Assessing Institutional Performance using Machine Learning on Arabic Facebook Comments

Zainab Alwan Anwer
Department of Software, Faculty of Computer Science and Information Technology, Wasit University, Iraq
zainab.alwan@uowasit.edu.iq (corresponding author)

Ahmad Shaker Abdalrada
Department of Software, Faculty of Computer Science and Information Technology, Wasit University, Iraq
aabdalra@uowasit.edu.iq

Received: 10 June 2024 | Revised: 27 June 2024 | Accepted: 30 June 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.8079*

ABSTRACT

Social networks have become increasingly influential in shaping political and governmental decisions in Middle Eastern countries and worldwide. Facebook is considered one of the most popular social media platforms in Iraq. Exploiting such a platform to assess the performance of institutions remains underutilized. This study proposes a model to help institutions, such as the Iraqi Ministry of Justice, evaluate their performance based on sentiment analysis on Facebook. Different machine learning algorithms were used, such as Support Vector Machine (SVM), Logistic Regression (LR), Extreme Gradient Boosting (XGBoost), Naive Bayes (NB), and Random Forest (RF). Extensive experimental analysis was performed using a large dataset extracted from Facebook pages belonging to the Iraqi Ministry of Justice. The results showed that SVM achieved the highest accuracy of 97.774% after retaining certain stop words, which proved to have a significant impact on the accuracy of the algorithms, ensuring the correct classification of comments while preserving the sentence's meaning.

Keywords-sentiment analysis; social media; facebook ;machine learning; TF-IDF

I. INTRODUCTION

Social networks have become an essential part of modern life, changing the way people connect, interact, and obtain information. It allows people to broadcast their thoughts, experiences, and interests to a global audience. The ability of users to create, share, and interact with content provides numerous benefits to individuals and businesses [1, 2]. Social media are one of the important parts of human life that has to do with social activities in the modern world and benefit decision-making, communication, information, interchange, and business marketing [3, 4]. It is one of the fundamental components in the digital age, providing data about customer likes, dislikes, etc. on various products and services. Companies and organizations have realized the importance of social media data in formulating effective marketing strategies, improving customer satisfaction, and building brand reputation [5]. Exploiting such data in various fields could be valuable to decision-makers. However, manually analyzing large amounts of social media data can be a challenging task that requires significant time, resources, and expertise.

Social media platforms have a measurable effect on shaping public discourse, particularly in issues related to governance and politics [6]. For institutions such as the Iraqi Ministry of Justice, these platforms offer valuable insight into the collective thoughts of the public. By actively following discussions on platforms such as Facebook, Twitter, and Instagram, the Ministry can stay informed about the public's views and concerns on legal and political issues in real time. By analyzing these data, organizations can tailor their policies according to public preferences, conduct surveys on legal issues, and provide legal advice. Social media act as a transparent channel of communication, promoting the interaction between citizens and the government and facilitating participation in decision-making processes. Facebook is one of the most popular social media platforms in Iraq [7]. More people have been involved in politics since the introduction of Facebook, as they receive news updates directly from political pages instead of using the Internet or television [8]. This study aimed to analyze the opinions of people on Facebook. Many studies have analyzed the contents of such platforms [8-16]. However, the Iraqi dialect for assessing Iraqi institutions remains largely untapped.

Machine learning has emerged as a powerful tool in various industries, revolutionizing the way data are processed, analyzed, and utilized [17]. By leveraging sophisticated algorithms and statistical models [18], machine learning enables systems to automatically learn from data, identify patterns, and make predictions or decisions with minimal human intervention [19]. In fields such as finance, healthcare [20, 21], marketing, and cyber security, machine learning techniques have been instrumental in extracting valuable insights from large datasets, improving efficiency, and driving innovation. Sentiment analysis is a specific application of machine learning and Natural Language Processing (NLP) that uses contextual mining to identify and extract subjective data from textual content, allowing businesses to understand social sentiment in their products or services [22]. Many people express their feelings online via social media channels. As a result, the data collected by these platforms can be used to analyze the attitudes expressed by users on various apps [23]. Using machine learning to assess and anticipate performance metrics, institutions can improve their efficiency, competitiveness, and long-term sustainability.

This study presents an approach for evaluating the performance of institutions, such as the Iraqi Ministry of Justice, using Facebook comments. The main contributions of this study are as follows:

- Develop a predictive model to assess the performance of institutions under the Iraqi Ministry of Justice.
- Collect and extract a public dataset.

II. LITERATURE REVIEW

Several studies have explored text classification in many types of text data, such as news, social media postings, and political publications, in different text languages. In [8], a system was presented to extract public opinions about the Iraqi government and politicians, analyzing comments on Facebook, employing three machine learning algorithms: Naïve Bayes (NB), K-Nearest Neighbor (KNN), and AdaBoost ensemble. Within the NB approach, two models were implemented: Bernoulli and multinomial models. The NB algorithm using the multinomial model achieved the highest accuracy (94%). In [9], the challenge of sentiment analysis in Arabic, due to its linguistic complexities, was addressed using a model that combined a CNN layer for local feature extraction with two LSTM layers to capture long-term dependencies. This new method achieved an impressive accuracy of 90.75% over a wide range of applications on different datasets. In [10], the challenge of analyzing sentiment in Arabic text, especially in noisy social media contexts such as Twitter, was addressed using a method based on the Discriminant Multinomial NB (DMNB) approach combined with N-gram tokenizer, stemming, and TF-IDF techniques. The DMNB classifier performed remarkably well, outperforming previous methods by achieving an accuracy improvement of 0.3%. A relevant study was also presented in [11], comparing three classifiers: Logistic Regression (LR), KNN, and Decision Tree (DT). The results showed that LR achieved the highest accuracy (93%), especially on large datasets, outperforming the other classifiers.

In [12], a method was proposed that involved the need for sentiment analysis on social media to better understand people's feelings and opinions on a variety of public topics, including women's rights and violence against women. Various traditional classification algorithms, such as KNN, Support Vector Machine (SVM), NB, and DT, were examined, with SVM yielding the most promising accuracy of 78.25%, while NB performed the poorest. In [13], the difficulty of analyzing Arabic tweets about COVID-19 for sentiment and topic classification was examined. A machine learning system that included sentiment analysis and topic categorization was presented. Sentiment analysis algorithms were developed and evaluated on a dataset obtained by KAUST, achieving a high F1-score of 97% with the NB classifier. For topic classification, an LSTM model was trained and tested using the AITD dataset, achieving an F1-score of 93%. In [14], it was stated that there is a lack of comprehensive comparative studies on the evaluation of time complexity in Arabic sentiment analysis using machine learning and deep learning models. This gap hinders the understanding of the computational efficiency of these algorithms for processing Arabic text. This study empirically determined the time complexity of seven popular machine learning algorithms to classify positive and negative Arabic sentences. This study collected Twitter data in Arabic, trained MLP, SVM, and LR models, and evaluated their time complexity to assess computational efficiency. SVM achieved the highest accuracy of 81%.

In [15], a comprehensive framework was presented to tackle the difficulty of interpreting emotions conveyed in Arabic tweets about the COVID-19 outbreak in Saudi Arabia, using two deep learning methods: Bi-directional Long-Short-Term Memory (BiLSTM) and Convolutional Neural Networks (CNN). The results showed that the CNN and BiLSTM models performed admirably, with 92.80% and 91.99% accuracy, respectively. In [16], a system was presented to analyze sentiments on social media during the global World Cup event, especially on platforms such as Twitter. This study aimed to measure the feelings toward the Qatar World Cup 2022 among Twitter users in Arab countries. A dataset was created, focusing on countries such as Egypt, Oman, Syria, Palestine, Algeria, Kuwait, Iraq, Sudan, Saudi Arabia, Jordan, Bahrain, Qatar, Yemen, and the United Arab Emirates. Using LR, Random Forest (RF), NB, and SVM, the results showed that LR performed best with 93% accuracy in sentiment analysis.

Despite the challenges of the Arabic language, it is important to recognize the benefits of methods to overcome obstacles. This study uses Facebook to assess the performance of government institutions in Iraq. Most studies have looked at text classification and sentiment analysis in different areas, such as news, social media posts, political publications, and COVID-19 sentiments. However, they have not focused much on using these methods to evaluate institutional performance.

III. SYSTEM MODEL

The approach followed to perform sentiment analysis on Iraqi Arabic datasets consists of five phases, as shown in Figure 1. The first step was to acquire data from Facebook. The dataset was subsequently cleaned and labeled. After that, data preprocessing and TF-IDF were applied to convert the text into vectors. Then, LR, RF, Extreme Gradient Boost (XGBoost), and NB were implemented to classify Facebook comments as positive or negative. Lastly, the performance of these models was evaluated using precision, accuracy, F-score, and recall.

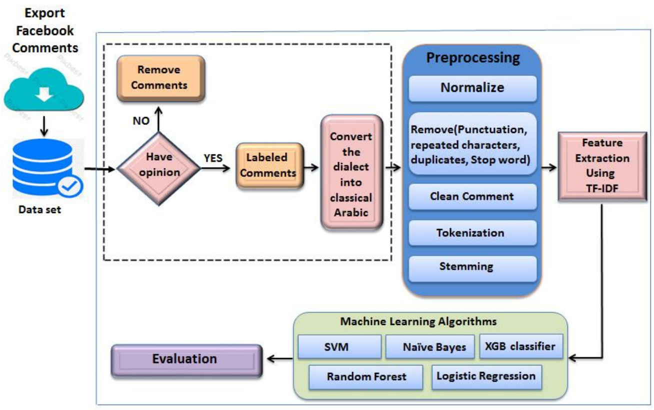

Fig. 1. The main steps of the proposed approach with its input and output.

A. *Data Pre-processing*

As this study deals with the noisy nature of the Arabic language on Facebook, the data need to be pre-processed to reduce the vast number of vocabulary words to be later used as features [24]. Text preprocessing is the most crucial step in any text classification system, as it can filter text by eliminating unwanted words and transforming words into appropriate representations. The steps involve pre-processing the dataset with the use of several libraries and tools, such as the Natural Language Toolkit (NLTK), which is a Python-based collection of applications and tools for statistical and symbolic NLP [25]. NLTK supports many tasks, such as classification, tagging, parsing, stemming, semantic reasoning, and tokenization [26]. Text preprocessing included the following operations:

- Normalization - Remove diacritics (Tashkeel). Diacritics are tiny markings applied to Arabic writing to represent vowels and other phonetic qualities. In text normalization, these diacritics were removed to simplify the text and reduce its complexity. Diacritics include symbols such as ◌ً ◌ٌ ◌ٍ ◌َ ◌ُ ◌ِ ◌ّ ◌ْ and others. These marks are often used in educational texts or poetry to aid pronunciation but are not essential for understanding the text.
- Restoring certain characters: Some characters in Arabic text may have variant forms, which are employed in specific circumstances or for historical reasons. For example:

  a) "آ,أ, إ" is typically replaced with "ا".

  b) "ئ,ؤ" is usually replaced with "ء".

  c) "ي" is typically replaced with."ى"

  These replacements were made to standardize the text and simplify processing, as the alternative forms are not commonly used in modern Arabic writing. These normalization steps ensure that the text is in a consistent and simplified form, which is more suitable for further processing and analysis, such as text classification tasks.

- Remove punctuation, repeated characters, duplicates, and stop words: In NLP, stopped words are meaningless words such as pronouns, prepositions, conjunctions, etc. These words were removed to decrease the amount of data and improve the efficiency of the algorithms used. However, it was observed that some stopped words affect the meaning of the sentence, and there was a difference in accuracy when deleting all stop words for Modern Standard Arabic, as accuracy increased when keeping stop words such as (لم, ليس, لا, غير). Repetitive letters in one word, such as (فاااااشل), were deleted to become (فاشل), but some words consist of two similar letters, so deleting the repetition of the letter such as (ممتازة , ممتاز , ممنون, ممنوع , سياسيين, نناشد , متكرر , متشدد ,متكرر, مميز ,etc.) was excluded. The punctuation marks were also deleted from the sentence because they do not express any analysis of feelings such as (#, @, $, %, &, |, etc.). Finally, the repetition of the word in one sentence was deleted to reduce the size of the data.
- Clean comment: The data cleaning step included deleting numbers and excess space in comments [27].
- Tokenization: This step divided the words in one sentence to benefit from them in the next step to compile the text.
- Stemming: This is a crucial pre-processing approach that improves the model's performance by removing words' affixes and reducing them to their most basic form. Stemming aims to alleviate the high dimensionality of text data by converting the words to their base forms. This reduction in word forms helps create a more manageable and less sparse dataset to improve the efficiency and effectiveness of text analysis [28]. Table I shows some examples.

TABLE I. EXAMPLES OF STEMMING COMMENTS

Comments	Stemming
قرار ممتاز ويقلل الازدحام والروتين	قرار ممتاز قل ازدحام روتين
جهود مباركة، وفقك الله وسدد قرارك العادل	جهد مبارك وفق الله قرار عادل
الرجل شريف ووطني والله يحفظه ويوفقه عمله	رجل شريف وطني الله حفظ وفق عمل

B. *Feature Extraction*

The Term Frequency-Inverse Document Frequency (TF-IDF) method uses two factors: the weight associated with the single word in the paragraph, used to determine its hesitation, and the inverse proportionality to the number of paragraphs in the set of paragraphs where the word appeared [30, 31]. The most popular technique, which is opposed to the other ways that skip this stage, counts the number of times the term appears in each document [29]. Since the dataset used is textual, it must be represented numerically to be fed into machine learning algorithms that will create the appropriate classifiers. The TF-IDF weighting system functions by assigning a low weight to phrases that are often found in the provided corpus. The inverse of the total number of times a certain phrase occurs in the corpus is called IDF. When a term is multiplied by TF, the outcome indicates how important it is to the specific document in question, representing how a phrase is document-specific [31]. The primary formula to determine the TF-IDF for each phrase in each document is [32]:

$$TF - IDF = TF \times IDF \quad (1)$$

TF counts the number of times a phrase appears in a document. It is always feasible for words to occur more frequently in longer texts than in shorter ones since documents vary in length.

$$TF = \frac{Frequency\ of\ term\ t\ in\ document\ d}{\text{Total number of terms in document d}} \quad (2)$$

The relevance of terms in a text is gauged using IDF. In determining TF, each term is given an equal weight. However, certain phrases, such as "when", "that is", and "at", despite being widespread, lack significant meaning.

$$IDF == \frac{Total\ number\ of\ documents}{Number\ of\ documents\ with\ term\ t\ in\ it} \quad (3)$$

C. Machine Learning Algorithms

1) Support Vector Machine (SVM)

The SVM is a linear model that can handle both regression and classification problems. SVM examines the data and finds the important information inside the input space. Two vectors, named classes, contain the imperative data. Every bit of data is expressed as a vector and categorized into classes. The border between the two classes is then established by the machine [26]. SVM creates a model that assigns new instances to one class based on a series of training examples classified into one of the two groups. Examples are represented as points in the feature space [33, 34].

2) Naive Bayes (NB)

NB is usually chosen for classification because of its simplicity and quickness. The NB classifier assumes that a feature's presence in a class is independent of the presence of any other feature. A fruit can be categorized as an apple, for instance, if it is spherical, yellow, and has a diameter of around three inches. The fact that this fruit is orange and gets its name "Naive" is due to all these characteristics, even if they are interdependent [35]. In mathematics, using the Bayes theorem, for a word w and class c:

$$P(c/w) = [P(w/c)P(c)]/P(w) \quad (4)$$

The probability of class c given word w is denoted by $P(c/w)$. $P(w)$ is the probability of the word w, and $P(c)$ is the probability of class c. Text classification situations are the main focus of the supervised learning technique known as the MNBC. This approach uses conditional probability and the multinomial distribution concept [36]. The probability calculation of the MNBC is shown in [37].

3) Extreme Gradient Boost (XGBoost) Classifier

XGBoost is the third classifier employed in this study [38, 39]. It is a scalable machine-learning method that has shown effectiveness in several data mining and machine-learning tasks. The XGBoost model serves as a sentiment analysis classifier to attain the best level of accuracy in emotion classification [26].

4) Random Forest (RF)

RF is a well-known ensemble learning technique and is widely used for both regression and classification tasks [40]. To provide a forecast that is more reliable and accurate, RF builds an ensemble of DTs or forests. This combination of tree predictors is based on the theory that the generalization error of the forest converges to a limit as the size of the forest trees increases [41].

5) Logistic Regression (LR)

LR is a predictive analytic method for classification issues, with its foundation being the idea of probability. LR uses the sigmoid function, which is a more intricate cost function. LR theory suggests that the cost function should be restricted to values between 0 and 1 [42].

IV. EVALUATION OF PREDICTIVE MODELS

A. Data Collection

Data were collected from the official Facebook page of the Iraqi Ministry of Justice using the comment extraction tool (Export Facebook Comments) [43]. About 7,716 comments were collected, and after cleaning the data from irrelevant comments, pictures, and stickers, 5,032 comments remained. Subsequently, psychology experts classified the comments into 3,000 positive comments and 2,032 negative comments. In the process of classifying comments, psychology experts utilized a systematic approach to discern between positive and negative sentiments expressed within them. The comments were meticulously analyzed, with positive ones often characterized by the presence of encouraging words, such as " "رائع", "ممتاز" موفق",, among others. Additionally, expressions of satisfaction, enthusiasm, or appreciation contribute to their classification. On the contrary, negative comments are identified through the usage of critical or disapproving language, such as " , "فاشل محبط"" , "سيء". Instances of dissatisfaction, frustration, or anger are key indicators of negativity. Finally, the comments were converted from the dialect to standard Arabic using Google Translate. Table II shows the dataset details.

TABLE II. DATASETS DETAILS

Dataset	Positive	Negative	Total
Facebook comments	3000	2032	5032

B. Experimental Setup

All experiments were carried out using Python on a PC with an Intel Core i7 3.4 GHz CPU and 16 GB of RAM running Windows 10. A 10-fold Cross-Validation (CV) was used to provide a balanced evaluation of the generalization error, randomly partitioning the total dataset into ten subsets, where nine were used for training (90%), and the remaining one was used for testing (10%). This technique was performed ten times, each time replacing the testing folds. CV is crucial in evaluating the reliability of a model and its ability to generalize to unseen data. It helps mitigate bias or random variance issues and provides a more accurate estimate of model performance across multiple tests [44, 45].

C. Performance Metrics

Standard metrics, such as precision, recall, F1-score, and accuracy were used to evaluate the performance of the prediction models [31, 46, 47].

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1-score = 2 * \frac{Precision*Recall}{Precision+Recall}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

V. RESULTS AND DISCUSSION

A. Performance Results by Removing All Stop Words

Table III shows the performance evaluation results using the five machine learning algorithms.

TABLE III. PERFORMANCE EVALUATION WHEN REMOVING ALL STOP WORDS

Facebook comments				
CV = 10				
Classifier	**Precision**	**Recall**	**F1-score**	**Accuracy**
SVM	97.603	97.576	97.58	97.576
NB	96.902	96.9	96.901	96.9
XGBoost	95.776	95.747	95.754	95.747
RF	96.441	96.443	96.441	96.443
LR	97.292	97.258	97.263	97.258

SVM emerged as the top-performing model with an accuracy of 97.58%. Additionally, precision was observed at 97.60%, with recall and F1-score also at 97.58%. SVM was closely followed by LR, with an accuracy of 97.26%, a precision of 97.29%, and a recall and F1-score of 97.26%. NB achieved an accuracy of 96.90%, with recall, F1-score, and precision all at 96.90%. This was slightly higher than for RF, which scored 96.44%. In RF, precision, recall, and F1-score were all 96.44%. XGBoost demonstrated the lowest accuracy among the models at 95.75%, precision of 95.78%, recall of 95.75%, and F1-score also of 95.75%. In summary, SVM was the leading model in this evaluation, followed by LR, NB, RF, and XGBoost. When it comes to the factors that should be taken into account when selecting a text classification technique, SVM is showing signs of outperforming other classifiers in text mining applications such as text categorization and text filtering [32].

B. Performance Results with Keeping Some Stop Words

Table IV presents the performance evaluation results achieved by the five machine learning algorithms.

TABLE IV. PERFORMANCE EVALUATION WITH KEEPING SOME STOP WORDS

Facebook comments				
CV = 10				
Classifier	**Precision**	**Recall**	**F1-score**	**Accuracy**
SVM	97.799	97.774	97.778	97.774
NB	97.042	97.039	97.040	97.039
XGBoost	96.209	96.184	96.190	96.184
RF	96.801	96.800	96.801	96.800
LR	97.452	97.417	97.422	97.417

After retaining some stop words, an increase in accuracy was observed across in five algorithms. Again, SVM emerged as the best-performing model, achieving an accuracy of 97.77%, with precision at 97.80%, recall at 97.77%, and F1-score at 97.78%. Following closely behind, LR achieved an accuracy of 97.42%, with precision at 97.45%, and recall and F1-score at 97.42%. NB achieved an accuracy of 97.04%, with recall, F1-score, and precision all at 97.04%, slightly surpassing RF, which scored 96.80%. For RF, precision was 96.80%, and recall and F1-score were 96.80%. XGBoost demonstrated the lowest accuracy of 96.18%, with precision at 96.21%, recall at 96.18%, and F1-score at 96.19%. In summary, SVM was the leading model in this evaluation, followed by LR, NB, RF, and XGBoost.

Analyzing sentiment in Arabic text has been a topic of interest in NLP research. Previous studies have focused on various methods to accurately classify sentiments, but one aspect that has been overlooked is the importance of retaining stop words. The results show that retaining some stop words improves accuracy and reduces the misclassification of sentiments. SVM was the highest-performing model in both sentiment analysis approaches, achieving the highest accuracy. As shown in Table IV, retaining some stop words increased the accuracy of the model and achieved better classification.

TABLE I. COMPARISON WITH PREVIOUS STUDIES

Ref.	**Dataset**	**Algorithm**	**Accuracy**
[8]	Facebook comments	Multinomial NB KNN	0.93544 0.80252
[9]	(from previously published works)	SVM NB Softmax KNN CNN–LSTM	90.75 88.75 87.43 86.83 90.75
[10]	Arabic tweets	DMNB	87.2%
[11]	Arabic tweets	DT KNN LR	74% 74% 93%
[12]	Arabic tweets	SVM KNN NB DT	78.25% 75.86% 71.07% 75.25%
[13]	Arabic tweets	NB LSTM NB + LSTM	0.97 0.93 91%
[14]	Arabic tweets	LR SVM DT RF KNN NB MLP	0.79 0.81 0.77 0.79 0.71 0.78 0.77
[15]	Arabic tweets	CNN BiLSTM	92.80 91.99
[16]	Arabic tweets	RF LR SVM NB	92 93 93 88
This study	Facebook comments	SVM Naïve Bayes XGB classifier Random Forest Logistic Regression	0.97774 0.97039 0.96184 0.96800 0.97417

The efficacy of the proposed model for Arabic sentiment analysis was contrasted with previous studies. Several criteria were used to choose the earlier studies. The datasets were sourced from Facebook, Twitter, and Instagram, the classifiers were machine learning or deep learning algorithms, and the datasets contained Arabic language content. Table V shows a discernible variance in the accuracy rates. In particular, the proposed model exhibited the highest accuracy, reaching 97.77% when using the SVM algorithm. For example, in [13], an accuracy of 97% was achieved using the NB algorithm on Twitter data comprising 55,000 instances. Similarly, in [8], 94% accuracy was achieved using the MNBC algorithm on 11,000 Facebook comments. In [16], 93% accuracy was achieved using LR and SVM on 464,124 tweets. Furthermore, a 93% accuracy rate was achieved in [11] when analyzing 66,666 tweets using LR. In [15], a CNN was used, achieving 92.8% accuracy on 90,187 Twitter messages. In [9], 90.75% accuracy was achieved using SVM. Interestingly, the use of a CNN in the same study produced an identical accuracy rate. In [10], the DMNB method achieved 87.2% accuracy with a small sample of 2,000 Twitter messages. In [14], 81% accuracy was achieved with SVM on a dataset of 85,751 tweets. Finally, in [12], 78.2% accuracy was achieved using SVM on a dataset consisting of 3,700 tweets. This analysis emphasizes the gradual improvements and nuances in accuracy rates when using different algorithms and dataset sizes.

VI. CONCLUSION AND FUTURE WORK

This study presented a method for evaluating institutional performance, focusing on the Iraqi Ministry of Justice, using sentiment analysis of Facebook comments. It conducted a comprehensive experimental study to investigate the performance of several machine learning algorithms to properly recognize feelings conveyed in the Arabic language. Notably, the results showed that preserving some stop words improved model accuracy and decreased sentiment misclassification. Among these methods, the best-performing model was SVM, achieving an accuracy of 97.774%, a precision of 97.80%, a recall of 97.77%, and an F1-score of 97.78%. NB, LR, and RF also performed well, but XGBoost had the lowest accuracy.

Future research should investigate modifications to the models and their application to different institutions and languages. Furthermore, increasing the number of comments and using more deep-learning approaches might increase the model's resilience and generalizability. Moreover, adding a third classification category for neutral sentiments in addition to positive and negative sentiments could provide a more nuanced understanding of public perception. Finally, it would be interesting to examine other criteria, such as interpretability.

REFERENCES

[1] I. M. Tarigan, M. A. K. Harahap, D. M. Sari, R. D. Sakinah, and A. M. A. Ausat, "Understanding Social Media: Benefits of Social Media for Individuals," *Jurnal Pendidikan Tambusai*, vol. 7, no. 1, pp. 2317–2322, Feb. 2023, https://doi.org/10.31004/jptam.v7i1.5559.

[2] A. Yohanna, "The influence of social media on social interactions among students," *Indonesian Journal of Social Sciences*, vol. 12, no. 2, pp. 34–48, 2020.

[3] G. Appel, L. Grewal, R. Hadi, and A. T. Stephen, "The future of social media in marketing," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 79–95, Jan. 2020, https://doi.org/10.1007/s11747-019-00695-1.

[4] Noureen, S. H. H. Huspi, and Z. Ali, "Sentiment Analysis on Roman Urdu Students' Feedback Using Enhanced Word Embedding Technique," *Baghdad Science Journal*, vol. 21, no. 2(SI), Feb. 2024, https://doi.org/10.21123/bsj.2024.9822.

[5] K. F. Ferine, S. S. Gadzali, A. M. A. Ausat, M. Marleni, and D. M. Sari, "The Impact of Social Media on Consumer Behavior," *Community Development Journal : Jurnal Pengabdian Masyarakat*, vol. 4, no. 1, pp. 843–847, Mar. 2023, https://doi.org/10.31004/cdj.v4i1.12567.

[6] A. M. A. Ausat, "The Role of Social Media in Shaping Public Opinion and Its Influence on Economic Decisions," *Technology and Society Perspectives (TACIT)*, vol. 1, no. 1, pp. 35–44, Aug. 2023, https://doi.org/10.61100/tacit.v1i1.37.

[7] "Top Social Media Networks Websites Ranking in Iraq in May 2024," *Similarweb*. https://www.similarweb.com/top-websites/iraq/computers-electronics-and-technology/social-networks-and-online-communities/.

[8] L. A. Habeeb, "Sentiment Analysis for Iraqis Dialect in Social Media," *Iraqi Journal of Information and Communication Technology*, vol. 1, no. 2, pp. 24–32, Jul. 2018, https://doi.org/10.31987/ijict.1.2.17.

[9] A. H. Ombabi, W. Ouarda, and A. M. Alimi, "Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks," *Social Network Analysis and Mining*, vol. 10, no. 1, Jul. 2020, Art. no. 53, https://doi.org/10.1007/s13278-020-00668-1.

[10] H. AlSalman, "An Improved Approach for Sentiment Analysis of Arabic Tweets in Twitter Social Media," in *2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, Mar. 2020, https://doi.org/10.1109/ICCAIS48893.2020.9096850.

[11] N. K. Bolbol and A. Y. Maghari, "Sentiment Analysis of Arabic Tweets Using Supervised Machine Learning," in *2020 International Conference on Promising Electronic Technologies (ICPET)*, Jerusalem, Palestine, Dec. 2020, pp. 89–93, https://doi.org/10.1109/ICPET51420.2020.00025.

[12] M. Alzyout, E. Al Bashabsheh, H. Najadat, and A. Alaiad, "Sentiment Analysis of Arabic Tweets about Violence Against Women using Machine Learning," in *2021 12th International Conference on Information and Communication Systems (ICICS)*, Valencia, Spain, May 2021, pp. 171–176, https://doi.org/10.1109/ICICS52457.2021.9464600.

[13] F. Alderazi, A. A. Algosaibi, and M. A. Alabdullatif, "The Use of Arabic Language COVID-19 Tweets Analysis in IoT Applications," in *2021 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, Dubai, United Arab Emirates, Dec. 2021, pp. 112–117, https://doi.org/10.1109/GCAIoT53516.2021.9693080.

[14] R. Kharsa and S. Harous, "Machine Learning Classification Algorithms for Sentiment Analysis in Arabic: Performance Evaluation and Comparison," in *2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, Ras Al Khaimah, United Arab Emirates, Nov. 2022, pp. 395–400, https://doi.org/10.1109/ICECTA57148.2022.9990108.

[15] A. Alqarni and A. Rahman, "Arabic Tweets-Based Sentiment Analysis to Investigate the Impact of COVID-19 in KSA: A Deep Learning Approach," *Big Data and Cognitive Computing*, vol. 7, no. 1, Mar. 2023, Art. no. 16, https://doi.org/10.3390/bdcc7010016.

[16] M. Faisal, Z. Abouelhassan, F. Alotaibi, R. Alsaeedi, F. Alazmi, and S. Alkanadari, "Sentiment Analysis Using Machine Learning Model for Qatar World Cup 2022 among Different Arabic Countries Using Twitter API," in *2023 IEEE World AI IoT Congress (AIIoT)*, Seattle, WA, USA, Jun. 2023, pp. 222–228, https://doi.org/10.1109/AIIoT58121.2023.10188463.

[17] A. S. Abdalrada, J. Abawajy, T. Al-Quraishi, and S. M. S. Islam, "Prediction of cardiac autonomic neuropathy using a machine learning model in patients with diabetes," *Therapeutic Advances in Endocrinology and Metabolism*, vol. 13, Jan. 2022, Art. no. 20420188221086693, https://doi.org/10.1177/20420188221086693.

[18] A. S. Abdalrada, J. H. Abawajy, M. U. Chowdhury, S. Rajasegarar, T. Al-Quraishi, and H. F. Jelinek, "Relationship Between Angiotensin Converting Enzyme Gene and Cardiac Autonomic Neuropathy Among Australian Population," in *Recent Advances on Soft Computing and Data Mining*, 2018, pp. 135–146, https://doi.org/10.1007/978-3-319-72550-5_14.

[19] A. S. Abdalrada, J. Abawajy, M. Chowdhury, S. Rajasegarar, T. Al-Quraishi, and H. F. Jelinek, "Meta learning ensemble technique for diagnosis of cardiac autonomic neuropathy based on heart rate variability features," in *30th International Conference on Computer Applications in Industry and Engineering, CAINE 2017*, 2017, pp. 169–175.

[20] T. Al-Quraishi, J. H. Abawajy, N. Al-Quraishi, A. Abdalrada, and L. Al-Omairi, "Predicting Breast Cancer Risk Using Subset of Genes," in *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, Paris, France, Apr. 2019, pp. 1379–1384, https://doi.org/10.1109/CoDIT.2019.8820378.

[21] A. Abdalrada, Ali Fahem Neamah, and Hayder Murad, "Predicting Diabetes Disease Occurrence Using Logistic Regression: An Early Detection Approach," *Iraqi Journal For Computer Science and Mathematics*, vol. 5, no. 1, pp. 160–167, Jan. 2024, https://doi.org/10.52866/ijcsm.2024.05.01.011.

[22] P. Karthika, R. Murugeswari, and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," in *2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, Tamilnadu, India, Apr. 2019, https://doi.org/10.1109/INCOS45849.2019.8951367.

[23] H. A. Alatabi and A. R. Abbas, "Sentiment Analysis in Social Media using Machine Learning Techniques," *Iraqi Journal of Science*, pp. 193–201, Jan. 2020, https://doi.org/10.24996/ijs.2020.61.1.22.

[24] H. Alamoudi *et al.*, "Arabic Sentiment Analysis for Student Evaluation using Machine Learning and the AraBERT Transformer," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11945–11952, Oct. 2023, https://doi.org/10.48084/etasr.6347.

[25] D. Elangovan and V. Subedha, "Adaptive Particle Grey Wolf Optimizer with Deep Learning-based Sentiment Analysis on Online Product Reviews," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10989–10993, Jun. 2023, https://doi.org/10.48084/etasr.5787.

[26] M. Nandan, S. Chatterjee, A. Parai, and O. Bagchi, "Sentiment Analysis of Twitter Classification by Applying Hybrid-Based Techniques," in *Proceedings of the 3rd International Conference on Communication, Devices and Computing*, 2022, pp. 591–606, https://doi.org/10.1007/978-981-16-9154-6_55.

[27] M. A. Kausar, S. O. Fageeri, and A. Soosaimanickam, "Sentiment Classification based on Machine Learning Approaches in Amazon Product Reviews," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10849–10855, Jun. 2023, https://doi.org/10.48084/etasr.5854.

[28] E. Refaee, "Sentiment Analysis for Micro-blogging Platforms in Arabic," in *Social Computing and Social Media. Applications and Analytics*, Vancouver, Canada, 2017, pp. 275–294, https://doi.org/10.1007/978-3-319-58562-8_22.

[29] Z. Lu, "Web Page Classification Using Features from Titles and Snippets," M.S. Thesis, University of Ottawa, 2015.

[30] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*, 2003, vol. 242, no. 1, pp. 29–48.

[31] S. A. Aljuhani and N. Alghamdi, "A comparison of sentiment analysis methods on Amazon reviews of Mobile Phones," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 608–617, 2019, https://doi.org/10.14569/ijacsa.2019.0100678.

[32] H. Syahputra and A. Wibowo, "Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 1, pp. 165–173, 2023.

[33] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, Pittsburgh, PA, USA, Apr. 1992, pp. 144–152, https://doi.org/10.1145/130385.130401.

[34] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, https://doi.org/10.1007/BF00994018.

[35] A. Goel, J. Gautam, and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, Dehradun, India, Oct. 2016, pp. 257–261, https://doi.org/10.1109/NGCT.2016.7877424.

[36] A. A. Farisi, Y. Sibaroni, and S. A. Faraby, "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier," *Journal of Physics: Conference Series*, vol. 1192, no. 1, Nov. 2019, Art. no. 012024, https://doi.org/10.1088/1742-6596/1192/1/012024.

[37] K. Dhola and M. Saradva, "A Comparative Evaluation of Traditional Machine Learning and Deep Learning Classification Techniques for Sentiment Analysis," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, Jan. 2021, pp. 932–936, https://doi.org/10.1109/Confluence51648.2021.9377070.

[38] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, May 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[39] B. Gaye and A. Wulamu, "Sentimental Analysis for Online Reviews using Machine Learning Algorithms," *International Research Journal of Engineering and Technology*, vol. 6, no. 8, pp. 1270–1275, 2019.

[40] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, https://doi.org/10.1023/A:1010933404324.

[41] R. S. Utsha, M. Keya, A. Hasan, and S. Islam, "Qword at CheckThat! 2021: An Extreme Gradient Boosting Approach for Multiclass Fake News Detection," in *CEUR Workshop Proceedings*, Bucharest, Romania.

[42] A. Poornima and K. S. Priya, "A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, Mar. 2020, pp. 493–496, https://doi.org/10.1109/ICACCS48705.2020.9074312.

[43] "Export Facebook, Instagram, Twitter, YouTube, TikTok, Vimeo Comments," *exportcomments.com*. https://exportcomments.com/.

[44] S. Dutt, S. Chandramouli, and A. K. Das, *Machine Learning*, 1st ed. Pearson Education, 2018.

[45] H. Smolic, "The Importance of Cross-Validation in Machine Learning," Feb. 19, 2024. https://graphite-note.com/the-importance-of-cross-validation-in-machine-learning/.

[46] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Computer Science*, vol. 57, pp. 821–829, Jan. 2015, https://doi.org/10.1016/j.procs.2015.07.523.

[47] A. S. Al-Jumaili, "A Hybrid Method of Linguistic and Statistical Features for Arabic Sentiment Analysis," *Baghdad Science Journal*, vol. 17, no. 1(Suppl.), pp. 385–390, Mar. 2020, https://doi.org/10.21123/bsj.2020.17.1(Suppl.).0385.