Journal of Information Systems Engineering and Management
2025, 10(21s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comparative Analysis of Feature Selection Methods for Twitter Sentiment Classification

Mohamed Omar
Department of Computer Science,
Faculty of Computers and Informatics, Zagazig University,
Zagazig, Sharkia, 44519, Egypt
Ahmad Salah
Department of Information Technology,
College of Computing and Information Sciences,
University of Technology and Applied Sciences,
Ibri, Ad-Dhahirah, 516, Oman
Department of Computer Science,
Faculty of Computers and Informatics, Zagazig University,
Zagazig, Sharkia, 44519, Egypt
Mahmoud A. Mahdi
Department of Computer Science,
Faculty of Computers and Informatics, Zagazig University,
Zagazig, Sharkia, 44519, Egypt

ARTICLE INFO

Received: 12 Dec 2024
Revised: 30 Jan 2025
Accepted: 15 Feb 2025

ABSTRACT

In this paper, we investigate the performance of machine learning models designed for sentiment analysis of tweets using different featurization techniques: BoW and TF-IDF. With the development of social media at the current rate, it is becoming increasingly important to assess sentiments from short texts written in informal language. We analyze how BoW and TF-IDF impact the performance of different ML approaches, focusing on F1-score, precision, accuracy, and recall as the most important metrics for performance evaluation. We demonstrate via extensive experimentation on a large Twitter dataset that the use of TF-IDF with advanced ML models, such as Random Forest, significantly improves performance. These results demonstrate that the TF-IDF-using models outperform the prior benchmarks, per the reported evaluation metrics to date on the utilized dataset. The results also indicate that the SVM model with a TF-IDF vectorizer has the best performance among all compared models and vectorizer combinations by having the highest accuracy, precision, recall, and F1 score at 98%. Logistic Regression and Random Forest also yield very good performance, especially with Bows and TF-IDF, standing consistently around 92% to 96% for the metrics. The model of Naive Bayes has a more moderate performance of about 81% to 82% with all metrics. Overall, TF-IDF performs best among it is an especially effective vectorizer with SVM, thus making the classification tasks very effective. The developed model using SVM along with TF-IDF in this work outperforms other methods on F1-score and accuracy evaluation metrics.

Keywords: Classification; Machine learning; Prediction; Sentiment; Analysis; Tweets.

I.INTRODUCTION

Sentiment analysis of tweets has emerged as a vital domain within the paradigm of computational social science, mainly as part of natural language processing and big data analytics by several researchers in [1], [2]. Positive, negative, or neutral sentiments extracted from the tweets summarize information on social trends, consumer behavior, and public opinions, indicating a real-time pulse for social sentiment analysis as seen in [3], [19]. This ability is of great importance in a wide variety of applications, from market research and brand management [4] to political analysis [5] and disaster response [6]. The fine-grained and timely information of the sentiment labels of tweets therefore indicates that it is greatly valuable for shaping strategies and policies in various fields [7], [20].

In essence, the different techniques and models underpin SA towards increasing its accuracies and performances regarding opinion understanding in the social media texts are paramount reasons it plays an important role within the machine learning model [21], [22]. For example, a study has reviewed Random Forest and TF-IDF for feature extraction with the inclusion of intelligent WordNet lemmatization to raise the quality of the tweets by the removal of noise and application of lemmatization that greatly enhances the accuracy of sentiment classification in multi-class emotion text data [8]. Another study compares the performance between the Support Vector Machine model and the Random Forest model, and it has been witnessed that the Support Vector Machine outperformed the Random Forest in sentiment classification on social media text [9]. Moreover, research addressing sentiment analysis in the case of the Monkeypox outbreak has demonstrated that machine learning could effectively categorize public opinion from Twitter data, stressing the need to monitor public sentiment during health emergencies [10]. Essentially, these research works highlight the use of machine learning for sentiment analysis in effective data preprocessing, feature extraction, and classification, thereby providing various methodologies and applications in practice. Techniques such as TF-IDF and BoW feature extraction are key to text data processing in sentiment analysis. BoW is a simple technique that involves counting words in a piece of text with no consideration for word order.

The current research utilizes a proposed framework to analyze two different featurization methods, namely 1) TF-IDF [11] and 2) BoWs [12], on a dataset consisting of tweets along with their respective sentiment labels. The research methodology begins with data preprocessing, where the textual data is cleaned by removing special characters, numeric values, and stop words. Following this, tokenization and lemmatization are performed to ensure uniformity. The preprocessed tweets are then converted into numerical values through the application of either TF-IDF or BoWs. Different machine learning algorithms, such as Support Vector Machine (SVM) and Random Forest (RF), are used to classify the sentiments in the tweets. The models are trained and tested on a particular subset of the dataset. To evaluate the performance of the models, different performance measures—such as accuracy, F1-score, and recall—are utilized, while cross-validation methods are used to prevent overfitting and improve generalizability.

The following are the contributions of this study:

1) Extract sentimental features from tweets using two different techniques to study how each technique impacts the execution of various machine learning models.

2) The proposed methodology was thoroughly evaluated on a real-life dataset SA analysis as a classification problem. The data collected indicate that the suggested methodology outperformed the existing state of-the-art.

The following text expose the paper layout. Section II discusses relevant papers that used ML in Sentiment Analysis of tweets. The proposed methodology is detailed in Section III. Section IV discusses the outcomes of the suggested. Finally, the paper is concluded in Section V.

II.RELATED WORK

The study by Ferdiansyah et al. [13] is concerned with the assessment of public opinion on accessibility for people with disabilities in public spaces, using data obtained from the X platform, namely tweets. A preprocessing step was carried out to classify the data into positive and negative sentiments before applying the Support Vector Machine (SVM) algorithm, which used a Radial Basis Function (RBF) kernel. The SVM model achieved an accuracy of 83.37%, reflecting a high competence of SVM in the context of sentiment analysis (SA). A major limitation of this study is the exclusion of neutral sentiment, which could have provided a more refined insight into public opinions by improving the classification of user-generated tweets [13]. Likewise, Styawati et al. performed sentiment analysis on consumer reviews of online ride-hailing apps (Gojek and Grab) obtained from the Google Play Store. They applied Word2Vec for feature extraction, combined with SVM for classification, with accuracy rates of 89% for Gojek and 87% for Grab. This study supported the efficacy of the Word2Vec-SVM combination for sentiment analysis. However, a reported limitation is related to the self-labeling approach, which could be time-consuming and cumbersome, involving aspects of subjective assessment [14].

In [15], the authors tackled the problem of detecting euphemisms in multiple languages by introducing a new model that combines contextual and behavior-related features. The proposed system uses an ensemble classifier based on outputs from behavior-related finely-tuned models, which achieved a macro-averaged F1-score of 69%. A limitation reported is related to its reliance on contextual awareness, which could affect performance when encountering unfamiliar or previously unseen euphemisms.

The paper in [16] addresses sentiment analysis on tweets collected from the X platform, which is complex due to the nature of tweets. It proposes an LSTM model to classify sentiments as positive, neutral, or negative. The model achieved 93% for both accuracy and F1-score metrics; these scores are better than other models of comparison. However, the effectiveness heavily depends on two criteria of the training dataset, namely, the quality and size, as well as the tuning of hyperparameters. These results matching the findings of studies in [17,18, 19].

In [26], authors discuss the classification of sentiment in social media texts using a BoW model combined with several machine learning algorithms. The main focus is on the delicacies involved in automatic identification of the emotional tendency-a positive, neutral, or negative-from texts. Some other contributing factors include semantic ambiguity and the variety of expression in sentiments. The source documents include data collected from social media with details about the source, date, and additional emotional labels categorized into five broader groups: surprise, excitement, admiration, annoyance, and satisfaction, and finally, into three broader sentiment classes. The methodology uses Bag-of-Words model to convert text into numerical vectors and uses algorithms like Logistic Regression, Random Forest, and SVM. These models were evaluated for their accuracy in classifying sentiments. However, the computer being used faces these limitations in accurately classifying sentiments: within the framework, one of them is the usage of the model for semantic complexities and variation in text lengths. Further, works might focus and involve deep learning mechanisms and advanced natural language processing techniques to cater to the challenges mentioned above and enhance the quality of classification.

The authors cited in [27] explore sentiment analysis (SA) in the context of Arabic social media data, focusing specifically on consumer reviews of coffee products. The complexity of the Arabic language, with its complicated morphology, poses a significant challenge to successful sentiment analysis. They used a dataset of 10,646 Twitter reviews of different coffee products. To improve the accuracy and reliability of sentiment classification, feature selection extraction techniques, such as term frequency-inverse document frequency (TF-IDF) and minimum redundancy maximum relevance (MRMR), were used. This study employed four supervised machine learning algorithms for sentiment analysis: k-nearest neighbor, support vector machine, decision tree, and random forest; the results were then combined using ensemble learning techniques. The results show a significant improvement in accuracy and indicate the potential of sophisticated machine learning methods in overcoming the challenges of Arabic text analysis. In addition, the study recognizes a significant limitation in the generalizability of the results to other product categories or languages. The authors suggest that future research studies attempt to generalize these methods to address more complex problems in sentiment analysis.

The methods discussed demonstrate various efforts to utilize SA of tweets for different applications. Each paper employs distinct approaches for the converting the textual data to numerical data, yet the impact of different feature extraction methods, i.e., featurization has not been thoroughly studied. This research gap motivated the current work.

III.METHODOLOGY

A. Overview

Algorithm 1 begins by downloading and preprocessing the training and validation datasets, which are text and sentiment labels (lines 1-2). A function preprocess_text is implemented to perform word tokenization, text lowercasing, removal of stopwords, and stemming (lines 3-6). The preprocessed text thus obtained is replaced with the original text in both datasets, and sentiment labels are represented as numeric values (lines 7-8). The data is split into features (X) and labels (y), and a list to hold the results is created (lines 9-10). The algorithm employs nested loops (lines 11-18) to try different combinations of models (Naive Bayes, Logistic Regression, SVM, Random Forest) and featurization methods (BoW, TF-IDF). For each combination, it trains the model on the training set and tests it on the validation set using metrics like precision, accuracy, recall, and F1 score. The best-performing pair in terms of F1 score is selected (line 19) and then retrained from scratch on the whole training set (line 20). Finally, the model's performance on the test set is measured in order to see how well it can generalize (line 21).

Algorithm 1 Sentiment analysis using Machine Learning for differe Featurization

```
1: Load training and validation datasets
2: Preprocess data (remove NaN, rename columns)
3: function PREPROCESS_TEXT(text)
4:     Tokenize, lowercase, remove stopwords, and lemma-
       tize
5:     return preprocessed text
6: end function
7: Apply preprocessing to 'text' column in both datasets
8: Encode sentiment labels
9: Split data into features (X) and labels (y)
10: Initialize results list
11: for each model in [Naive Bayes, Logistic Regression,
    SVM, Random Forest] do
12:     for each vectorizer in [BoW, TF-IDF] do
13:         Create pipeline with vectorizer and model
14:         Fit pipeline on training data
15:         Predict on validation data
16:         Calculate accuracy, precision, recall, and F1 score
17:         Append results to results list
18:     end for
19: end for
20: Identify best model and vectorizer based on F1 score
21: Retrain best model with best vectorizer on full training
    data
22: Evaluate final model on validation set
```

B. Dataset

We have collected the dataset of different stocks from the Kaggle website on the following link (www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis, last accessed on 8-Feb-25). The dataset consists of Twitter entity sentiment analysis with train and validation datasets. Each record contains features like tweet id and sentiment (positive, negative, neutral, irrelevant, etc.), entity type, text, etc. Your training data has the sentiment distribution, where positive sentiment account for about 50%, while the negative, neutral and irrelevant sentiments are almost evenly distributed. The entity types are diverse. With an average tweet length (in characters) around 100, tweet lengths and richness of content vary greatly. The same holds for the validation set to ensure that the sentiment and entity types for the validation set are as similar to them as possible. The dataset has a good balance of sentiments and entity classes, such as negative, positive, irrelevant, neutral, that provide a good foundation for training models to carry out sentiment analysis on social media data.

C. Text Preprocessing

Text preprocessing is an important step in sentiment analysis (SA) [15,16]. The text data went through various preprocessing steps before being ready for feature extraction. These included tokenization, creating single words, or tokens from the text; in this case, stop words, or very frequent low-information words generally not considered to contribute much to the interpretation of the sentiment analysis (e.g., 'and', 'the', 'is'), were also cleaned from the data to reduce noise. Stemming was then applied in order to bring the words in consideration to their root form so that different forms of a single word can be treated as one token. The case of words was uniformly converted to lower case so that words differing in case do not get treated as different tokens. Such kind of preprocessing pipeline helps in making a clean and uniform dataset, which is very vital for proper feature extraction and model training.

D. Feature Extraction

It is considered that feature extraction to be one of the most crucial aspects of Sentiment Analysis [23,24,25]. It is an important step in Sentiment Analysis since it helps us recognize the attributes that are pertinent for the most accurate classification with regard to opinions and feelings. The selection and extraction of features enable improving of lower performance models and elementary workloads through which the models perform. Techniques such as mutual information, chi-square test, and a recursive feature elimination process among others are basically employed to select features that are key for SA models. An efficient feature selection can lead to lower noise interference and irrelevant data that may result in less generalization and a weaker sentiment analysis performance. Feature selection boosts up the basic estimate of a model and encourages prolonged analysis into the underlying features affecting sentiment orientations. Features obtained post-preprocessing are outlined in terms of two categories of techniques: BoWs and TF-IDF techniques. The BoW model sets out the vocabulary of all unique words in the text corpus and defines other features in terms of frequency of each word within this vocabulary for each document.

E. Utilized Models

We experiment with different machine learning models to classify the sentiment of the tweets. The models used in this study include SVM, Naive Bayes, and Random Forest classifiers. SVM is known for its robustness in high-dimensional spaces and its ability to find the optimal hyperplane that maximizes the margin between different classes. For the Random Forest classifier, we used 100 tress (n estimators=100).

Naive Bayes, on the other hand, is a probabilistic classifier based on Bayes' theorem, known for its simplicity and speed, particularly suitable for text classification tasks. Random Forest, a powerful ensemble method, fuse the decisions from multiple decision trees to enhance the scores of the accuracy metric and control overfitting. These models were chosen for their proven effectiveness in text classification tasks and their complementary strengths.

G. Evaluation Metrics

For comparing and analyzing the performance of the models, we employed various evaluation metrics such as precision, accuracy, recall, and F1-score. Accuracy in Eq. 1 calculates the proportion of correct predictions out of the total cases. Precision computes the proportion of true positive predictions out of the total positives predicted, as in Eq. 2, indicating the capability of the model in avoiding false positives. Recall is presented in Eq. 3. Recall calculates the proportion of correctly predicted positive instances out of the total positive instances, which estimates how well the model is in getting all the relevant instances. F1-score is the harmonic mean of precision and recall, a composite measure that considers both false positive and false negative. F1-score is indicated by Eq. 4. The formulae for these metrics are as follows:

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

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

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

$$F1 - Score = 2.\frac{Precision\,.Recall}{Precision+REcall} \quad (4)$$

where, TP refer to instances where the model successfully identifies positive cases. This metric shows how well the model can recognize what it is supposed to detect as positive, indicating its strength in correctly identifying relevant items. TN, for true negative, occurs when the model accurately identifies negative cases. This measure reflects the model's effectiveness in excluding items that should not be classified as positive, preventing unnecessary alerts. FP, for false positive represents the instances where the model mistakenly labels a negative case as positive. This can be problematic, as it leads to unwanted interruptions and reduces user trust in the system's decisions, indicating a need for improved precision. Finally, FN represent cases where the model fails to detect a positive instance and incorrectly classifies it as negative. This type of error can be critical in certain applications, as it means that the model is not capturing all the relevant positive instances, reducing its recall performance.

IV.RESULTS

A. Setup

The research was conducted on a personal computer that had a 64-bit Windows 11 operating system. The computer had an Intel processor with 2 cores, each with a frequency of 2.6 GHz, and was also set up with 24 GB of RAM for effective data processing. In the execution of the machine learning models, the Scikit-learn libraries version 1.4.2 was utilized. The source code used throughout these experiments, as well as the data used, are publicly available. The source code of this research is downloadable at the following link: https://github.com/stars-of-orion/SA-Analysis-Featurization, last accessed on the 12th of October, 2024.

B. Results

Models attempted include Naive Bayes, Logistic Regression, SVM, Random Forest, and Gradient Boosting. They were evaluated based on two types of feature extraction techniques: WoB and TF-IDF Vectorizer. Accuracy, precision, recall, and F1 score were the key metrics on the basis of which their performances were evaluated. Accuracy was 98%, precision was 97%, recall was 98%, and F1 score was 98% for SVM with TF-IDF Vectorizer. It was the highest among all the models. Logistic Regression with TF-IDF Vectorizer also performed well, with accuracy being 92%, precision being 92%, recall being 92%, and F1 score being 92%. In general, models based on the TF-IDF Vectorizer performed better than those based on the WoBs. The results are presented in Table I. Overall, the SVM model combined with the TF-IDF Vectorizer proved to be the best-performing model for this task, highlighting the imperative nature of choosing appropriate feature extraction methods. This study highlights the need to find the right model and feature extraction method combination to achieve optimum effectiveness in sentiment analysis activities.

The performance metrics for the various models using the different vectorizers are provided in Table I: Bows and TF-IDF. Naive Bayes performed unusually well, apparently equally with both Bows and TF-IDF, with an average of about 82% performance for all four evaluation metrics utilized. This means that Naive Bayes is performing quite average, not that it is good or the best model. Logistic Regression yields better results (the highest), with 95% for BoWs and 92% for TF-IDF metrics. This means that Logistic Regression goes one notch up, but using TF-IDF yields slightly lower performance.

The SVM has performed quite impressively, especially using TFIDF, attaining 98% on all the metrics. This hints at SVM performing exceptionally with TF-IDF features. It also has a solid performance with WoBs, reaching 96%, but not quite as good as TF-IDF. Random Forest model is on par for BoWs and TF-IDF, scoring 96% for all four evaluation metrics utilized. This consistency indicates that Random Forest model is a strong choice across vectorization methods. SVM with TF-IDF will likely have the best classification accuracy and reliability for this is a specific dataset.

Table I qualifies TF-IDF as the best vectorization method as the models meeting this benchmark are generally achieving higher or at least similar performance compared to BoWs. For instance, the SVM model, with TF-IDF, attained the highest scores—across the four evaluation metrics considered—close to 98%, which is greater than any other combination. Similarly, Logistic Regression and Random Forest models also produce strong TF-IDF performance, yielding consistent performance at all metrics. Quite simply, this suggests that TF-IDF captures more meaningful features from the text data, hence making it the most useful vectorizer for this dataset. Using TF-IDF allows for better model performance as it focuses on unique and relevant terms to devalue the significance of high-frequency, less-informative words.

The performance metrics for classification shown in Table II display consistently high results for all sentiment categories. The model achieves an overall accuracy of 98%, show its robust ability to differentiate between various sentiment classes. Precision, recall, and F1-scores are uniformly elevated across the four sentiment categories—Irrelevant, Negative, Neutral, and Positive—with scores ranging from 97% to 98%. This uniformity indicates that the model is effectively identifying all types of sentiment without a noticeable preference for any specific class. The Positive sentiment category exhibits slightly lower metrics, with precision, recall, and F1-score all at 97%, whereas the other categories achieve a solid 98% across these metrics. The macro average scores of 98% for precision, recall, and F1-score emphasize the model's balanced performance across all categories.

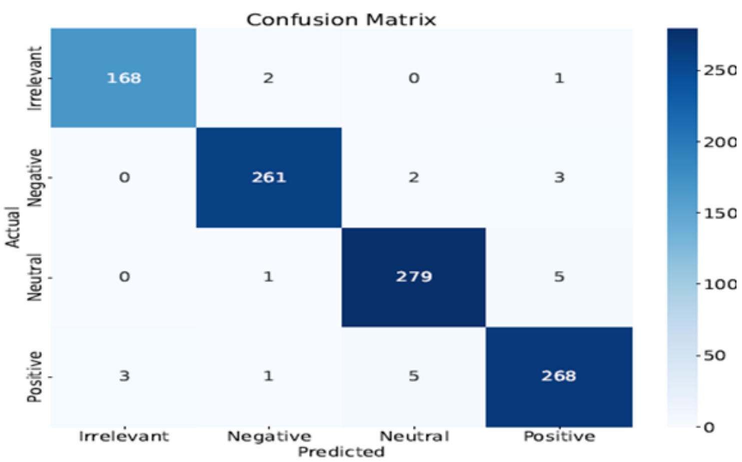

Fig. 1: Confusion matrix of the SVM and TF-IDF model.

The SVM model's confusion matrix has been shown in Fig. 1. The SVM model confusion matrix from TF-IDF represents four categories: Irrelevant, Negative, Neutral, and Positive. The model correctly classified 168 cases as Irrelevant and misclassified only three instances for this class. The model performed fairly well with the Negative category, classified 261 cases with only slight misclassifications of 2 cases as Neutral and 3 cases as Positive. In the Neutral category, the model predicted 279 instances correctly and misclassified only 6 instances. For the Positive category, the model reported 268 cases as correct only 9 instances misclassified. The results of this confusion matrix imply that the model did well at identifying the commons or broader classes but still confuses a few categories. Most confusing between Positive and Neutral shows that these classes share some commonality features. The indeterminate di-vision might also arise due to overlapping features of the text data. Feature extraction method used or tuning the model parameters can serve to reduce these mistakes. Overall, it is a strong model, but connectivity tweaks can help improve performance.

TABLE I. PERFORMANCE METRICS FOR DIFFERENT MODELS AND VECTORIZERS.

Model	Vectorizer	Accuracy	Precision	Recall	F1-Score
Naive Bayes	**Bows**	82%	83%	82%	82%
	TF-IDF	81%	84%	81%	81%
Logistic Regression	**Bows**	95%	95%	95%	95%
	TF-IDF	92%	92%	92%	92%
SVM	**WoBS**	96%	96%	96%	96%
	TF-IDF	98%	98%	98%	98%
Random Forest	**Bows**	96%	96%	96%	96%
	TF-IDF	96%	96%	96%	96%
Gradient Boosting	**Bows**	58%	64%	58%	57%
	TF-IDF	58%	64%	58%	57%

TABLE II. CLASSIFICATION PERFORMANCE METRICS PER EACH CLASS.

Class	Precision (%)	Recall (%)	F1-score (%)
Irrelevant	98%	98%	98%
Negative	98%	98%	98%
Neutral	98%	98%	98%
Positive	97%	97%	97%
Accuracy	98%		
Macro avg	98%	98%	98%
Weighted avg	98%	98%	98%

TABLE III. COMPARISON BETWEEN THE PROPOSED WORK AGAINST THE EXISTING METHODS.

Model	F1-score	Accuracy
Ours (SVM+TF-IDF)	98%	98%
[15]	95%	96%
[16]	Not Reported	93%
[18]	Not Reported	87%

Observed results in Table III confirm the promising performance of the proposed work against existing methods due to F1-score and accuracy values. The proposed model achieved an F1-score of 98% and accuracy of 98%, which far exceeded the work in [15] which delivered a score of 95% and accuracy of 96%. Also, as not all had reported the F1-scores but delivered lower accuracy of 93% and 87%, respectively. The results showed the robust and more performant behavior of the proposed model over the given task indicating a considerable improvement from previous works. Whenever a metric was not yet effectively analyzed, it was reported as being not applicable (not reported).

V.CONCLUSIONS

To conclude, this study quantitatively demonstrated a significant impact of these featurization techniques on the quality of machine learning models to perform a sentiment analysis of tweets. By comparing BoW and TF-IDF methods, the SVM classifier models using TF-IDF performed significantly better than the BoW across the board, according to relevant evaluation metrics like accuracy, precision, recall, and F1-score, with noise negative, neutral, and positive indicators taken into account. In our study, the SVM model with TF-IDF was capable of producing the best performance, culling out the importance of selecting adequate feature extraction techniques for favorable model performance. The results should inspire future research in this area and provide useful information for practitioners, to improve sentiment analysis systems basically based on social media data. Future work can focus on other featurization methods and propagation of those featurization methods through the use of advanced machine learning models to further improve sentiment analysis.

ACKNOWLEDGEMENT(S)

The authors would like to thank Mr. Eslam Elhadidi for his invaluable technical assistance and expert advice on the coding aspects of this project.

REFERENCES:

[1] X. Zhou, X. Tao, J. Yong, and Z. Yang, "Sentiment analysis on tweets for social events," in Proceedings of the 2013 IEEE 17th international conference on computer supported cooperative work in design (CSCWD). IEEE, 2013, pp. 557–562.

[2] P. William, A. Shrivastava, P. S. Chauhan, M. Raja, S. B. Ojha, and K. Kumar, "Natural Language processing implementation for sentiment analysis on tweets," in Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022. Springer, 2023, pp. 317–327.

[3] K. Darshan, J. Samuel, C. Manjunatha Swamy, P. Koparde, and N. Shivashankara, "NLP-Powered Sentiment Analysis on the Twitter," Saudi J Eng Technol, vol. 9, no. 1, pp. 1–11, 2024.

[4] L. M¨ahnert, C. Meyer, U. R. Orth, and G. M. Rose, "Brand heritage on Twitter: a text-mining stereotype content perspective," Journal of Product & Brand Management, vol. 33, no. 3, pp. 370–384, 2024.

[5] L. Hagemann and O. Abramova, "Sentiment, we-talk and engagement on social media: Insights from Twitter data mining on the US presidential elections 2020," Internet Research, vol. 33, no. 6, pp. 2058–2085, 2023.

[6] J. Ai, X. Lv, and D. Gursoy, "Impact of social media posts on travelers' attitudes and behaviors towards a destination after a natural disaster: moderating role of the source of the post," Journal of Sustainable Tourism, vol. 31, no. 11, pp. 2460–2478, 2023.

[7] K. L. Tan, C. P. Lee, and K. M. Lim, "A survey of sentiment analysis: Approaches, datasets, and future research," Applied Sciences, vol. 13, no. 7, p. 4550, 2023.

[8] S. Saranya and G. Usha, "A Machine Learning-Based Technique with IntelligentWordNet

Lemmatize for Twitter Sentiment Analysis," Intelligent Automation & Soft Computing, vol. 36, no. 1, 2023.
[9] T. A. Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. B. M. Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," Journal of Informatics and Web Engineering, vol. 3, no. 1, pp. 67–75, 2024.
[10] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A machine learning-sentiment analysis on monkeypox outbreak: An extensive dataset to show the polarity of public opinion from Twitter tweets," IEEE Access, vol. 11, pp. 11811–11826, 2023.
[11] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information processing & management, vol. 24, no. 5, pp. 513–523, 1988.
[12] C. D. Manning, Introduction to information retrieval. Syngress Publishing, 2008.
[13] H. Ferdiansyah, N. Komaria, and I. Arief, "The Application of Support Vector Machine Method to Analyze the Sentiments of Netizens on Social Media Regarding the Accessibility of Disabilities in Public Spaces,"Journal of Information System, Technology and Engineering, vol. 1, no. 1, pp. 6–10, 2023.
[14] Styawati, A. Nurkholis, A. A. Aldino, S. Samsugi, E. Suryati, and R. P. Cahyono, "Sentiment analysis on online transportation reviews using Word2Vec text embedding model feature extraction and support vector machine (SVM) algorithm," in 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE). IEEE, 2022, pp. 163–167.
[15] F. Vitiugin and H. Paakki, "Ensemble-based Multilingual Euphemism Detection: a Behavior-Guided Approach," in Proceedings of the 4th Workshop on Figurative Language Processing (FigLang 2024), 2024, pp. 73–78.
[16] N. O. Aljehane, "A New Approach to Sentiment Analysis on Twitter Data with LSTM," in 2023 3rd International Conference on Computing and Information Technology (ICCIT). IEEE, 2023, pp. 657–663.
[17] W. Stigall, M. A. A. H. Khan, D. Attota, F. Nweke, and Y. Pei, "Large Language Models Performance Comparison of Emotion and Sentiment Classification," in Proceedings of the 2024 ACM Southeast Conference, 2024, pp. 60–68.
[18] A. C. M. V. Srinivas, C. Satyanarayana, C. Divakar, and K. P. Sirisha, "Sentiment analysis using neural network and LSTM," in IOP conference series: materials science and engineering, vol. 1074, no. 1. IOP Publishing, 2021, p. 012007.
[19] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," Knowledge and Information Systems, vol. 60, pp. 617–663, 2019.
[20] M. Rodr´ıguez-Ib´anez, A. Cas´anez-Ventura, F. Castej´on-Mateos, and P.- M. Cuenca-Jim´enez, "A review on sentiment analysis from social media platforms," Expert Systems with Applications, vol. 223, p. 119862, 2023.
[21] R. K. Das, M. Islam, M. M. Hasan, S. Razia, M. Hassan, and S. A. Khushbu, "Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models," Heliyon, vol. 9, no. 9, 2023.
[22] D. S. Asudani, N. K. Nagwani, and P. Singh, "Impact of word embedding models on text analytics in deep learning environment: a review," Artificial intelligence review, vol. 56, no. 9, pp. 10345–10425, 2023.
[23] C. P. Chai, "Comparison of text preprocessing methods," Natural Language Engineering, vol. 29, no. 3, pp. 509–553, 2023.
[24] K. S. Eljil, F. Nait-Abdesselam, E. Hamouda, and M. Hamdi, "Enhancing Sentiment Analysis on Social Media with Novel Preprocessing Techniques," J. Adv. Inf. Technol, vol. 14, no. 6, pp. 1206–1213, 2023.
[25] J. Cao, J. Li, M. Yin, and Y. Wang, "Online reviews sentiment analysis and product feature improvement with deep learning," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 8, pp. 1–17, 2023.
[26] S. Lee, "Social Media Text Sentiment Classification Based on Bag-of-Words Models with Multiple Machine Learning Algorithms," in 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT). IEEE, 2024, pp. 1775–1779.
[27] O. Alsemaree, A. S. Alam, S. S. Gill, and S. Uhlig, "Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions," in Heliyon, vol. 10, no. 9. Elsevier, 2024, p. 012007.