Decoding Sentiment: How Machine Learning Maps Emotions Across Domains

Dilasha Shakya
Minnesota State University
228 Wiecking Center
Mankato, MN

Rushit Dave
Minnesota State University
229 Wissink Hall
Mankato, MN

Dishant Thakkar
Tesla Inc.
San Francisco
California

ABSTRACT

The internet has become a useful platform for individuals to share their experiences, emotions and gather information. Whether it is getting a good coffee among a variety of options or searching for the ideal restaurant for dinner, people enjoy listening to other users' experiences and opinions before deciding to go there. Aside from personal uses, this digital space has become an excellent tool for businesses, politicians, health care sector and activists to understand their audience, take their feedback and take actions accordingly. It has also become an essential outlet for individuals to express their thoughts, share videos, and advocate against injustice, amplifying their voices in the pursuit of justice. Another usage of this might be in the areas like market research, political analysis, stock trends, and others, which will be further explored in the paper.

Sentiment analysis serves this purpose by utilizing machine learning techniques and natural language processing to identify emotions from textual data. The first stage is preprocessing of data. This involves removal of unnecessary words, punctuation, repeats, standardization of words and correction of spelling errors from unstructured data. Crucial features and adjectives are then extracted to determine their polarity. Various research uses different methods for calculating overall polarity. Rule-based approaches assign predefined scores to words and their synonyms, while others rely on sentence structure, word frequency, and intensity. Feature extraction methods include bag of words, word embedding, word count and noun count. These extracted features are then classified into machine learning models which help with analysis and prediction processes. The measure of performance is then evaluated using precision, recall, accuracy along with tools such as confusion and classification matrix.

Overall, sentimental analysis has emerged as a significantly growing area in recent years with great potential across various fields such as business, politics, healthcare, social media, fashion, crisis management, tourism. Therefore, proper research and increased model performance in sentiment analysis systems ensures the model can provide accurate insights aiding in better decision-making and planning across a wide range of applications.

General Terms

Sentiment, analysis, insights, research, opinions, features, machine learning

Keywords

Emotion detection, Natural Language Processing, Polarity detection

1. INTRODUCTION

In this era of technology, people have used social media and messaging as tools to express their ideas and raise concerns about events, businesses, products, politicians, etc. [1]. The platforms Twitter, Instagram, Reddit, and Facebook have not only become platforms for information sharing but for getting crucial ideas through the sentiments expressed in those sentences [2]. From businesses trying to sell products to politicians trying to win votes, knowing people's opinions is key. With so much information and influence on the internet, it is essential to have technology that can analyze the sentiments from opinions, and trends, differentiate fake and real data, and provide us with accurate information [3].

In the past decades, research has been ongoing on the study of using online reviews/messages on various subjects such as health, education, market, and tourism to gain useful information. Machine learning techniques have been used to take the unstructured data, clean, process those data, and use methods to analyze the sentiments underlying those reviews [4]. This process of analyzing data and tracking the sentiments which could be positive/neutral, or negative is sentiment analysis [4,5]. With this, various features can be extracted in different levels; Document, sentence, and feature levels which are subjected to how the data is structured and the specific objectives of the analysis [6]. Document level involves analyzing the document as a whole document while sentence level involves analyzing each sentence by sentence to come up with the sentiments expressed in those texts. Depending on the research and methods, these levels have been incorporated by research specialists.

This paper highlights different areas where sentiment analysis can be used and demonstrates the benefits of utilizing various machine-learning techniques to enhance its effectiveness. By mapping sentiments geographically, the paper aims to see where people feel good or bad about certain topics. Another significant contribution of our research is the investigation of demographic, news source, and geographic factors that influence sentiment dynamics. The research also includes the development of practical applications for sentiment analysis, such as monitoring brand perception on social media platforms like Twitter [7,8].

Most papers for research have been focused on the study of using Twitter data to comprehend opinions on various topics. In a paper, data collected from April to May 2013 from Twitter has been used to understand user choices for electronic devices such as mobile phones, and laptops. Tweets have been used to analyze data and tweets about people's opinions on tourism in Thailand. With COVID-19, and changes in travel restrictions, reviews were used to see sentiments around Thai tourism and ways to boost the economy post-pandemic [9].

For businesses, social media serves as an important channel to promote products, understand current trends, and customer reviews, and make decisions that can make an impact in the market [10]. The paper also takes us through the research done on Amazon reviews in real-time about beauty products and musical instruments [11]. Numerous studies have also been conducted on mixed language reviews. YouTube comments have been scraped to comprehend mixed language reviews and see which machine learning model works best.

Along with sentiment analysis on text messages and recommendation systems, the paper contributes to the growing body of knowledge in sentiment analysis, offering novel methodologies, insights, and applications to enhance our understanding of human emotions and opinions in textual data.

2. REVIEW METHODS

2.1 Research questions

The literature review aims to identify and analyze the applications of sentiment analysis models developed using various machine learning techniques. The review follows the dataset information, layout of how machine learning model works, sentiment analysis levels, approaches used for sentiment analysis, validation models. It then follows the application used in various research and applied in different areas with various challenges discussed and a conclusion.

Research Question (RQ): How have sentiment analysis techniques been successfully applied across various areas, and what machine learning methodologies, datasets, evaluation metrics, and challenges are associated with these applications?

2.2 Literature Selection

The literature search was conducted in three primary steps: identifying phase, diversity phase and summary phase.

2.21 Identifying phase:

In the identifying phase, the review investigated journal, research papers for academic search engines from 2010 to 2023. These search engines included IEE, Springer Link, Google Scholar, Science Direct, ACM Digital Library, ResearchGate, Web of Science etc. Important keywords such as sentiment analysis, machine learning, applications, challenges were used for the search process. This helped with selection of multiple papers for research and insights.

2.22 Diversity phase:

For this phase, the papers selected had different applications or used varied machine learning approaches to assess sentiment analysis. Scanning the title and abstract of the papers gave an idea of the content which helped decide which articles were to be selected and had diverse or varied applications. Around 16 articles were reviewed for detailed summary phase.

2.23 Summary phase:

Each paper reviewed was summarized in my own words in research notes, including sections on the introduction, methodologies, challenges, and research questions for analysis. Key insights and potential gaps were also noted to gain a deeper understanding of each study's contribution to the field.

3. DATASET

Table 1 provides a variety of datasets that scholars have utilized. The most used data sets were from twitter API, Stanford Sentiment Treebank (SST), SemEval, ISEAR and available public domains such as Apple, ICIC etc. Other datasets used were texts taken from YouTube comments, Amazon reviews, Google feedback forms and social media sites such as Twitter, Instagram, Facebook where Twitter has been used most often.

Table 1 Sentiment Analysis dataset information

Number	Datasets	Data Size	Domain
1.	Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis [12]	19,340 tweets 18,340 used for training and 1,000 for testing.	Twitter data
2.	Cornell Movie Review Dataset (polarity dataset v2.0)	1,000 positive and 1,000 negative movie reviews collected from IMDb	Movie reviews
3.	Application of Machine Learning Techniques to Sentiment Analysis [13]	Twitter API, including different domains like IT (Apple), Banking (ICICI), and Telecom (BSNL). The dataset sizes vary from 200 to 4,000 tweets. 70% of the dataset is used for training and 30% for testing.	Twitter data
4.	Analysis of Student Sentimental Feedback using Machine Learning Techniques [6]	dataset that consisted of student feedback 80 responses collected via Google Forms	Grading feedback
5.	Sentimental analysis from imbalanced code-mixed data using machine learning approaches. [14]	uses a bilingual dataset (Tamil-English) created by Chakravarthi et al. data was collected from YouTube comments and consists of 15,744 sentences.	Social media data for mixed languages

6.	Application of Support Vector Machine (SVM) in the Sentiment Analysis of Twitter Dataset [15]	Dataset: http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip 1.6 million records in the dataset, without empty records. 50% of the data has negative tags while the other 50% has positive tags	Twitter data
7.	Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning [16]	English-language tweets related to tourism in Thailand from July 1 to December 31, 2020. collected using the Twitter API and filtered with tourism-related keywords, such as "travel," "trip," and "tour," along with location-specific terms.	Tourism information in Thailand
8.	Movie recommendation and sentiment analysis using machine learning [17]	tmdb_5000_movies.csv and tmdb_5000_credits.csv), which are merged to form a single dataset containing columns like movie_id, title, and tags. Another dataset, reviews.txt, is used for sentiment analysis, which contains two columns: one for reviews and another for comments, with positive comments labeled as 1 and negative ones as 0 3,943 positive comments and 2,975 negative comments.	Movie selection
9.	Real-time Sentiment Analysis on E-Commerce Application [18]	product reviews collected from Amazon.com, specifically focusing on beauty products and musical instruments. total of 6,918 reviews, with 3,943 labeled as positive comments and 2,975 labeled as negative comments.	Product reviews
10.	A review on sentiment analysis and emotion detection from text [19]	Datasets used from Stanford Sentiment Treebank (SST), SemEval, ISEAR (International Survey of Emotional Antecedents and Reactions), and others like EmoBank and SS-Tweets.	Social media, business, healthcare
11.	Automated Framework for Real-Time Sentiment Analysis [20]	100 tweets gathered through the Twitter API then converted to csv file.	Advertising, politics, business

The paper utilizes various datasets from numerous sources to articulate ideas and come up with conclusions. Most datasets have used labelled data, which made it easier to determine the polarity. Twitter has been a popular platform for data collection. Tweets have been collected over the years using the Twitter API. The paper has also used the Cornell movie review dataset, which has been widely used in natural language processing to analyze sentiments. The data here are reviews collected from the Internet Movie Database (IMDb). Here, minimal pre-processing has been done to ensure the dataset is close to natural language for testing in practical scenarios. The dataset also contains challenging expressions such as sarcasm and complex sentence structures, which make it both a good and challenging dataset. Datasets from the Stanford Sentiment Treebank have also been used, which contain 1.6 million instances with movie reviews. This dataset helps understand sentiments at different levels. Apart from available datasets, the paper also includes datasets created and taken from Amazon reviews, Google Forms, a bilingual dataset, etc. All these datasets have contributed to providing various insights and information on sentiment analysis using machine learning and have helped in understanding the process in a detailed manner.

4. LAYOUT OF MACHINE LEARNING PROCESSES IN SENTIMENT ANALYSIS

4.1 Different levels of sentiment analysis: Sentimental analysis can be performed at various levels of granularity according to the goal and scope of the analysis. The types include: [12]

4.11 Document-level sentiment analysis: In this, the whole document is taken as the basic information unit to determine the sentiment as positive, negative or neutral. The sentiment of the entire document such as the review, article, is analyzed as a whole.

4.12 Sentence-level sentiment analysis: The sentiment of each sentence in the document or test is taken as basis to classify the sentiment as positive, negative or neutral. This allows for more fine-grained analysis to capture variations within a document.

4.13 Aspect-level sentiment analysis: For this, specific features or aspects of the product, service or topic are extracted and classified to give a positive, negative or neutral score. This helps where different features/ products are given different scores as they can have varied sentiments.

4.2 Different approaches for Sentiment Analysis

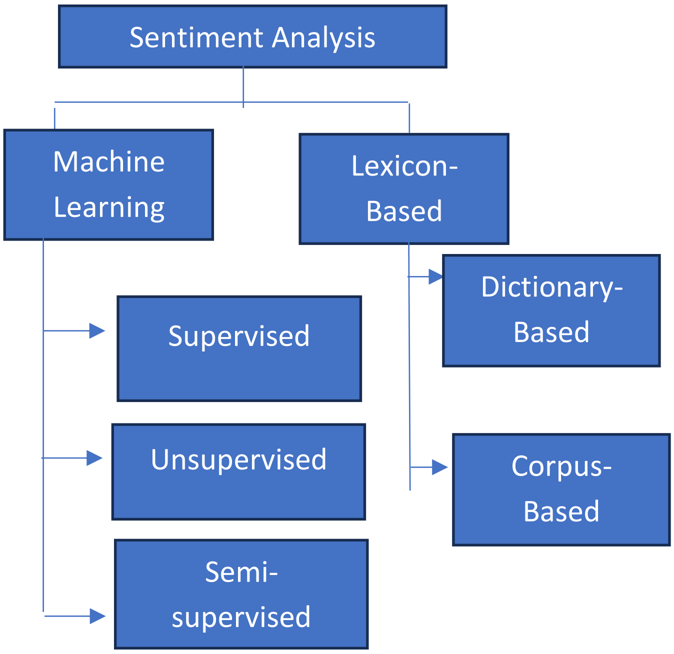

Figure 1. Different approaches

In Figure 1, you can observe the different approaches that can be used to perform sentiment analysis.

4.21 Lexicon-based:

This is an unsupervised approach that uses a set of pre-defined words from a dictionary or set that have been associated with a certain sentiment or score such as positive, negative or neutral words. The sentiment of the text is then calculated by sum or average of the scores or words.

To determine whether words are positive or negative, opinion lexicons are created in the following ways:

4.21.1 Dictionary-based approach: A limited set of opinion words with known orientations is initially collected manually. Synonyms and antonyms of these words are then sourced from corpora such as WordNet or a thesaurus and incorporated into the set. The dictionary expands until no new words are discovered.

4.21.2 Corpus based approach: They rely on large corpora for syntactic and semantic patterns of opinion words. The generated words are context specific and could require a large, labelled data set.

Lexicon based approach has been successfully incorporated to analyze news articles where they used a dictionary of pre-defined words assigned a sentiment. This technique was applied to a dataset of BBC news articles, calculating the overall sentiment of each article by summing the polarities of individual words. This helped reveal different sentiments expressed in different news categories [21]

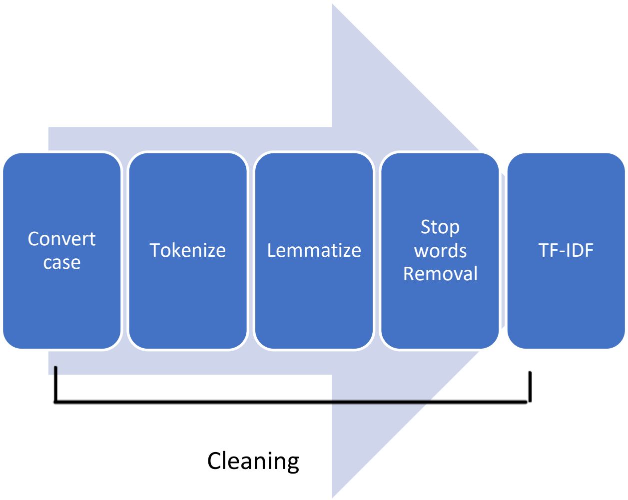

Figure 2. stepwise diagram of Naive Bayes performing sentiment analysis Cleaning

4.22 Machine learning-based: This method uses two sets of data; train data and test data. This method relies on a training model that learns from labelled data and predicts sentiments. After the trained data is tuned and ready, it is then applied to test data which is real-word data to see how the model functions. This can be divided into two types of learning: supervised and unsupervised learning.

4.22.1 Supervised learning methods: These methods rely on labeled data to perform predictions. Traditional models such as Naive Baiyes, Support Vector Machines, Decision trees have been used in various applications for sentiment analysis with their pros and cons. While SVM has proved out to have higher accuracy in most cases mostly because of its ability to find the optimal hyperplane to separate different classes particularly effective in high-dimensional spaces, making them suitable for sentiment analysis tasks with large feature spaces [15]. It has thus been able to provide more correct results than other models very often [22].

Naive Baiyes Method has been used for its simplicity in training and classifying data. It can learn patterns by examining set of documents that have been categorized. An example would be the review comment "I am happy' provides positive polarity as result [23]. It is used to describe an event's probability based on prior knowledge [12].

It works well for large-scale sentiment analysis because it trains quickly. However, it relies heavily on its training data. If unobserved words or new words are used, it may not do exact predictions. It can also suffer if there is bias or imbalance in the data. Figure 3 can be used to observe how Naive Baiyes performs the process sentiment analysis.

Maximum entropy handles features that are overlapped features. It aims to find the most unbiased way to make predictions based on the information available. Decision trees and random forests are great for large datasets. They work by comparing choices and opinions and separating them into different classes. They are helpful in sentiment analysis as they help visualize the decisions that are being carried out to perform the predictions. Random forests are also not sensitive to outliers and unbiased in decision making.

The K-nearest neighbor (KNN) classifier is an instance-based model that uses training documents' class labels that are comparable to the test document as its basis. An instance is categorized by its neighbors using a similarity-based vote system. KNN locates the k closest neighbors among training documents given a test document. The weight of the neighbor document's classes is determined by how similar each closest neighbor document is to the test document. [24]

4.3 Model Evaluation

Various performance measures are used to check how well the model performs with the training data. Some of the measures are: The matrix displays the number of instances produced by the model on the test data. This helps summarize and check results on test data. Fig6 can be taken as reference to see how a confusion matrix is combined of. **True Positive (TP):** The model correctly predicted a positive outcome. **True Negative (TN):** The model correctly predicted a negative outcome. **False Positive (FP):** The model incorrectly predicted a positive outcome. It has a negative outcome in actual value. **False Negative (FN):** The model incorrectly predicted a negative outcome. It has a positive outcome in actual value.

a) Accuracy: Fraction of correct predictions over total predictions, typically aiming for 70% to 90%. 100% accuracy or more may indicate overfitting.

b) Precision: Measures how accurately the model predicts each class, calculated by correct predictions over true positives and true negatives.

c) Recall: Measures the model's completeness for each class, determined by correct predictions over true positives and false negatives.

d) F-1 score: It is measured as:

F-score = (2* Precision*Recall)/ (Precision + Recall)

5. APPLICATIONS

5.1 Applications of sentimental analysis in different areas

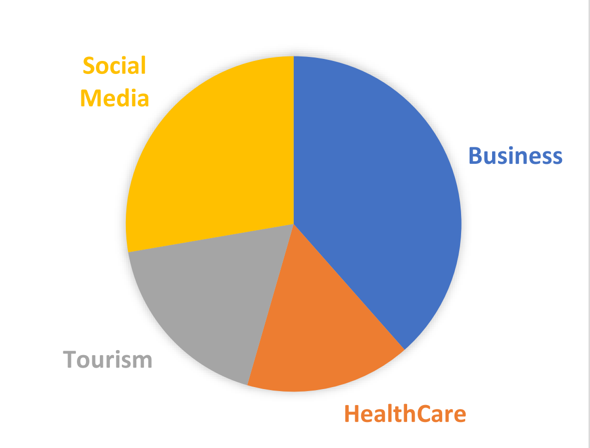

Fig 3. Pie chart for applications in different domains

5.11 Tourism Domain

Sentiment Analysis has been used to determine various places that attract tourists, places that tourists are less likely to come to, features that tourist do not like through twitter comments to assess where the tourism department needs to improve. This has been an efficient hands-on review from people who have visited the place and have experiences that they have shared through twitter.

It has taken into account the diverse ideas that people share in such a public platform where free voice of speech is practiced, taking this as an opportunity for them to improve their services.

In 2022, researchers analyzed English-language tweets to understand tourist sentiments and intentions in Thailand’s popular destinations such as Bangkok, Chiang Mai, and Phuket especially during COVID-19 [4]. They used machine learning algorithms (support vector machine, decision tree, and random forest) to classify tweets as positive, neutral, or negative, and to identify intentions to visit or avoid these areas. The study provided specific recommendations for tourism recovery, such as targeting marketing efforts on social media, highlighting safe travel experiences, and promoting the beauty of natural attractions. In these ways, sentiment analysis has been a great tool for getting useful information and insights from social media or texts and boost tourism by working on facilities that are most needed by the people.

5.12 Social Media Domain

Social media and online platforms are strong tools for ideas, opinions and reviews on various topics. The popular clothing brand, the new sparkling drink at the new restaurant or the buzzing news about a top star are all found on the internet. It is a place where people read news and freely share their thoughts and daily life activities. It is a place where people take and give inspiration. Business companies have social media accounts to promote their new launch or events. It extends to stock market performance, where investor sentiment, facilitated by social media interactions, can swiftly impact market trends. Political parties these days use social media to influence people for votes and most doctors also use social media to share prescriptions and health benefits to people. However, everything on the internet is not true and should not be trusted. This is why many companies use sentiment analysis to detect real sentiments as well as understand the hot topic, trends to focus on to help them stay aligned with what customers are looking forward to. Twitter is the best place where people frequently share updates and opinions and has been used by most researchers for the purpose of sentiment analysis. Shanshan Yi1 and Xiaofang Liu used social media, especially Twitter, to analyze customer reviews and opinions on products and shopping experiences [25]. The results helped researchers gain insights into customer satisfaction and preferences, which improved the accuracy of the product recommendations.

5.13 Business Domain

Sentiment Analysis has been used by business companies to get the most effective summary of customer reviews, assess customer’s likes and dislikes and make decisions on products based on customer reactions. When purchasing products, customers also view other customers reviews to understand how the product is and then purchase it. Hence, companies understand that it is essential to fix their product problems and stay on the young and old trend to attract customers. Researchers have performed a real-time sentiment analysis of amazon reviews of beauty products and musical instruments using SVM, where they were able to classify the reviews and share useful insights. [18]

5.15 Healthcare domain

Sentiment analysis has been recently used in healthcare and medicine sectors. This has helped gather more information about the patient's symptoms, disease treatments, medicine impacts, and surgery outcomes etc. [26]. Using platforms such as Twitter and Facebook to take user's feedback, providers understand patient experience. Using sentiment analysis, health care providers can monitor patient attitudes toward treatment plans and health services. This would help the providers give better support and take decisions that can help boost the medicine sector. An example is: there has been a study conducted to assess the mental health during COVID-19 through various online platforms which helped understand the overall feelings of people during the lockdown and take initiatives accordingly. [27]

6. LIMITATIONS/CHALLENGES

Despite the variety of methods for sentiment and emotion analysis, researchers encounter challenges such as contextual interpretation, dealing with mixed emotions, slang, and linguistic ambiguity which has been represented in Fig10. People have different ways to share their ideas which can be hard when it comes to analysing which emotion they are depicting. Teasing language can specifically be hard to interpret. Other challenges include manually labelling of large dataset which is time consuming and less reliable. Irony in sentences make it a tedious task to find the actual sentiment. The other challenge is the use of multiple emotions in a single sentence. It is difficult to determine various aspects and their corresponding sentiments or emotions from the multi-opinionated sentence. For instance, the sentence “view at this site is so serene and calm, but this place stinks” shows two emotions, 'disgust' and 'soothing' in various aspects [19]. Another challenge is that it is hard to detect polarity from comparative sentences. For example, consider two sentences 'Phone A is worse than phone B' and 'Phone B is worse than Phone A.' The

word ’worse’ in both sentences will signify negative polarity, but these two sentences oppose each other. It is a great challenge for researchers to develop a technique that can efficiently work in all domains. Some challenges the paper discusses is the lack of sentiment analysis lexicons and resources for other languages. Analysis also changes with time especially for election related events. Hence, real-time sentiment analysis and dynamic opinion mining is presented as a challenge in the paper. Working with labelled data is very expensive and therefore many are shifting towards unlabelled data for research. Another challenge that the papers have come across is the use of mixed languages in reviews which has been difficult to comprehend. For example, mixing Hindi and English language to write a review. Researchers have been focusing on such hybrid language models as well as models that can fix issues of web slang, language, sarcasm, irony and multiple expressions to take sentiment analysis and enhance it applicability. These advanced models improve accuracy in understanding nuanced human emotions, making sentiment analysis more versatile across various fields.

7. CONCLUSION

This paper provides a literature review of sentiment analysis process in stepwise order. Initially, the review discusses the various sentiment analysis levels which is sentence level, document level and sentence level. Different approaches for conducting machine learning such as lexicon-based approach, machine learning approaches and hybrid approaches were discussed. The stepwise layout of how the machine learning model works has been described through pre-processing, feature selection and model application processes. After this the model’s validation is checked. The applications of sentiment analysis in business, political, social media and health care domain were highlighted showcasing how sentiment analysis can be applicable in almost many different areas. Despite its flexibility in usefulness and advantages, sentiment analysis faces challenges in language, readability, context and dataset labelling. Recently, researchers have been working on going deeper into such challenges and building models that address such issues. Sentiment analysis models have the potential to understand the sentiments behind text and assist various areas in gaining valuable insights that can be incorporated into their respective fields. With continued research and development, the model could become an excellent tool for advancement.

8. REFERENCES

[1] Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human-centric Computing and Information Sciences, 7*, Article 32. https://doi.org/10.1186/s13673-017-0116-3

[2] Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for Twitter accounts. *Mathematical and Computational Applications, 23*(1), Article 11. https://doi.org/10.3390/mca23010011

[3] Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International Conference on the World Wide Web* (pp. 519–528). ACM Press. https://doi.org/10.1145/775152.775226

[4] Flores-Ruiz, D., Elizondo-Salto, A., & Barroso-González, M. d. l. O. (2021). Using social media in tourist sentiment analysis: A case study of Andalusia during the COVID-19 pandemic. *Sustainability, 13*(7), 3836. https://doi.org/10.3390/su13073836

[5] Coşkun, M., & Ozturan, M. (2018). europehappinessmap: A framework for multi-lingual sentiment analysis via social media big data (A Twitter case study). *Information, 9*(5), 102. https://doi.org/10.3390/info9050102

[6] Nasim, Z., Rajput, Q., & Haider, S. (2017). Sentiment analysis of student feedback using machine learning and lexicon-based approaches. *International Journal of Computer Applications, 161*(6), 1–5. https://doi.org/10.5120/ijca2017913229

[7] Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 497–501). Association for Computational Linguistics. https://aclanthology.org/N13-1053

[8] Cui, H., Mittal, V., & Datar, M. (2006). Comparative experiments on sentiment classification for online product reviews. In *Proceedings of the 21st National Conference on Artificial Intelligence* (pp. 1265–1270). AAAI Press.https://www.aaai.org/Papers/AAAI/2006/AAAI06-202.pdf

[9] Lu, Y., & Zheng, Q. (2021). Twitter public sentiment dynamics on cruise tourism during the COVID-19 pandemic. *Current Issues in Tourism, 24*(7), 892–898. https://doi.org/10.1080/13683500.2021.1889480

[10] Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation* (pp. 1320–1326). European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf

[11] Rain, C. (2013). Sentiment analysis in Amazon reviews using probabilistic machine learning. *Swarthmore College*. https://www.swarthmore.edu/sites/default/files/assets/documents/engineering/CS97_2013_Rain.pdf

[12] Gautam, G., & Yadav, D. (2014). Sentiment analysis of Twitter data using machine learning approaches and semantic analysis. In *2014 Seventh International Conference on Contemporary Computing (IC3)* (pp. 437–442). IEEE. https://doi.org/10.1109/IC3.2014.6897213

[13] Jain, A. P., & Dandannavar, P. (2016). Application of machine learning techniques to sentiment analysis. In *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)* (pp. 628–632). IEEE. https://doi.org/10.1109/ICATCCT.2016.7912109

[14] Srinivasan, R., & Subalalitha, C. N. (2023). Sentimental analysis from imbalanced code-mixed data using machine learning approaches. *Distributed and Parallel Databases, 41*, 37–52. https://doi.org/10.1007/s10619-021-07331-4

[15] Han, K.-X., Chien, W., Chiu, C.-C., & Cheng, Y.-T. (2020). Application of support vector machine (SVM) in the sentiment analysis of Twitter dataset. *Applied Sciences, 10*(3), 1125. https://doi.org/10.3390/app10031125

[16] Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., Yudha, A. K., & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon, 8*(1), e08795. https://doi.org/10.1016/j.heliyona.2022.e08795

[17] Pungliya, V., Raut, A., Bhonsle, R., Purohit, A., Patel, A., & Shashidhar, R. (2022). Movie recommendation and sentiment analysis using machine learning. *Global Transitions Proceedings, 3*, 279–284. https://doi.org/10.1016/j.gltp.2022.01.017

[18] Jabbar, J., Urooj, I., JunSheng, W., & Azeem, N. (2019). Real-time sentiment analysis on e-commerce applications. In *2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)* (pp. 391–396). IEEE. https://doi.org/10.1109/ICNSC.2019.8743331

[19] Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining, 11*, Article 81. https://doi.org/10.1007/s13278-021-00776-6

[20] Khanchi, I., Ahmed, E., & Sharma, H. K. (2020). Automated framework for real-time sentiment analysis. In *Proceedings of the 5th International Conference on Next Generation Computing Technologies (NGCT-2019)*. SSRN. https://doi.org/10.2139/ssrn.3702238

[21] Taj, S., Shaikh, B. B., & Meghji, A. F. (2019). Sentiment analysis of news articles: A lexicon-based approach. In *2019 2nd International Conference on Computing Mathematics and Engineering Technologies (iCoMET)* (pp. 1–5). https://doi.org/10.1109/ICOMET.2019.8673428

[22] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297. https://doi.org/10.1007/BF00994018

[23] Mubarok, M. S., & Adiwijaya, A. (2017). Aspect-based sentiment analysis to review products using Naïve Bayes. In *AIP Conference Proceedings* (Vol. 1867, No. 1, p. 020060). AIP Publishing. https://doi.org/10.1063/1.4994463

[24] El Mrabti, S., Al Achhab, M., & Lazaar, M. (2018). Comparison of feature selection methods for sentiment analysis. In Tabii, Y., Lazaar, M., Al Achhab, M., & Enneya, N. (Eds.), *Big Data, Cloud and Applications. BDCA 2018, Communications in Computer and Information Science* (Vol. 872). https://doi.org/10.1007/978-3-319-96292-4_21

[25] Yi, S., & Liu, X. (2020). Machine learning-based customer sentiment analysis for recommending shoppers, shops based on customers' reviews. *Complex & Intelligent Systems, 6*, 621–634. https://doi.org/10.1007/s40747-020-00147-2.

[26] Lai, S. T., & Mafas, R. (2022). Sentiment analysis in healthcare: Motives, challenges & opportunities pertaining to machine learning. In *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* (pp. 1–4). https://doi.org/10.1109/ICDCECE53908.2022.9792766.

[27] Dangi, D., Dixit, D. K., & Bhagat, A. (2022). Sentiment analysis of COVID-19 social media data through machine learning. *Multimedia Tools and Applications, 81*(29), 42261–42283.