Grenze International Journal of Engineering and Technology, Jan Issue

A Study on Various Sentiment Analysis for Mixed Transliterated Indigenous Language using Machine Learning Algorithms

Rishikesh J. Sutar[1] and Dr. Kamalakar R. Desai[2]
[1]Research scholar, E&TC Engg., DoT, Shivaji University, Kolhapur & Asst. Prof. in E&TC Engg., SCTR's PICT, Pune
[2]Professor, Electronics and Telecommunication Engineering, Bharati Vidyapeeth's College of Engineering, Kolhapur

***Abstract*—The evolution of Information Technology has led to the collection of large amounts of data, the volume of which has increased to the extent that in the last two years the data produced is greater than all the data ever recorded in human history. This has necessitated the use of machines to understand, interpret and apply data, without manual involvement. Sentiment Analysis has become the area of deep research due to the necessity wrought about by the advent of social media tools such as Twitter, Facebook, WhatsApp, and so on. In this survey, we will analyze 25 literature works concentrating on machine learning approaches associated with sentiment analysis for mixed transliterated languages, bringing into light the various shortcomings of the existing methodologies. Here, the analysis of various methods will be facilitated based on several factors, such as performance metrics, year of publication and journals, achievements of the techniques in numerical evaluations, and so on. On the other hand, the analysis of the methods with respect to the merits and demerits of the methods will be presented. Lastly, the paper discusses the potential future research directions and challenges in achieving better accuracy for sentiment analysis. The motivation of the research will be described in detail through the comparative discussion of the methods.**

***Index Terms*— Sentiment analysis, transliterated language, motivation, Machine learning, journals.**

I. INTRODUCTION

With the rapid development of information technology, the internet has taken on vital importance in people's lives. Most often, people express their ideas about various things, such as products and services, on online forums, blogs, social networking sites, and so on. [1][2]. Since more people are using social media, it is critical to protect the site from offensive content in order to promote personal safety online. Online content is more aggressive than offline content because anonymity is made possible by the internet [3]. These platforms contain useful data on a range of topics, including business, politics, and social applications [4][2]. The COVID-19 lockdown has significantly increased the number of online users on social media. Social networking platforms also offer free, open-source, and user-friendly methods for individuals to post material in their own local language or with mixed-code data[5]. Code-mixing is the use of linguistic units, words, phrases, and clauses at the sentence or utterance level across multiple languages. Typically, it is seen in a casual context like social media. Data that has been code-mixed can be easily collected from a number of social networking sites,

Grenze ID: 01.GIJET.9.1.324_1

including Facebook, Twitter, blogs, and various other platforms. Processing and analyzing code-mixed data still appears to be a difficult task [6] despite the vast amounts of data.

To extract irrational information from a source, computational linguistics and natural language processing methods are used in sentiment analysis [7][8]. Sentiment analysis and opinion mining have become hard and active areas of study for both languages with resources and those with limited resources. The broad concept of sentiment, evaluation, appraisal, or attitude of a piece of information that expresses the writer's or speaker's view is covered by the term "opinion" [9]. In the field of text, sentiment analysis of social media data has grown in popularity in recent years. One could determine someone's opinion by ascertaining the sentiment associated with a phrase or sentence [6]. Additionally, information for the sentiment analysis assignment can be gathered via social media, newspaper articles, online reviews, online discussion forums, and so on. But before doing sentiment analysis, the raw user-generated data from these platforms must be cleaned and standardized through a preprocessing procedure. This is because unstructured user-generated data are typically unsuitable for such analysis. Some of the pre-processing activities include stop word removal, filtering of URL links and special symbols.

Special symbols like emoticons, exclamation points, and other symbols can be employed as a feature in sentiment analysis when analyzing social media datasets [10]. When used on social data that is multilingual and code-mixed, traditional sentiment analysis methods that are effective for monolingual data suffer from a number of drawbacks. Therefore, there is enormous potential for creating unique algorithms for code-mixed sentiment analysis that make use of the ideas of grammatical transitions and code-switching [8]. CNN has been used to perform sentiment analysis on Hindi text. CNN's are used since they recently outperformed image classification and NLP in terms of performance [2].

II. LITERATURE REVIEW

Machine learning is a form of artificial intelligence, which helps software applications to predict outcomes more accurately without manual intervention. Sentiment analysis examines texts for polarity, ranging from positive, negative, and neutral. Machine learning technologies can automatically learn how to recognize sentiment without human input by being trained with samples of emotions in text. In recent data, due to the emergence of social media, sentiment analysis plays a crucial role in the machine learning field and multiple types of research are enabled based on this analysis, which is interpreted in detail in the below sections. The researchers are categorized based on the methods or classifier used shown in figure 1.

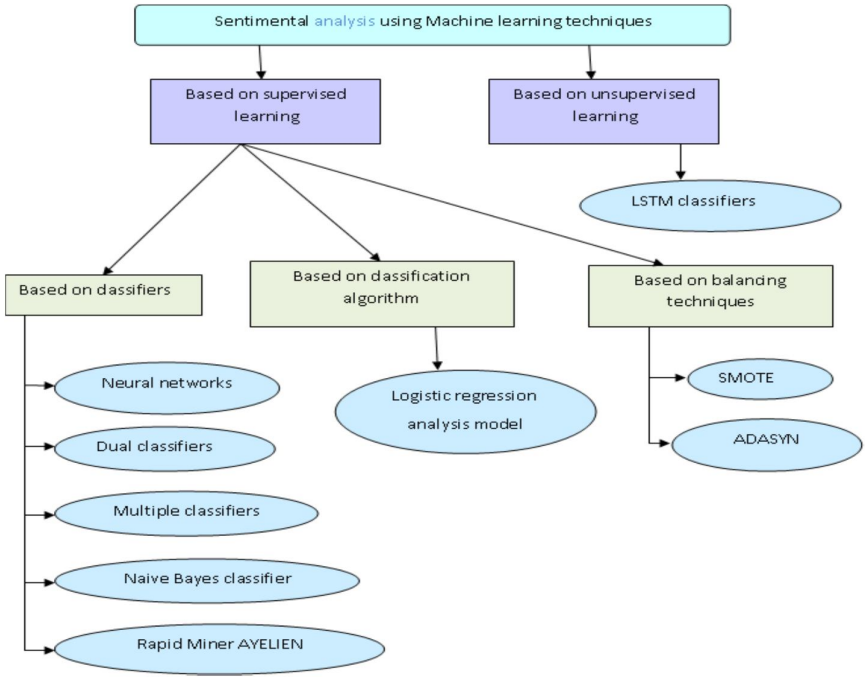

Figure 1. Categorization of the machine learning techniques based on sentimental analysis

A. Sentiment analysis based on Naïve Bayes classifier

The major objective of M. Vadivukarassi *et al.*[8] is to connect on Twitter and look for tweets that contain a specific keyword. After that, they assessed the polarity of the tweets and classified them as either positive or negative. Based on the feature selection of the score words, the sentiments of the online tweets are assessed. The Chi Square test is used to determine the best features, and the Naive Bayes classifier is used to train and evaluate the features in addition to more accurately determining the polarity of the comment. The difference between the legitimate candidate, who is expected to win the election based on their popularity with other metrics, and those who have only references with money power was underlined by B. P. Aniruddha Prabhu *et al.* [11]. Naive Bayes is a straightforward method for building classifier models that assign class labels to issue occurrences, represented as vectors of feature values, where the class labels are chosen at random from a small collection and a prediction formula based on their education, social status, and criminal activities, previous election records are also evaluated, which provides accurate classification.

B. Based on Dual classifiers

Mohammed Arshad Ansari and Sharvari Govilkar *et al.*[3] utilized multiple classifiers such as Naïve Bayes and Support Vector Machine (SVM) that classified the Hindi and Marathi texts, which is more flexible and adaptable. The words are identified on the daily usage of the individuals, which does not solely rely upon the dictionary for the identification of words, which acts as an add-on advantage. Ruchi Mehra *et al.* [5] used the naive Bayes and fuzzy Classifier to categorize the tweets into specific individual comments as positive, negative, or neutral. The authors created an ontology model that facilitates knowledge transfer between people and software agents, which made the domain more useful. Hewan Shrestha *et al.*[12] performed sentimental analysis of Hindi texts using the Naïve Bayas and decision tree classifier. The fundamental difference between the Hindi language and other languages was determined. The polarity of a text is determined by the fundamental structural differences between Hindi and English. The polarity of the words in the text is affected by the same set of words with minor modifications and changes in word order. As a result, in order to perform sentiment analysis on the Hindi language, a more thorough linguistic investigation is performed that effectively segregated the positive and negative comments. To determine the severity of the extremist feelings, Muhammad Asifb *et al.*[13] performed an emotional analysis of social media's multilingual text data using the classifier Multinomial Naïve Bayes and Linear Support Vector Classifiers. The study also divided the textual viewpoints it had included into four groups: high, low, moderate and neutral. By limiting the use of the same language, the technique also addressed the issue of repeated sentences which acts as an advantage for the research.

C. Based on neural networks

Merin Thomas and Latha [14] analyzed the sentiments using the Recurrent Neural Networks Long Short-Term Memory that achieved better accuracy even in increased and depth quality of datasets. The method mainly focused on the Malayalam tweets but is also flexible for analyzing the sentiments for other languages with larger data sizes. The technology employed in diverse locations comprehended the language's slang and the predictability of the sentiment analyzer, local syntax was better captured using regional slang and meaningful collocations. Sujata Rani and Parteek Kumar [15] performed the analysis based on Hindi movie reviews and annotated the data manually by three native speakers of Hindi. The sentiments are analyzed using the Convolutional neural network (CNN) and the results obtained showed that this model outperformed the basic machine learning models in classifying the texts into positive, negative, and neutral.

D. Based on LSTM classifiers

J. Shobana and M. Murali [16] established LSTM (long short-term memory) to recognize complex patterns in textual input. This method used Skip-gram architecture to extract word semantic and contextual information. This model reduced computational complexity and required less memory space and yielded higher accuracy. The majority of sentiment analysis research has been conducted in one language, and even the strongest monolingual approaches struggle in dwelling with that code-mixed language. When analyzing sentiment in code-mixed text, Siddharth Yadav and Tanmoy Chakraborty [1] proposed a technique that effectively transfers knowledge from monolingual text to code-mixed text using various types of multilingual and cross-lingual embeddings. Using the LSTM classifier, this technique also handled code-mixed text quite effectively in a zero-shot manner. Using LSTM classifiers, S. Shekhar *et al.*[4] created a classification model based on coding mixed social media data. The code mixed data included in this method dwelled with the languages English and Hindi and the words in the text are represented by word embedding features. The LSTM model's initial parameters are optimized using

artificial immune systems. The linguistic issue of polysemy is resolved by this method of ambiguous word intent detection.

E. Based on the Rapid Miner AYELIEN method

Ankita Sharma and Udayan Ghose [17] analyzed the sentiments using the Rapid-Miner AYLIEN which helped in extracting the information from the text without complexities. The AYLIEN also helped in entity extraction, language detection, hashtag suggestion, and related phrase operators are all included in this research. Since the majority of English words have a neutral polarity, the lexicons included in this package may not contain all of the English words. Since it solely uses unigrams, it cannot detect sarcasm or another emotive language. Another benefit is that the positive and negative sentiment scores aren't averaged to zero because the data is sentence-sized, which gives positive results. Deb Dutta Das *et al.*[18] concentrated on the sentimental analysis in the airline industry and obtained the opinion of the people using the comments provided. The sentiments are analyzed using the Rapid miner that helped in improvising the productivity by analyzing the positive and negative comments. The method helped in the easy interpretation of the data and provided information about the disliked aspects of the user to the owner. Better results are achieved by enabling the Naïve Baye's classifier in segmenting the data.

F. Based on data imbalance techniques

Wael M.S. Yafooz and Abdullah Alsaeedi [19] created a model that recognized and discovered the emotions or opinions of YouTube users on herbal treatment videos. In addition, an innovative Arabic Dataset on Herbal Treatments for Diabetes (ADHTD) is presented. This dataset is based on user feedback from several YouTube videos. The authors also resolved the issue of data imbalance using the SMOTE technique by initiating synthetic data points rather than duplicate data, which helped to increase accuracy. R. Srinivasan and C. N. Subalalitha resolved the issues aroused due to class imbalance and also initiated a new method called code mix for the classification of sentiments. The normalization of the base word was performed using the Levenshtein distance and the oversampling is performed using the SMOTE and ADASYN techniques. The advantage of the research was that it classified the languages such as Tamil, Hindi, Telugu, and Malayalam words.

G. Based on data imbalance techniques

Arnab Roy and Muneendra Ojha *et al.* [7] implemented three well-known deep learning models such as attention-based Bidirectional LSTM, Google BERT, and convolutional neural networks (CNNs) for Twitter sentiment analysis and preprocessed data to remove noise from data and increased the accuracy of models. Ranjan Kumar Behera *et al.* [20] introduced a hybrid technique combining two deep learning architectures, namely Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) for the purpose of classifying the sentiment of reviews. This Co-LSTM model is mainly aimed at two objectives in sentiment analysis. The primary benefit of this system was a sequential analysis of crucial aspects of a review was used for predicting the sentiment. Imane Guellil *et al.*[21] analyzed the sentiments in Arabic along with dialects. The Facebook messages that are written in the standard Arabic form, Algerian dialect, and Arabizi are analyzed. The sentiments are identified using the Random Forest, LSTM, CNN, and Logistic Regression Model. The polarities are classified using Word2Vec and fast text. Improved results are obtained by expanding both the sentiment lexicon and the sentiment annotated corpus.

H. Based on logistic regression analysis model

Multilingual or code-mixed data is quickly spreading over social media platforms due to the growing informal use of regional languages. When people post comments on social networks in more than one language, mixed code is produced. Such data poses a considerable barrier for sentiment analysis applications and has not yet been properly investigated by researchers. Pravalika A *et al.*[22] dwelled with multilingual, mixed language and classified the sentiments using lexicon-based approaches. Mixed language training data from Facebook was used to learn its grammatical transitions and frequently recurring rules. The analysis showed that the method worked well for code-mixed social data in a real-time environment. Bharathi Raja Chakravarthi *et al.*[23] focused on detecting the sentiments from the reviewer's comments made for the videos. Social media comments frequently use scripts that are not native to the author and do not strictly adhere to grammar rules. This issue is made more challenging by the lack of annotated code-mixed data for a language with limited resources like Tamil. The authors introduced a logistic regression analysis along with that the mixed Tamil English dataset also annotated for the polarity classification.

I. Others

Khaled Mohammad Alomari *et al.*[2] mined the opinion of the individuals using the SVM classifier and the weighing scheme is also introduced. The weighting scheme of the classifier based on the term frequency-inverse document frequency (TFIDF) was used, which helps in a better understanding of the contents present in the text. The combined usage of the weighting scheme of the classifier helped in achieving higher performance in classifying the Arabic text from the tweets. Twitter is a key tool for expressing ideas, opinions, and sentiments at various times. Twitter Sentimental Analysis is a technique for examining the feelings from tweets that might be useful in obtaining the user's sentimental values. Shobana G *et al.*[9] analyzed the ID and hashtag of famous individuals for comprehending the mindset of the people, which helps in the detection of positive and negative comments of the people more efficiently. The polarity of the text was identified using the Textblob, which offered a lot of features for the analysis of polarity. Raghavendra and K G Mohan [10] segregated and classified the tweets for providing optimized results to the individual through web minimization. The methodology used the K Nearest Neighboring (KNN) strategy to extract similar tweet strength using log threshold values. The framework was extended to pre-processed datasets from major social networking sites in order to retrieve data with a higher degree of accuracy. It also performed well with multiple clustering and cross-domain linkage using the tweeter API to separate positive and negative tweets. Ekagra Ranjan *et al.*[24] identified the abusive words in the Indic languages and resolved the challenges such as code-mixing along with transliteration. The predictions are made using the Random Forest classifier and the predictions are made robust if there exists an incorrect word, or unnoticed wrong words during training, it can also be corrected. Due to the non-standard spelling and grammar differences, automatic hate speech identification confronts many difficulties. This hate information would be in code-mixed form, which makes the process challenging, particularly for a country like India with a large population that speaks multiple languages and is bilingual. In order to identify hate speech in social media material that is combined with Hindi and English, our study proposes a machine learning algorithm. A machine learning model was created by Sreelakshmi *et al.* [25] that uses the Support Vector Machine (SVM)-Radial Basis Function (RBF) classifier to categorize Hindi-English tweets as either hateful or not. The findings indicated that character-level features are more informative for classifying code-mixed text than word and document-level features.

III. ANALYSIS BASED ON LITERATURE WORKS

The Sentiment analysis for mixed transliterated indigenous language using machine learning is analyzed in this section. In the below sections the analysis based on the parameter metrics used, dataset availed, software, year of publication, and journals published are analyzed.

A. Analysis based on the metrics

The analysis is made using the parameter metrics used by various researchers for proving the efficiency of the model. The metrics such as accuracy, precision, recall, f1 measure, Bilingual Evaluation Understudy (BLEU), sentimental score, processing time, confidence level, Mean Absolute Error (MAE), and Root-Mean-Squared Error (RMSE) are observed. The observations show that the accuracy of the metric is highly used by the researchers and the metrics precision, recall, and f1 measure are frequently used by the researchers, which is shown in table I.

TABLE I. ANALYSIS BASED ON METRICS

Metrics	**Paper**
Accuracy	[21][2][5][8][9][6][19][17][3][14][15][25][16][20][7][4]
Precision	[21][5][13][23][22][15][25][16][20][7]
Recall	[21][5][13][23][22][15][25][16][20][7]
F1 measure	[21][2][24][3][23][15][25][16][20][7][4]
BLEU	[21]
Sentimental score	[8][17][4]
Processing time	[10][15][16]
Confidence level	[17]
MAE	[15]
RMSE	[15]
Others	[12][18][11][1]

B. Analyses based on year of publication

In this section, the analysis is performed based on the year the journal was published. The works from the year 2017 to 2022 are reviewed. Most of the literary works relevant to the sentimental analysis are from the year 2020 and the detailed information is interpreted in figure 2.

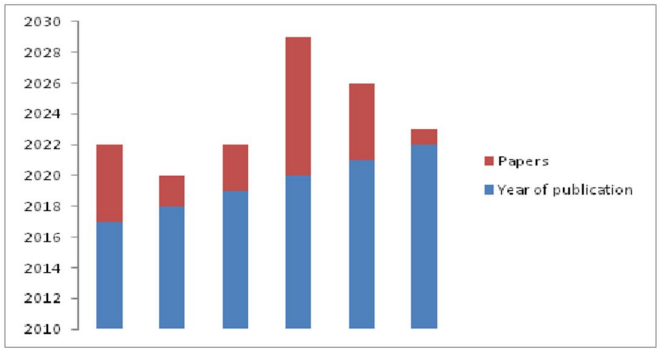

Figure 2. Analysis based on year of publication

In this section, the analysis is performed based on the year the journal was published. The works from the year 2017 to 2022 are reviewed. Most of the literary works relevant to the sentimental analysis are from the year 2020 and the detailed information is interpreted in figure 2.

C. Analyses based on journals

In this section, the papers that are published in the specified journals are investigated relevant to the sentimental analysis. The papers relevant to the sentiment analysis and classification are downloaded from various journals such as Springer, Academia, IEEE, researchgate, ArXiV, and Elsevier. Most of the papers from the journal researchgate are reviewed and the detailed analysis is interpreted in table II.

TABLE II: ANALYSIS BASED ON THE PUBLISHED JOURNALS

Journal name	**Papers**
Springer	[21][6][12][14][16]
Academia	[2][8]
IEEE	[5][22][7]
Researchgate	[9][19][11][3][15][20]
ArXiV	[24][23][1]
Elsevier	[13][10][17][25][4]
Others	[18]

D. Analysis based on the dataset

The analysis based on the dataset used by various researchers is interpreted in table 4, where various datasets such as corpora, Twitter, multilingual, monolingual, movie review dataset, Arabic Dataset on Herbal Treatments for Diabetes (ADHTD), and so on are reviewed. The datasets used by the reviewers are interpreted in figure 3.

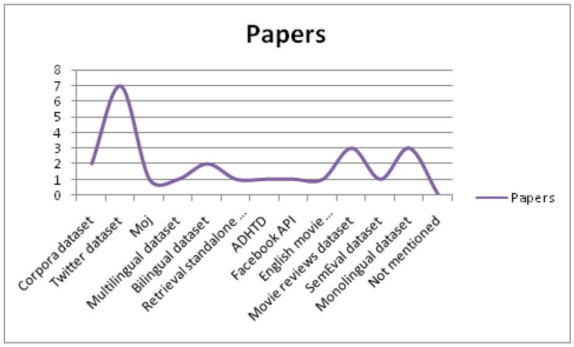

Figure 3. Analysis based on datasets used

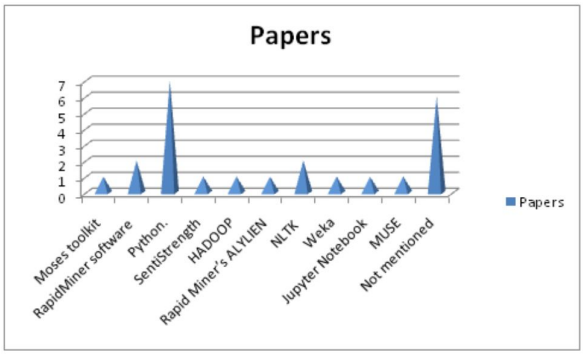

Figure 4. Analysis based on the software used

E. Analysis based on the software used

The analysis is performed based on the software used and most of the sentimental analysis methods use python for the classification of polarity in the sentiments. The significance of the software are graphically illustrated in figure 4.

IV. Research Gaps and Challenges

- The majority of studies on sentiment analysis have been conducted in English, whereas other languages had less research done on it [21]. Arabic language sentiment analysis research is few, and this lack is due to the complexity, multiplicity, and difficulty of assessing the Arabic language's composition [2].
- The difficulty with automatic annotation is coming up with a method that combines text, emoticons, and other features to increase annotation precision [8].
- Systems for sentiment analysis are often created for single-form languages like English. International sentiment, meanwhile, necessitates text analysis in a variety of local language forms. Additionally, the process of text analysis is hampered by dialectal diversity and linguistic richness issues [9].
- Users of Twitter, a social microblogging platform for social networking, can send tweets, and real-time messages to other users. Tweets feature a variety of unique properties, including brand-new difficulties and ways to conduct sentiment analysis on them in comparison to other domains [5].
- On the Internet, a variety of platforms are exploited for hostile purposes. These platforms are being used to create hate groups, biassed communities, disseminate radical ideologies, incite rage, and engage in violent behavior. Due to the enormous volume of data, automatic identification of online radicalization is a technically difficult task [13].
- One of the difficult issues for a Machine Learning (ML) application is the sentiment analysis of social data that has been code-mixed. Stemming, Parts of Speech Tagging (POS), and morphological analysis are examples of preprocessing techniques that are typically utilized in the emotional analysis of a monolingual text but are insufficient to assess feelings in code-mixed data. The cause of this is that code-mixed data contains more hidden lexicons and has poorly defined grammatical sentences [6].
- Since the tweets were gathered from an internet source, no sentiment descriptors were included. The lack of sentiment markers and the raucous nature of the tweets made the research more complicated [17].
- Due to the widespread use of irregular spellings and grammatical contractions between the languages utilized, analyzing multilingual social media information can be particularly difficult [22].
- The deep learning model is computationally expensive and needs a significant amount of data for proper training [20].

V. Conclusion

Using sentiment analysis, one can detect whether a text has feelings that are neutral, positive, or both. It is a type of text analytics that incorporates machine learning and natural language processing (NLP). Sentimental analysis is a crucial factor in today's world due to the emergence of social media in terms of analyzing consumer loyalty, satisfaction, the effectiveness of advertising and marketing, and product acceptance. The sentimental analysis could also be described as emotion detection or opinion mining process and here the investigation of sentimental

analysis using machine learning techniques is made using 25 research papers. The papers are analyzed in detail to provide a deeper insight into the techniques available for sentiment classification. The detailed analysis of the techniques paves a way for the researchers to further enhancement of the sentiment classification techniques and in the future, more papers relevant to sentiment classification could be reviewed for more efficient classification techniques to be innovated.

REFERENCES

[1] Yadav, Siddharth, and Tanmoy Chakraborty, "Unsupervised sentiment analysis for code-mixed data," arXiv preprint arXiv, vol. 2001, pp. 11384, 2020.

[2] Alomari, Khaled Mohammad, Hatem M. ElSherif, and Khaled Shaalan, "Arabic tweets sentimental analysis using machine learning," In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 602-610, Springer, Cham, 2017.

[3] Ansari, Mohammed Arshad, and Sharvari Govilkar, "Sentiment analysis of mixed code for the transliterated hindi and marathi texts," International Journal on Natural Language Computing (IJNLC), vol. 7, 2018.

[4] Shekhar, S., D. K. Sharma, D. K. Agarwal, and Y. Pathak, "Artificial immune systems-based classification model for code-mixed social media data," IRBM, 2020.

[5] Mehra, Ruchi, Mandeep Kaur Bedi, Gagandeep Singh, Raman Arora, Tannu Bala, and Sunny Saxena, "Sentimental analysis using fuzzy and naive bayes," In 2017 International Conference on Computing Methodologies and Communication (ICCMC), pp. 945-950, 2017.

[6] Srinivasan, R., and C. N. Subalalitha, "Sentimental analysis from imbalanced code-mixed data using machine learning approaches," Distributed and Parallel Databases, pp. 1-16, 2021.

[7] Roy, Arnab, and Muneendra Ojha, "Twitter sentiment analysis using deep learning models," In 2020 IEEE 17th India Council International Conference (INDICON), pp. 1-6. IEEE, 2020.

[8] Vadivukarassi, M., N. Puviarasan, and P. Aruna, "Sentimental analysis of tweets using Naive Bayes algorithm," World Applied Sciences Journal, vol. 35, no. 1, pp. 54-59, 2017.

[9] Shobana, G., B. Vigneshwara, and A. Maniraj Sai, "Twitter sentimental analysis," International Journal of Recent Technology and Engineering (IJRTE), vol. 7, 2018.

[10] Raghavendra, T. S., and K. G. Mohan, "Web mining and minimization framework design on sentimental analysis for social tweets using machine learning," Procedia Computer Science, vol. 152, pp. 230-235, 2019.

[11] Aniruddha Prabhu, B. P., B. P. Ashwini, Tarique Anwar Khan, and Arup Das, "Predicting election result with sentimental analysis using Twitter data for candidate selection," In Innovations in computer science and engineering, pp. 49-55. Springer, Singapore, 2019.

[12] Shrestha, Hewan, Chandramohan Dhasarathan, Shanmugam Munisamy, and Amudhavel Jayavel, "Natural language processing based sentimental analysis of Hindi (SAH) script an optimization approach," International Journal of Speech Technology, vol. 23, no. 4, pp. 757-766, 2020.

[13] Asif, Muhammad, Atiab Ishtiaq, Haseeb Ahmad, Hanan Aljuaid, and Jalal Shah, "Sentiment analysis of extremism in social media from textual information," Telematics and Informatics, vol. 48, pp. 101345, 2020.

[14] Thomas, Merin, and C. A. Latha, "Sentimental analysis of transliterated text in Malayalam using recurrent neural networks," Journal of Ambient Intelligence and Humanized Computing 12, no. 6, pp. 6773-6780, 2021.

[15] Rani, Sujata, and Parteek Kumar, "Deep learning based sentiment analysis using convolution neural network," Arabian Journal for Science and Engineering, vol. 44, no. 4, pp. 3305-3314, 2019.

[16] Shobana, J., and M. Murali, "An efficient sentiment analysis methodology based on long short-term memory networks," Complex & Intelligent Systems, vol. 7, no. 5, pp. 2485-2501, 2021.

[17] Sharma, Ankita, and Udayan Ghose, "Sentimental analysis of twitter data with respect to general elections in india," Procedia Computer Science, vol. 173, pp. 325-334, 2020.

[18] Das, Deb Dutta, Sharan Sharma, Shubham Natani, Neelu Khare, and Brijendra Singh, "Sentimental analysis for airline twitter data," In IOP Conference Series: Materials Science and Engineering, vol. 263, no. 4, pp. 042067, 2017.

[19] Yafooz, W. M., and Abdullah Alsaeedi, "Sentimental analysis on health-related information with improving model performance using machine learning," Journal of Computer Science, vol. 17, no. 2, pp. 112-122, 2021.

[20] Behera, Ranjan Kumar, Monalisa Jena, Santanu Kumar Rath, and Sanjay Misra, "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data," Information Processing & Management, vol. 58, no. 1, pp. 102435, 2021.

[21] Guellil, Imane, Ahsan Adeel, Faical Azouaou, Fodil Benali, Ala-Eddine Hachani, Kia Dashtipour, Mandar Gogate, Cosimo Ieracitano, Reza Kashani, and Amir Hussain, "A semi-supervised approach for sentiment analysis of arab (ic+ izi) messages: Application to the algerian dialect," SN Computer Science, vol. 2, no. 2, pp. 1-18, 2021.

[22] Pravalika, A., Vishvesh Oza, N. P. Meghana, and S. Sowmya Kamath, "Domain-specific sentiment analysis approaches for code-mixed social network data," In 2017 8th international conference on computing, communication and networking technologies (ICCCNT), pp. 1-6. IEEE, 2017.

[23] Chakravarthi, Bharathi Raja, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John P. McCrae, "Corpus creation for sentiment analysis in code-mixed Tamil-English text," arXiv preprint arXiv, vol. 2006, pp. 00206, 2020.

[24] Ranjan, Ekagra, and Naman Poddar, "Multilingual Abusiveness Identification on Code-Mixed Social Media Text," arXiv preprint arXiv, vol. 2204, pp. 01848, 2022.
[25] Sreelakshmi, K., B. Premjith, and K. P. Soman, "Detection of hate speech text in Hindi-English code-mixed data," Procedia Computer Science, vol. 171, pp. 737-744, 2020.