments connected with important economic decisions and is an example of how sentiment analysis may be used in financial policy texts[63]. A complete prediction model was developed specifically for the Colombo Stock Exchange (CSE) to improve stock trend predictions. This model determines the effect of economic news items on market moves by combining text categorization with an advanced LSTM-based RNN architecture. The model is unique for combining sentiment analysis and quantitative measures, demonstrating the combination of quantitative prediction methods with a particular emphasis on sentiment analysis[64]. This combination makes the model more adept at navigating the complexities of financial sentiment and, as a result, increases its predictive power for Colombo Stock Exchange stock patterns. SA in the financial sector, from policy evaluations to stock market forecasts, highlights its applicability and flexibility in gaining valuable data for informed decision-making.

G. E-Commerce

The organized and unorganized data that is easily accessible on the internet has helped the e-commerce business grow by providing sellers with significant information and enabling prospective consumers to make smart product selections[70]. Malki Kothalawala and Samantha Thelijjagoda present a hybrid methodology combining language resources and machine learning for aspect-based opinion mining on hair care product evaluations. Utilizing the Naïve Bayes technique, they dynamically display aspect-wise polarity of reviews on an eBay product page. Acknowledging the challenges of sentiment analysis in customer reviews, the [65] propose that additional training data from diverse websites could enhance the system further. Similarly, sentiment analysis in mobile phone reviews on platforms like Flipkart, employing the Naïve classification method, classifies data using features such as polarity, age, and helpfulness score, ultimately represented through star ratings[66]. This trained model aids users in making informed decisions when selecting products. Furthermore, Haque et al. emphasize the utility of machine learning methods in e-commerce sentiment analysis, employing feature extraction and pool-based active learning strategies. Their study showcases the prowess of Support Vector Machine (SVM) classifiers, particularly the linear variant, with impressive performance across datasets, including Electronics (94.02%), Music Instruments (94.02%, precision 0.9889, recall 0.971, F1 0.98), and Cell Phone & Accessories (93.57%, precision 0.96, recall 0.97, F1 0.97)[67].

H. Other Industries

In a study by Charlie *et al.* introduces an innovative approach for using SA in businesses to identify possible issues early on. The ASEP framework, which examines how employees express their emotions in emails, was developed to create profiles by examining the emotions and people they interact with to identify any strange or troubling behaviour[71]. They discovered it functioned successfully after testing it on emails from a business named Enron. As a result, ASEP is a powerful tool for identifying issues within organisations through SA of email correspondence.

Fahim and Musleh present an advanced automated system that uses AI techniques to analyse world events in multiple dimensions. The system receives unstructured descriptions of world events from a variety of sources, such as social media and news organisations, and analyses them using entity detection, SA, and anomaly detection. With a high overall classification accuracy of 99.4% for entity detection and an area under curve (AUC) score of 94.1% for anomaly detection, the methodology demonstrates extraordinary robustness and accuracy[72]. The system is notable for its explainable AI capabilities, which provide intelligent answers to questions and provide textual and visual explanations for unusual activity. The detailed evaluation shows the system's smooth integration and exceptional precision in identifying unexpected worldwide events.

In another study the researcher addresses the use of labelling and collecting few tweets regarding certain politicians to use SA on Twitter to forecast election outcomes. They used SVM and Multinomial NB algorithms, where SVM confirmed the best accuracy. As part of the approach, a ML system is trained to forecast election results and analyse popular opinion. The authors propose an active learning model for efficient labelling and recommend combining their model with other statistical methods and offline techniques for enhanced predictions[73].

Furthermore, Patel *et al.* provides a comprehensive review of the most recent AI techniques for interpreting facial expressions to determine emotions. The authors investigate multiple approaches for handling various emotions in facial photographs, including the use of CNN and soft labelling. They analysed earlier studies in the field and develop a classification scheme based on techniques for face detection, feature extraction, and emotion classification. They propose a successful model that achieves excellent accuracy on several datasets using multi-signal CNN (91.16%), Model Fusion, and Part-based Hierarchical bidirectional RNN. The results of the research have a big impact on applications in the industry, especially in fields like interaction between humans and machines, emotionally motivated robots, and healthcare[74].

VI. Purpose And Challenges

The different types of data, user behaviour, and context create challenges for SA across various industries. Table 2 lists the purpose and difficulties specific to various areas.

Conclusion

In conclusion, the implementation of ML for SA has yielded favorable outcomes across various professional domains.

In the sphere of social media, studies have demonstrated the effectiveness of SVM and LSTM techniques in capturing and comprehending user sentiments and emotions. The adaptability of SA is evident in its application during natural disasters, contributing to crisis monitoring and response through AI-driven strategies. In healthcare, SA emerges as a valuable tool, advancing medical research and offering insightful glimpses into the emotional states of patients. Within the educational context, the study accentuates the significance of employing domain-specific strategies for SA, particularly in the evaluation of student feedback.

In the tourism and hospitality sectors, SA plays a pivotal role in categorizing and interpreting visitor remarks, shedding light on the efficiency of various ML algorithms, such as SVM and NB. The financial sector strategically employs sentiment analysis, utilizing SVM and LSTM, to forecast stock market movements and evaluate economic policies, showcasing its versatility and utility. Lastly, in the dynamic landscape of e-commerce, sentiment analysis facilitates aspect-based opinion mining, aiding customers in making informed product choices.

TABLE2 CHALLENGES OF SA [50], [51], [54], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [71], [72].

INDUSTRIES	PURPOSE	LIMITATION
Social Media	1. Develop a system that can automatically determine the ratio of positive and negative feedback in the Bangla Language.	1. The dataset of comments and feedback in Bangla is challenging to extrapolate, which might limit how broadly the results can be applied to other languages or situations.
	2. To explore the effectiveness of different ML algorithms for SA.	2. The paper only considers four algorithms, and there may be other algorithms that could perform better for this task.
Natural Disaster	1. To provide a method for utilising information from Twitter for conducting SA to identify the feelings of the public during emergencies.	1. The moral implications of SA via social media in emergencies, including privacy issues and possible biases in the data, are not covered by the study.
	2. To provide a completely automated approach for extracting location-oriented public views on world emergencies using AI and NLP.	2. Disaster monitoring does not raise any significant ethical issues, such as privacy or consent issues.
Healthcare	1. To present a distributed platform implementing a variety of text mining and ML techniques for sentiment analysis and classification of cancer medical records.	1. Some of the model's outcomes are unreliable, and they simplify too much the variety of feelings and experiences that cancer patients may express.
	2. To give an overview, provide the key findings of clinical narrative sentiment analysis in the healthcare field.	2. Lack of an ideal sentiment language adapted to the unique features of clinical accounts.
Education	1. To develop a model for SA of student comments by utilising a hybrid technique that integrates lexicon-based and ML techniques.	1. The difficulties of a certain domain in SA and the value of domain-specific resources in the context of educational evaluation generate precise results.
	2. To demonstrate SA as a substitute method for examining qualitative data in educational environments.	2. the opinions and views of early childhood educators on the use of ICT for young children, but it skips over their real-world methods.
Tourism and Hospitality	1. To examine the effectiveness of NB and SVM algorithms in analysing people's opinions regarding tourist destinations in Bandung.	1. The problem of imbalanced data can be addressed by utilising the SMOTE, which is essential for enhancing the SA model's accuracy.
	2. Evaluating the opinions that visitors have given on different tourist attractions in the different provinces.	2. The study concentrates on five Indonesian provinces, which might not accurately reflect the country's perception of tourism.
Finance	1. To provide an ML approach for financial market prediction from Twitter releases.	1. There is a need for improvement in the unclear word usage and insufficient dataset in tweets about the Indian banking sector.
	2. To conduct SA of the annual budget for the financial year 2016-17.	2. The robustness and generalizability of the outcomes are limited by the lack of statistical validation and other SA methodologies.
E-commerce	1. To apply aspect-based SA to analyse hair care product reviews to identify their aspect-wise polarity.	1. The system needs to be enhanced further to accommodate additional features for hair care products.
	2. To provide information regarding the SA methods used by ML and the possible uses of these techniques in the field of e-commerce.	2. Data scraping may be problematic because there is enough data to treat it as an actual public evaluation of various products.
Other Industries	1. To develop a novel approach to detect possible threats from insiders in corporations.	1. The absence of relevant benchmark datasets for the insider threat identification.
	2. To demonstrate an automated system for multidimensional analysis of events around the globe.	2. The limitations or problems encountered while creating and implementing the automated multidimensional analysis system must be thoroughly addressed.

The examination of SA presented here is just the beginning, indicating that there might be many undiscovered applications in various fields. This provides researchers with a potential roadmap, outlining successful methods used in different areas. The findings also indicate the possibility of further exploration and the discovery of new uses for sentiment analysis. This may lead to improved understanding and methods specifically tailored to different professional fields.

FUTURE WORK

The identified challenges in various domains underscore the complexities associated with SA methodologies. In social media, the scarcity of Bangla language datasets and a limited exploration of ML algorithms emphasize the necessity for broader linguistic diversity and comprehensive algorithmic considerations. Ethical dilemmas in privacy and biases emerge in the context of natural disaster-related SA, highlighting the need for a balanced ethical framework. Healthcare encounters challenges in oversimplifying patient sentiments, signaling a requirement for a more nuanced sentiment language. Educational sentiment analysis faces domain-specific hurdles and a dearth of insight into educators' methods. Tourism and hospitality grapple with imbalanced data and limited geographical representation. Financial SA confronts issues in unclear word usage and dataset insufficiency. E-commerce challenges include accommodating additional product features and addressing data scraping limitations. These challenges pave the way for future research, fostering a more sophisticated, ethical, and domain-specific approach to SA across diverse professional fields.

ACKNOWLEDGEMENT

We would like to thank Bournemouth University's Department of Computing and Informatics for allowing us to present our research work. We would especially like to express our gratitude to Dr. Philip Davies and Ogochukwu Ann Ijezie for their kind assistance and useful advice during the duration of the research. Their knowledge and support have been crucial in determining how to proceed with our research. We are grateful to our classmates for their contributions given through peer review, which is a cooperative atmosphere. This paper expresses acknowledgement for the provision of chances for academic growth and information sharing, which have enhanced our research experience.

REFERENCES

[1] Š. Zapletalová and H. Starzyczná, *Customer Behaviour in ECommerce: Case Studies from the Online Grocery Market*. Springer Nature, 2023.

[2] C.-L. Liu, W.-H. Hsaio, C.-H. Lee, G.-C. Lu, and E. Jou, "Movie rating and review summarization in mobile environment," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 42, no. 3, pp. 397-407, 2011.

[3] A. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis," in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2016: IEEE, pp. 628-632.

[4] D. Alessia, F. Ferri, P. Grifoni, and T. Guzzo, "Approaches, tools and applications for sentiment analysis implementation," *International Journal of Computer Applications,* vol. 125, no. 3, 2015.

[5] S. Shayaa *et al.*, "Sentiment analysis of big data: methods, applications, and open challenges," *Ieee Access,* vol. 6, pp. 37807-37827, 2018.

[6] T. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: A review," in *2017 International conference on inventive communication and computational technologies (ICICCT)*, 2017: IEEE, pp. 298-301.

[7] S. Vohra and J. Teraiya, "A comparative study of sentiment analysis techniques," *Journal Jikrce,* vol. 2, no. 2, pp. 313-317, 2013.

[8] K. Mouthami, K. N. Devi, and V. M. Bhaskaran, "Sentiment analysis and classification based on textual reviews," in *2013 international conference on Information communication and embedded systems (ICICES)*, 2013: IEEE, pp. 271-276.

[9] M. Araújo, A. Pereira, and F. Benevenuto, "A comparative study of machine translation for multilingual sentence-level sentiment analysis," *Information Sciences,* vol. 512, pp. 1078-1102, 2020, doi: 10.1016/j.ins.2019.10.031.

[10] P. Balaji, O. Nagaraju, and D. Haritha, "Levels of sentiment analysis and its challenges: A literature review," in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 2017: IEEE, pp. 436-439.

[11] N. Nandal, R. Tanwar, and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products," *Spatial Information Research,* vol. 28, no. 5, pp. 601-607, 2020, doi: 10.1007/s41324-020-00320-2.

[12] F. Wu, Y. Huang, Y. Song, and S. Liu, "Towards building a high-quality microblog-specific Chinese sentiment lexicon," *Decision Support Systems,* vol. 87, pp. 39-49, 2016.

[13] N. S. Joshi and S. A. Itkat, "A survey on feature level sentiment analysis," *International Journal of Computer Science and Information Technologies,* vol. 5, no. 4, pp. 5422-5425, 2014.

[14] M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment analysis: a comparative study on different approaches," *Procedia Computer Science,* vol. 87, pp. 44-49, 2016.

[15] E. H. Houssein, R. E. Mohamed, and A. A. Ali, "Machine learning techniques for biomedical natural language processing: a comprehensive review," *IEEE Access,* vol. 9, pp. 140628-140653, 2021.

[16] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems,* vol. 226, p. 107134, 2021.

[17] A. Abirami and V. Gayathri, "A survey on sentiment analysis methods and approach," in *2016 Eighth International Conference on Advanced Computing (ICoAC)*, 2017: IEEE, pp. 72-76.

[18] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 8, no. 4, p. e1253, 2018.

[19] A. Shaikh, N. A. Mahoto, and M. A. Unar, "Bringing shape to textual data - a feasible demonstration," *Mehran University Research Journal Of Engineering & Technology,* vol. 38, no. 4, pp. 901-914, 2019.

[20] J. Kocoń *et al.*, "ChatGPT: Jack of all trades, master of none," *Information Fusion,* p. 101861, 2023.

[21] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," *IEEE Computational intelligence magazine,* vol. 9, no. 2, pp. 48-57, 2014.

[22] E. Tyagi and A. K. Sharma, "Sentiment analysis of product reviews using support vector machine learning algorithm," *Indian Journal of Science and Technology,* vol. 10, no. 35, pp. 1-9, 2017.

[23] Y. H. Ko, P.-Y. Hsu, Y.-C. Liu, and P.-C. Yang, "Confirming Customer Satisfaction With Tones of Speech," *IEEE Access,* vol. 10, pp. 83236-83248, 2022.

[24] A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," in *2017 7th International annual engineering seminar (InAES)*, 2017: IEEE, pp. 1-4.

[25] S. Hammi, S. M. Hammami, and L. H. Belguith, "Advancing aspect-based sentiment analysis with a novel architecture combining deep learning models CNN and bi-RNN with the machine learning model SVM," *Social Network Analysis and Mining,* vol. 13, no. 1, p. 117, 2023.

[26] Z. Madhoushi, A. R. Hamdan, and S. Zainudin, "Sentiment analysis techniques in recent works," in *2015 science and information conference (SAI)*, 2015: IEEE, pp. 288-291.

[27] A. Mudinas, D. Zhang, and M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis," in *Proceedings of the first international workshop on issues of sentiment discovery and opinion mining*, 2012, pp. 1-8.

[28] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent systems,* vol. 28, no. 2, pp. 15-21, 2013.

[29] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep convolution neural networks for twitter sentiment analysis," *IEEE access,* vol. 6, pp. 23253-23260, 2018.

[30] A. Montoyo, P. Martínez-Barco, and A. Balahur, "Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments," *Decision Support Systems,* vol. 53, no. 4, pp. 675-679, 2012.

[31] T.-J. Lu, "Semi-supervised microblog sentiment analysis using social relation and text similarity," in *2015 International conference on big data and smart computing (BigComp)*, 2015: IEEE, pp. 194-201.

[32] D. R. Kawade and K. S. Oza, "Sentiment analysis: machine learning approach," *International journal of engineering and technology,* vol. 9, no. 3, pp. 2183-2186, 2017.

[33] B. Probierz, P. Stefański, and J. Kozak, "Rapid detection of fake news based on machine learning methods," *Procedia Computer Science,* vol. 192, pp. 2893-2902, 2021.

[34] Z. Lu *et al.*, "Natural language processing and machine learning methods to characterize unstructured patient-reported outcomes: validation study," *Journal of Medical Internet Research,* vol. 23, no. 11, p. e26777, 2021.

[35] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal,* vol. 5, no. 4, pp. 1093-1113, 2014.

[36] N. Mukhtar, M. A. Khan, and N. Chiragh, "Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains," *Telematics and Informatics,* vol. 35, no. 8, pp. 2173-2183, 2018, doi: 10.1016/j.tele.2018.08.003.

[37] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis," *HP Laboratories, Technical Report HPL-2011,* vol. 89, pp. 1-8, 2011.

[38] B.-H. Nguyen and V.-N. Huynh, "Textual analysis and corporate bankruptcy: A financial dictionary-based sentiment approach,"

Journal of the Operational Research Society, vol. 73, no. 1, pp. 102-121, 2022, doi: 10.1080/01605682.2020.1784049.
[39] J. Zhang, "A combination of Lexicon-based and classified-based methods for sentiment classification based on Bert," in *Journal of Physics: Conference Series*, 2021, vol. 1802, no. 3: IOP Publishing, p. 032113.
[40] C. S. Khoo and S. B. Johnkhan, "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons," *Journal of Information Science,* vol. 44, no. 4, pp. 491-511, 2018.
[41] V. Bonta, N. Kumaresh, and N. Janardhan, "A comprehensive study on lexicon based approaches for sentiment analysis," *Asian Journal of Computer Science and Technology,* vol. 8, no. S2, pp. 1-6, 2019.
[42] A. Gilmore and N. Millar, "The language of civil engineering research articles: A corpus-based approach," *English for Specific Purposes,* vol. 51, pp. 1-17, 2018, doi: 10.1016/j.esp.2018.02.002.
[43] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: a big data-ai integration perspective," *IEEE Transactions on Knowledge and Data Engineering,* vol. 33, no. 4, pp. 1328-1347, 2019.
[44] Ö. Canay and Ü. Kocabıçak, "An innovative data collection method to eliminate the preprocessing phase in web usage mining," *Engineering Science and Technology, an International Journal,* vol. 40, p. 101360, 2023.
[45] M. Nesca, A. Katz, C. K. Leung, and L. M. Lix, "A scoping review of preprocessing methods for unstructured text data to assess data quality," *International Journal of Population Data Science,* vol. 7, no. 1, 2022.
[46] T. Steuer, A. Filighera, and T. Tregel, "Investigating educational and noneducational answer selection for educational question generation," *IEEE Access,* vol. 10, pp. 63522-63531, 2022.
[47] M. E. Meguellati, R. B. Mahmud, S. B. A. Kareem, A. O. Zeghina, and Y. Saadi, "Feature Selection for Location Metonymy Using Augmented Bag-of-Words," *IEEE Access,* vol. 10, pp. 81777-81786, 2022.
[48] Q. Mao *et al.*, "Adaptive Pre-Training and Collaborative Fine-Tuning: A Win-Win Strategy to Improve Review Analysis Tasks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 30, pp. 622-634, 2022.
[49] M. H. Abd El-Jawad, R. Hodhod, and Y. M. Omar, "Sentiment analysis of social media networks using machine learning," in *2018 14th international computer engineering conference (ICENCO)*, 2018: IEEE, pp. 174-176.
[50] M. A. Shafin, M. M. Hasan, M. R. Alam, M. A. Mithu, A. U. Nur, and M. O. Faruk, "Product review sentiment analysis by using nlp and machine learning in bangla language," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, 2020: IEEE, pp. 1-5.
[51] S. Zahoor and R. Rohilla, "Twitter sentiment analysis using machine learning algorithms: a case study," in *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*, 2020: IEEE, pp. 194-199.
[52] D. Vidanagama, A. Silva, and A. Karunananda, "Ontology based sentiment analysis for fake review detection," *Expert Systems with Applications,* vol. 206, p. 117869, 2022.
[53] Y. W. Rabby and Y. Li, "Landslide inventory (2001–2017) of Chittagong hilly areas, Bangladesh," *Data,* vol. 5, no. 1, p. 4, 2019.
[54] H. J. Kaur and R. Kumar, "Sentiment analysis from social media in crisis situations," in *International Conference on Computing, Communication & Automation*, 2015: IEEE, pp. 251-256.
[55] F. K. Sufi and I. Khalil, "Automated disaster monitoring from social media posts using AI-based location intelligence and sentiment analysis," *IEEE Transactions on Computational Social Systems,* 2022.
[56] D. Georgiou, A. MacFarlane, and T. Russell-Rose, "Extracting sentiment from healthcare survey data: An evaluation of sentiment analysis tools," in *2015 Science and Information Conference (SAI),* 2015: IEEE, pp. 352-361.
[57] D. C. Edara, L. P. Vanukuri, V. Sistla, and V. K. K. Kolli, "Sentiment analysis and text categorization of cancer medical records with LSTM," *Journal of Ambient Intelligence and Humanized Computing,* vol. 14, no. 5, pp. 5309-5325, 2023.
[58] K. Denecke and D. Reichenpfader, "Sentiment analysis of clinical narratives: A scoping review," *Journal of Biomedical Informatics*, p. 104336, 2023.
[59] M. R. Yaakub, F. Z. M. Zaki, M. I. A. Latiffi, and S. Danby, "Sentiment analysis of preschool teachers' perceptions on ICT use for young children," in *2019 IEEE International Conference on Engineering, Technology and Education (TALE)*, 2019: IEEE, pp. 1-6.
[60] Z. Nasim, Q. Rajput, and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," in *2017 international conference on research and innovation in information systems (ICRIIS)*, 2017: IEEE, pp. 1-6.
[61] A. Arfilinia, R. Andreswari, F. Hamami, and J. M. F. Machado, "Multidimensional Sentiment Analysis of Tourism Object in DKI Jakarta, Banten, East Java, Central Java and West Java using Support Vector Machine Algorithm," in *2023 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, 2023: IEEE, pp. 1-6.
[62] Y. B. P. Pamukti and M. Rahardi, "Sentiment Analysis of Bandung Tourist Destination Using Support Vector Machine and Naïve Bayes Algorithm," in *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2022: IEEE, pp. 391-395.
[63] M. Shakeel and V. Karwal, "Lexicon-based sentiment analysis of Indian Union Budget 2016–17," in *2016 International Conference on Signal Processing and Communication (ICSC)*, 2016: IEEE, pp. 299-302.
[64] M. B. D. Pavithya, G. S. D. Perera, S. L. Munasinghe, S. N. Karunarathna, and S. L. A. A. th International Conference on Information and Automation for Sustainability Negambo, "Quantitative Analysis and Sentiment Analysis for Stock Price Forecast: The Case of Colombo Stock Exchange," in *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*: IEEE, 2021, pp. 512-517.
[65] M. Kothalawala and S. Thelijjagoda, "Aspect-based sentiment analysis on hair care product reviews," in *2020 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 2020: IEEE, pp. 228-233.
[66] P. V. Rajeev and V. S. Rekha, "Recommending products to customers using opinion mining of online product reviews and features," in *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*, 2015: IEEE, pp. 1-5.
[67] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," in *2018 IEEE international conference on innovative research and development (ICIRD)*, 2018: IEEE, pp. 1-6.
[68] H.-X. Shi and X.-J. Li, "A sentiment analysis model for hotel reviews based on supervised learning," in *2011 International Conference on Machine Learning and Cybernetics*, 2011, vol. 3: IEEE, pp. 950-954.
[69] M. Ghobakhloo and M. Ghobakhloo, "Design of a personalized recommender system using sentiment analysis in social media (case study: banking system)," *Social Network Analysis and Mining,* vol. 12, no. 1, 2022, doi: 10.1007/s13278-022-00900-0.
[70] S. Vanaja and M. Belwal, "Aspect-level sentiment analysis on e-commerce data," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2018: IEEE, pp. 1275-1279.
[71] C. Soh, S. Yu, A. Narayanan, S. Duraisamy, and L. Chen, "Employee profiling via aspect-based sentiment and network for insider threats detection," *Expert Systems with Applications,* vol. 135, pp. 351-361, 2019.
[72] F. K. Sufi and M. Alsulami, "Automated multidimensional analysis of global events with entity detection, sentiment analysis and anomaly detection," *IEEE Access,* vol. 9, pp. 152449-152460, 2021.
[73] J. Ramteke, S. Shah, D. Godhia, and A. Shaikh, "Election result prediction using Twitter sentiment analysis," in *2016 international conference on inventive computation technologies (ICICT)*, 2016, vol. 1: IEEE, pp. 1-5.
[74] K. Patel *et al.*, "Facial sentiment analysis using AI techniques: state-of-the-art, taxonomies, and challenges," *IEEE Access,* vol. 8, pp. 90495-90519, 2020.