Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM
e-ISSN: 2788-5879 p-ISSN: 2788-5887

Assessing Institutional Performance Using Machine Learning Algorithms

Zainab Alwan Anwer [1]*, Ahmad Shaker Abdalrada [2] Ihtiram Raza Khan [3]

[1] Department of Software, Faculty of Computer Science and Information Technology, Wasit University, Iraq
[2] Department of Software, Faculty of Computer Science and Information Technology, Wasit University, Iraq
[3] Department of Computer Science & Engineering, School of Engineering Sciences & Technology Jamia Hamdard Delhi, India

*Corresponding Author: Zainab Alwan Anwer

DOI: https://doi.org/10.31185/wjcms.263
Received 20 July 2024; Accepted 19 Augest 2024; Available online 30 September 2024

ABSTRACT: In Middle Eastern nations, social media has grown in importance in influencing political and governmental choices. In Iraq, Facebook is regarded as one of the most widely used social networking sites. The underutilization of this tool for evaluating institutional performance persists. Thus, using sentiment analysis on Facebook, this study suggests a methodology that aids organizations like the Ministry of Justice in Iraq in assessing their own performance. The model makes use of a variety of machine learning methods, including Naive Bayes, Logistic Regression, and Support Vector Machine. TF-IDF (Term Frequency-Inverse Document Frequency) was used to convert the textual data into numerical features, which is essential for effective text analysis. Additionally, features were carefully managed by utilizing both unigram and bigram models. Using datasets from (Facebook pages belonging to the Iraqi Ministry of Justice), a thorough experimental investigation was conducted. The results of our experiments showed that the SVM algorithm produced the best accuracy, at 98.311%. following the suggested model's retention of a few stop words, which was shown to significantly improve the algorithm's performance and guarantee accurate categorization of comments while maintaining the content of the phrase.

Keywords: Sentiment analysis; Social media; Machine Learning Algorithms ; TF-IDF

1. INTRODUCTION

Social media has become an integral part of modern life, fundamentally reshaping how individuals connect, communicate, and consume information. Through platforms such as Facebook, Twitter, Instagram, and LinkedIn, people can effortlessly share their thoughts, experiences, and interests with a global audience [1]. One of the most significant aspects of human life is social media, which is used for social activities in the modern world. It has developed to the point where some of us have benefited from it even in our earlier years in terms of decision-making, communication, information exchange, marketing, and business marketing, particularly in the business sector [2, 3].

In the digital age, it has become an essential component of everyday life, offering insights about consumer preferences, attitudes, and views about a range of goods and services. As a result, businesses and organizations now understand how crucial social media data is to developing successful marketing plans, raising consumer satisfaction levels, and enhancing brand recognition [4]. It gives an importance in every community.

The 2016 US presidential election serves as an example of this, since social media platforms like Facebook, Twitter, and YouTube significantly influenced popular opinion and voting behavior [5]. It results in more accountability and openness. Social media allows citizens to keep track on government acts and hold elected officials accountable in real time [6]. Protest movements have benefited greatly from social media organization and mobilization; the Arab Spring is one well-known example. Social media sites like Facebook and Twitter were used to disseminate information and plan protests, which had a significant impact on political changes in a number of nations [7].

*Corresponding author: zainab.alwan@uowasit.edu.iq

Social media campaigns on certain issues may force governments to act. To counteract sexual assault and harassment, for instance, the #MeToo movement altered public opinions and spurred laws and government initiatives [8]. Real-time feedback helps with the required adjustments to meet public concerns and makes it easier to grasp the repercussions of policies. Social media platforms, for instance, provide governments with the opportunity to interact with information and opinion as it becomes available (in real-time) during emergencies, including natural disasters or public health crises. This can improve the efficiency and promptness of government responses [9].

For organizations such as the Iraqi Ministry of Justice, these platforms provide a unique perspective about how people are feeling as a collective. Social media acts as a transparent channel of communication, promoting interaction between citizens and the government and facilitating participation in decision-making processes. In Iraq, Facebook is one of the most widely used social networking platforms [10]. Thus, the results of this study will be based on what public civils have to say on Facebook. Numerous studies have been conducted in the literature [11-16] to analyze the contents of such platforms, but the use of the Iraqi dialect language to assess Iraqi institutions is yet completely unexplored.

Machine learning has become a potent instrument in many different fields, transforming the handling, interpretation, and use of data [17]. By the use of complex statistical models and algorithms [18]. It is allows systems to learn from data automatically, see patterns, and make decisions or predictions with little help from humans [19]. Machine learning approaches have been essential in gaining new insights from enormous datasets, improving efficiency, and spurring innovation in several industries, including finance, healthcare [20, 21] marketing, and cyber security. Sentiment analysis is a specialized usage of natural language processing and machine learning that leverages contextual mining to find and extract subjective information from textual content, This helps companies to comprehend how consumers feel about their offerings [22]. Everyone uses social media platforms to share their emotions online. Therefore, the information gathered by these platforms might be used to analyze the opinions users have voiced on different apps [23]. Institutions may improve efficiency, competitiveness, and long-term sustainability by using machine learning to evaluate and predict performance measures [24, 25].

In This paper, develops a seamless model that may be used to assess an organization's performance, such the Ministry of Justice in Iraq, utilizing comments from Facebook. The following are our principal contributions:

- Create a prediction model to evaluate how well the Iraqi Ministry of Justice's institutions are performing.
- Gather and take a public dataset.

This paper outline was organized as follows: Section 2 gave a summary of the relevant studies that have already been done. In Section 3, the recommended sentiment analysis prediction model was developed further. Section 4 included an overview of the proposed model and an assessment of its functionality. Section 5 presented the results, while Section 6 included a conclusion and recommendations for further study.

2. RELATED WORKS

Many research on sentiment analysis have been conducted in a variety of text data formats, including news articles, political publications, social media posts, and texts written in Arabic and English. Papers on sentiment analysis and related fields will be discussed.

The researcher Bolbol and Maghari [11], who investigated the difficulty of interpreting Arabic language data on social media networks. Arabic text mining is a relatively untapped field because there aren't many research in this area. The study uses machine learning methods for data categorization in order to address this problem. In particular, a comparison is conducted between three classifiers: Decision Tree (DT), K-Nearest Neighbors (KNN), and Logistic Regression (LR). According to the results, logistic regression outperforms the other classifiers by 93% and has the best accuracy rate, particularly when used to big datasets.

Alzyout, et al. [12] addressed sentiment analysis and offered a solution that centred on the necessity of sentiment analysis on social media in order to better understand people's thoughts and sentiments regarding a range of public issues, such as violence against women and women's rights. Examined are a number of conventional classification techniques, including Decision Trees (DT), Naive Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). According to the study, Naive Bayes produces the worst results, while the Support Vector Machine method produces the most promising accuracy of 78.25%.

Also, Alderazi, et al. [13] to address the challenge of classifying topics and assessing mood in Arabic tweets around COVID-19. The researchers proposed a machine learning method that combined sentiment analysis and subject classification as two models to address this. Using a dataset provided by KAUST, they created and assessed sentiment analysis algorithms, and the Naïve Bayes classifier yielded a high F1-score of 97%. Using the AITD dataset, an LSTM model was trained and tested for topic categorization, and it obtained an F1-score of 93%.

Moreover, Kharsa and Harous [14] explained the dearth of thorough comparison research on the evaluation of temporal complexity in Arabic sentiment analysis with ML and DL models. Understanding the computational effectiveness of these methods for processing Arabic text is hampered by this gap. The temporal complexity of seven well-known machine learning methods for identifying positive and negative Arabic phrases is empirically determined in this work to close the gap. To gauge computing efficiency, the researchers trained models, gathered Arabic-language Twitter data, and evaluated time complexity. The MLP, SVM, and LR algorithms are the ones employed in this study. SVM had the greatest accuracy, coming in at 81%.

Another authors Alqarni and Rahman [15], achieved noteworthy progress by releasing an extensive framework for this study makes an effort to address the challenge of deciphering the emotions expressed in Arabic tweets on the COVID-19 pandemic in Saudi Arabia. The study employed convolutional neural networks (CNN) and bi-directional long short-term memory (BiLSTM) as two deep learning techniques. The results showed that the CNN and BiLSTM models worked quite well. The accuracy of CNN was 92.80%, whereas the accuracy of BiLSTM was 91.99%.

Finally, Faisal, et al. [16], proposed a new method for assessing mood on social media during the World Cup, particularly on sites like Twitter. This survey attempts to gauge how Arab Twitter users feel about the 2022 World Cup in Qatar. A targeted dataset was assembled by the researchers from a number of nations, including Saudi Arabia, Jordan, Bahrain, Qatar, Yemen, Algeria, Egypt, Oman, Syria, Palestine, Algeria, Kuwait, Iraq, and Sudan. Results show that Logistic Regression outperformed Random Forest, Naive Bayes, SVM machine learning algorithms, with 93% accuracy in sentiment analysis.

Despite these difficulties, it's critical to acknowledge the advantages of electronic technologies and make efforts to remove any barriers to their adoption. The research's listed above are restricted to using Facebook to evaluate the effectiveness of government institutions in Middle Eastern nations such as Iraq. The majority of research have examined text categorization and sentiment analysis across several domains, including news, political publications, social media posts, and COVID-19 attitudes. They haven't, however, given much thought to applying these techniques to assess institutional effectiveness.

3. SYSTEM MODEL

In this paper, the system model utilized in this work has five steps, as illustrated in **Figure 1**, for sentiment analysis on dialect Iraqi Arabic datasets. Facebook data collection was the first step. Following this, the datasets were cleaned and labelled. Subsequently, the text was converted into vectors using TF-IDF and data pre-processing. Afterwards, to categorize the Facebook comments as positive or negative, the machine learning techniques Logistic Regression, Support Vector Machine, and Naive Bayes were used. The experiment's success was then assessed using a number of metrics, including recall, F-score, accuracy, and precision. We will go into great detail about each step.

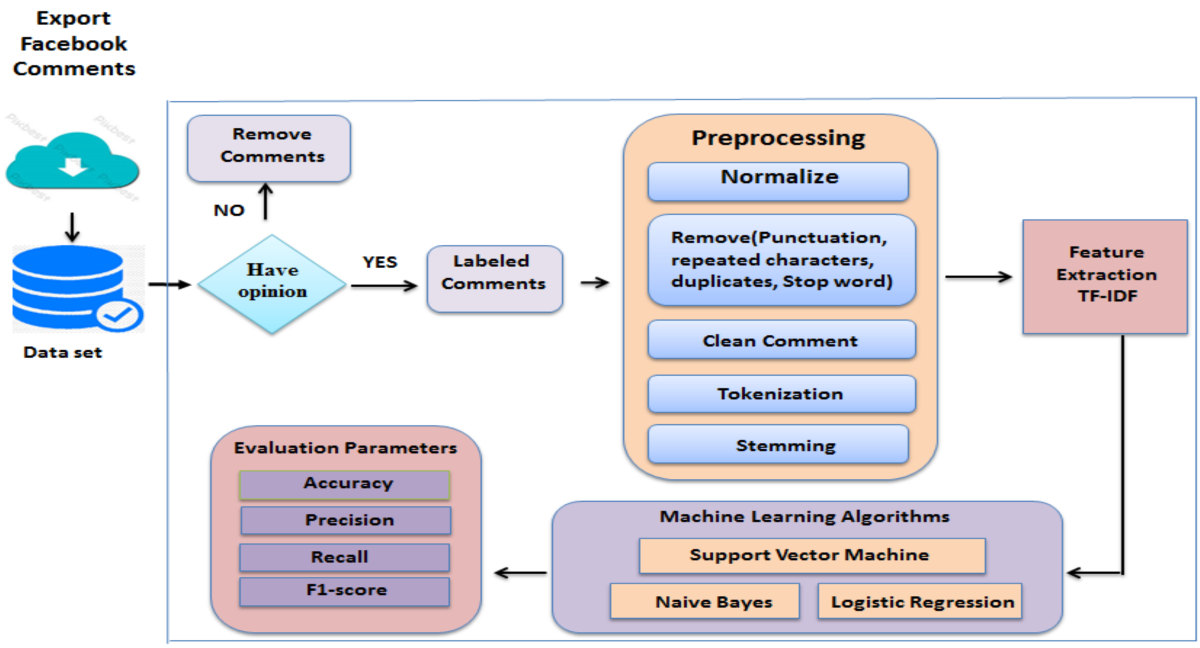

FIGURE 1. - System Model

Each step of the recommended process was explained in the next subsections.

3.1 Data Pre-processing

The complex morphology, script, and orthographic variants of Arabic provide a variety of unique challenges for Arabic text preparation[26]. Natural language processing (NLP) preprocessing is a crucial step that prepares text for further analysis or machine learning activities. When properly preprocessed, NLP models can perform substantially better [27]. Preprocessing entails several steps, including the following:

3.1.1 Data Cleaning

Data cleaning is the practice of removing data that doesn't enhance a dataset's overall quality and duplicate entries. Erroneous, insufficient, or unnecessary data are frequently removed from the dataset [28].

- Duplicates.
- Eliminate URLs.
- Repeated Characters.
- Remove Punctuation.
- Remove Numbers.
- Eliminating non-Arabic characters.
- Remove Whitespaces.

3.1.2 Stopword Removal

Stopped words in natural language processing are useless words such as conjunctions, pronouns, and prepositions, which are eliminated to reduce data volume and boost algorithm performance [27]. But it was noted that some stopped words changed the meaning of the phrase, and that accuracy differed when all stop words were removed from Arabic sentences; accuracy increase when stop words such (لم, ليس, لا, غير) were retained.

3.1.3 Normalization

- Removing Diacritics: Arabic script is phonetically structured, with diacritical symbols designating short vowels. They are often removed in order to reduce sparsity in the text data.
- Standardizing Characters: Various versions of the letter "Alef" (اً, أ, إ, آ) are normalized to become a single form (ا) in order to assure uniformity.

3.1.4 Tokenization

The process of breaking a sentence up into its component words, phrases, characters, and symbols is known as tokenization [29].

3.1.5 Stemming

Text pre-processing requires the crucial step of stemming, which is the reduction of words to their original root [30]. By reducing words to their basic forms, stemming attempts to reduce the high dimensionality of text data. This makes the dataset easier to handle and less sparse, which enhances the efficacy and efficiency of text analysis [27].

3.2 Feature Extraction

For machine learning algorithms to work directly with text, it must first be converted into sets of numbers (vectors); this process is referred to as vector representation or feature extraction in machine learning. A group of techniques used in natural language processing (NLP) for language modeling and representation training are referred to as vector representations [31]. The TF-IDF approach uses two factors: the weight of the single word in the paragraph, which is used to calculate its hesitation; and, in another sense, the inverse proportionality to the number of paragraphs in the collection of paragraphs where the word surfaced. In contrast to alternative approaches that omit this step, the most often used method counts the instances of the word in each document inside the collection of documents [32]. Phrases that are often present in the supplied corpus are given a low weight by the TF-IDF weighting mechanism. Inverse document frequency, or IDF, is the reciprocal of the total number of times a certain term appears in the corpus. The product of a word's term frequency (TF) and its importance should be expressed in terms of the particular text in question. It illustrates how a term is unique to a certain document [33]. **Equation (1)** provides the main formula for calculating the TF-IDF for each phrase in each text [34].

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

When a term appears in a document, TF keeps track of how often it occurs. Given the variance in length of documents, it is always possible for terms to appear more frequently in longer texts than shorter ones.

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

IDF is used to determine a term's relevancy inside a text. Every phrase has the same weight when calculating TF. However, some expressions, such "when," "that is," and "at," It is widely used, yet it has no real significance.

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

3.3 Machine Learning Algorithms

The performance of the following classifiers has been examined in this subsection.

3.3.1 Support Vector Machine (SVM):

Support Vector Machine (SVM) is an optimization-based learning method that uses high-dimensional data to create imaginary spaces that are represented as linear functions [35, 36]. SVM is frequently used to regression and binary classification issues [37]. SVM is also a family of powerful classifiers that have shown efficacy in a range of tasks related to natural language processing [34]. Support Vector Machine (SVM) technology builds a model based on a set of training examples categorized into one of the two groups, which then allocates fresh instances to one class. In the SVM models feature space, examples are represented as points. Learn more at the source [38, 39].

3.3.2 Naive Bayes:

An approach for supervised classification is the naive Bayes algorithm. Naive Bayes is the most popular choice for classification because to its speed and ease of use. The assumption made by the naive Bayes classifier is that a feature's existence in a class is unrelated to the existence of any other feature. For example, a fruit that is round, yellow, and about three inches in diameter may be called an apple. Even though some of these traits are interrelated, all of these factors contribute to the fruit's orange color and the moniker "Naive" [41]. In the field of mathematics, the Bayes theorem is used for a word w and class c.

$$\mathrm{P(c/w) = [P(w/c)P(c)]/P(w)} \qquad (4)$$

P(c/w) is the probability of class c given word w. P(w) represents the likelihood of the word w, whereas P(c) represents the likelihood of the class.

Probabilistic supervised learning techniques like the Multinomial Naïve Bayes Classifier are mostly focused on text classification scenarios. This method makes use of the multinomial distribution notion and conditional probability [42]. The multinomial Naive Bayes classifier's probability computation in reference [43].

3.3.3 Logistic Regression

Logistic Regression is a classification approach for machine learning that forecasts the likelihood of a categorical dependent variable. The dependent variable in logistic regression is a binary variable with data that is either recorded as 0 (NO) or 1 (YES) [44]. It is a predictive analytical technique used to address classification problems. Probability is the fundamental concept of it. In logistic regression, a more complex cost function called the "sigmoid function" is employed [45].

4. THE EVALUATION OF THE PREDICTIVE MODEL

The suggested prediction model underwent evaluation and testing. The measurements, experimental setup, and dataset used are all covered in this section.

4.1 Data collection

Using the comment extraction tool (Export Facebook Comments), data was gathered from the Iraqi Ministry of Justice's official Facebook page. After removing unnecessary remarks, images, and stickers from the data, 5,032 comments were left out of the approximately 7,716,000 that were initially gathered. Experts in psychology have divided the comments into 3,000 positive and 2,032 negative categories. Psychology professionals use a methodical strategy to classify remarks in order to differentiate between positive and negative comments. Positive comments are carefully examined; they are frequently identified by the use of terms that are uplifting, such "ممتاز ,"رائع"," موفق", and so on. Expressions of contentment, zeal, or gratitude also play a role in categorizing them. On the other hand, disparaging remarks can be distinguished by the use of phrases like "محبط" ",فاشل," and "سيء." Examples of wrath, annoyance, or discontent are important markers of negative. The dataset is preprocessed before being analyzed using machine learning models for the comments. **Table 1** displays the details of the dataset.

Table 1. - Datasets Details

Dataset	Positive	Negative	Total
Facebook Comments	3000	2032	5032

4.2 Experimental Setup

The Python programming language was used for all of the tests on a Windows 10 PC with an Intel Core 7 processor, 3.4 GB of RAM, and 16 GB of CPU. A number of metrics were developed to evaluate the effectiveness of the proposed model, including confusion matrix evaluations, accuracy, precision, recall, and F1-score. The assessment tools used in this work are described as follows: A fair assessment of the generalization error was obtained through the use of 10-fold cross-validation (CV). The whole dataset was randomly divided into ten sub-sets using 10-fold cross-validation (CV). Nine of these sub-sets were utilized for training, while the remaining sub-set was used for testing (10%). The tested folds are then substituted each time through 10 iterations of this approach.

4.3 Performance Metrics

Standard metrics, such as precision, recall, F1-score, and accuracy, were computed to assess the prediction model's performance [33, 46].

The precision is determined by dividing the fraction of positively categorized reviews that really forecast by the entire number of favorably categorized reviews, as given by **Equation 5** [47].

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

The percentage of TP value with the total number of TP and FN is known as recall.

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

Equation's F1-score incorporates recall and accuracy.

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{7}$$

Finally, accuracy is defined by **Equation 8** as the proportion of reviews correctly classified divided by the total number of reviews [48]. where TN, FN, TP, and FP stand for true negatives, false negatives, true positives, and false positives, respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

5. RESULTS AND DISCUSSION

The experimental findings and discussions were provided in this section.

5.1 Model Performance Results using three algorithm by remove all stop word (Max_df = 0.5 , Min_df = 5, Max_features = 5000, N-grams(1,2) (Unigram and Bigram)):

Table 2 displayed the results of the performance evaluation utilizing the three machine learning methods. The accuracy of the SVM algorithm was a higher.

By removing uncommon and extremely common words, condensing the amount of features to the most crucial ones, and utilizing bigrams and unigrams to capture more context, TF-IDF parameters assist enhance text analysis.

- **Max_df = 0.5**

Terms that are present in more than half of the documents will be eliminated in order to eliminate overly frequent terms that don't contribute to the analysis.

- **Min_df = 5**

Exclusion of uncommon terms that could not be helpful in statistical analysis occurs when a word appears in at least five texts.

- **Max_features = 5000**

Restricts the number of features to 5000 and uses the TF-IDF scores to determine which words are most relevant in order to minimize the amount of the data while preserving crucial characteristics.

- **N-grams (1,2) (Unigram and Bigram)**

Takes into account both single words (unigrams) and word pairs (bigrams) in order to extract additional contextual information from the texts.

Table 2. - Performance evaluation for three machine learning algorithms (Experiment 1).

Facebook comments				
Cross Validate , cv=10				
Classifier	***Precision***	***Recall***	***F1-score***	***Accuracy***
SVM	**0.98141**	**0.98132**	**0.98134**	**0.98132**
Naive Bayes	0.97915	0.97913	0.97914	0.97913
Logistic Regression	0.97452	0.97436	0.97440	0.97436

The best-performing model was Support Vector Machine (SVM), which had an accuracy of 98.132%. Recall was at 98.132, and F1 score was recorded at 98.134%, while precision was reported at 98.141%. Naive Bayes came in second place with accuracy of 97.913%, precision of 97.915%, recall of 97.913%, and F1 score of 97.914%. The results of the logistic regression showed that the accuracy was 97.436%, recall was 97.436%, F1 score was 97.440, and precision was 97.452%. In conclusion, SVM emerges as the top model in our analysis, trailed, in that order, by Naive Bayes and Logistic Regression, respectively. This performs better than other classifiers because of the theoretical guarantee. Text categorization and text filtering are two text mining applications where SVM is outperforming other classifiers when it comes to the elements that should be considered when choosing a text classification algorithm [34, 49].

5.2 Model Performance Results using three algorithm with keeps some stop words (Max_df = 0.5 , Min_df = 5, Max_features = 5000, N-grams(1,2) (Unigram and Bigram)):

Table 3 displayed the performance assessment results obtained from using three different machine learning methods. The SVM algorithm stands out among the others in terms of accuracy.

Table 3. - Performance evaluation for three machine learning algorithms (Experiment 2).

Facebook comments				
Cross Validate , cv=10				
Classifier	***Precision***	***Recall***	***F1-score***	***Accuracy***
SVM	**0.98328**	**0.98311**	**0.983131**	**0.98311**
Naive Bayes	0.98059	0.98052	0.98054	0.98052
Logistic Regression	0.97597	0.97579	0.97579	0.97576

All three of the algorithms in **Table 3** showed improved accuracy after keeping a few stop words. Support Vector Machine (SVM), which achieved an accuracy of 98.311%, precision of 98.328%, recall of 98.311%, and F1 score of 98.313%, was the best-performing model among these techniques. Naive Bayes came in second, scoring 98.052% accuracy, 98.059% precision, 98.052% recall, and 98.054% F1 score. Logistic regression performed quite well With a recall of 97.576, an F1 score of 97.579, accuracy of 97.576%, and precision of 97.597%. In summary, SVM emerges as the top model in our analysis, trailed, in that order, by Naive Bayes and Logistic Regression.

Research on natural language processing has shown interest in the analysis of sentiment in Arabic text. Prior research has concentrated on several approaches to precisely categorize feelings; nevertheless, one factor that has been disregarded was the significance of keeping stop words. The tables showed that keeping certain stop words helps improve accuracy and decreases the number of positive and negative attitudes that are incorrectly classified. With the best accuracy in both sentiment analysis methodologies, The Support Vector Machine (SVM) algorithm is the most accuracy model in both sentiment analysis approaches. It performs the best.

The effectiveness of the proposed Arabic sentiment analysis model was compared with previous studies. The earlier research were selected based on a number of criteria: Arabic language material was present in the datasets, which were sourced from Facebook, Twitter, and Instagram. The classifiers used in the datasets were machine learning or deep learning techniques.

Looking at the results shown in **Table 4**, we can see that there is a noticeable difference in the accuracy rates between the various methods. Notably, the suggested model outperforms previous research with the highest accuracy, achieving 98.311% when using the Support Vector Machine (SVM) method.

For instance, in [11] achieved a 93% accuracy rate when analyzing 66,666 Twitter occurrences using logistic regression. Additionally, [12] used the SVM approach on a dataset of 3,700 Twitter occurrences and achieved a 78.2% accuracy. Moreover, in [13] 55,000 Twitter instances were used to achieve a 97% accuracy rate using the Naïve Bayes (NB) algorithm. Furthermore, using a bigger dataset of 85,751 Tweets, [14] reported an 81% accuracy rate using the SVM system. In contrast, [15] used the Convolutional Neural Network (CNN) method, which produced 92.8% accuracy on 90,187 tweets sent on Twitter. Finally, [16] showed 93% accuracy on a sample of 464,124 tweets using logistic regression and SVM techniques. This in-depth examination highlights the subtleties and incremental gains in accuracy rates when using various dataset sizes and techniques.

Table 4. - Comparison of the proposed system with previous research in the same fields

References	Dataset	Algorithm	Accuracy
Bolbol and Maghari [11] 2020	Arabic tweets	(Decision Trees)DT (k-nearest neighbors)KNN **Logistic Regression**	74% 74**%** **93%**
Alzyout, et al. [12] 2021	Arabic tweets	**SVM** KNN Naive Bayes classifier(NB) Decision Tree Classifier	**78.25%** 75.86% 71.07% 75.25%
Alderazi, et al. [13] 2021	Arabic tweets	**NB** (Long Short-Term Memory)LSTM NB + LSTM	**0.97** 0.93 91%
Kharsa and Harous [14] 2022	Arabic tweets	LR **SVM** DT RF KNN NB MLP	0.79 **0.81** 0.77 0.79 0.71 0.78 0.77

Alqarni and Rahman [15] 2023	Arabic tweets	**CNN** BiLSTM	**92.80** 91.99
Faisal, et al. [16] 2023	Arabic tweets	Random Forest **Logistic Regression** Classifier **Support Vector Machine – Classifier** Naive Bayes Classifier	92 **93** **93** 88
This study	Facebook comments	**SVM** Naïve Bayes Logistic Regression	**98.311** 98.052 97.576

6. CONCLUSION AND FUTURE WORKS

This study offered a method for assessing institutional performance using sentiment analysis from Facebook comments, with a particular focus on the Ministry of Justice in Iraq. To show how different machine learning algorithms perform in precisely detecting the emotions expressed in Arabic, a thorough experimental investigation was carried out. Retaining specific stop phrases greatly increased the model's accuracy and decreased sentiment misclassification, according to one important study. The Support Vector Machine (SVM) was the most successful approach examined, with an accuracy of 98.311%, precision of 98.328%, recall of 98.311%, and F1 score of 98.313%. Naive Bayes and Logistic Regression also showed strong performance.

For future research, it might be useful to investigate potential model updates and how the model can be used with other institutions and languages. Furthermore, using more deep learning techniques and analyzing a larger number of comments might improve the model's resilience and generalizability. Lastly, in addition to positive and negative feelings, a third categorization group for neutral sentiments might offer a more thorough knowledge of public opinion.

Funding

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

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, 2023.

[2] A. M. A. Ausat, "The Role of Social Media in Shaping Public Opinion and Its Influence on Economic Decisions," *Technology and Society Perspectives*, vol. 1, no. 1, pp. 35-44, 2023.

[3] E. A. Sukma, A. N. Hidayanto, A. I. Pandesenda, A. N. Yahya, P. Widharto, and U. Rahardja, "Sentiment analysis of the new indonesian government policy (omnibus law) on social media twitter," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2020, pp. 153-158: IEEE.

[4] 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, 2023.
[5] L. Buccoliero, E. Bellio, G. Crestini, and A. Arkoudas, "Twitter and politics: Evidence from the US presidential elections 2016," *Journal of Marketing Communications,* vol. 26, no. 1, pp. 88-114, 2020.
[6] J. Garside, H. Watt, and D. Pegg, "The Panama Papers: how the world's rich and famous hide their money offshore," *The Guardian,* vol. 3, 2016.
[7] K. Z. MERAL and Y. MERAL, "The Role of Social Media in Arab Spring," *Istanbul Medipol University,* 2017.
[8] S. E. Garcia, "The woman who created# MeToo long before hashtags," *The New York Times,* vol. 20, no. 20, 2017.
[9] I. Volkmer, "Social media and COVID-19: A global study of digital crisis interaction among Gen Z and millennials," *University of Melbourne. Retrieved May,* vol. 5, p. 2022, 2021.
[10] "Top Websites Ranking," https://www.similarweb.com/top-websites/iraq/computers-electronics-and-technology/social-networks-and-online-communities/, 2024.
[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)*, 2020, pp. 89-93: IEEE.
[12] M. Alzyout, E. A. 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)*, 2021, pp. 171-176: IEEE.
[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)*, 2021, pp. 112-117: IEEE.
[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)*, 2022, pp. 395-400: IEEE.
[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, p. 16, 2023.
[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)*, 2023, pp. 0222-0228: IEEE.
[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, p. 20420188221086693, 2022.
[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: Proceedings of the Third International Conference on Soft Computing and Data Mining (SCDM 2018), Johor, Malaysia, February 06-07, 2018*, 2018, pp. 135-146: Springer.
[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, 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)*, 2019, pp. 1379-1384: IEEE.
[21] A. Abdalrada, A. F. Neamah, and H. 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, 2024.
[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)*, 2019, pp. 1-5: IEEE.
[23] H. A. Alatabi and A. R. Abbas, "Sentiment analysis in social media using machine learning techniques," *Iraqi Journal of Science,* pp. 193-201, 2020.
[24] I. M. Enholm, E. Papagiannidis, P. Mikalef, and J. Krogstie, "Artificial intelligence and business value: A literature review," *Information Systems Frontiers,* vol. 24, no. 5, pp. 1709-1734, 2022.
[25] M. Schrage, D. Kiron, F. Candelon, S. Khodabandeh, and M. Chu, "The Future of Strategic Measurement: Enhancing KPIs With AI," *Future,* 2024.
[26] M. O. Hegazi, Y. Al-Dossari, A. Al-Yahy, A. Al-Sumari, and A. Hilal, "Preprocessing Arabic text on social media," *Heliyon,* vol. 7, no. 2, 2021.
[27] E. A. A. Refaee, "Sentiment analysis for micro-blogging platforms in Arabic," Heriot-Watt University, 2016.

[28] F. N. A. Hussein, "CYBERBULLYING CLASSIFICATION AND DETECTION IN TWITTER USING DATA MINING TECHNIQUES," University of Kerbala, 2024.
[29] M. H. Al-Hashimi, "Proposed Deep Learning Model for Rumour Detection in Facebook Posts in Arabic Language," *Journal of Algebraic Statistics,* vol. 13, no. 3, pp. 3249-3254, 2022.
[30] B. M. Sabbar, N. T. Yousir, and L. A. Habeeb, "Sentiment Analysis for Iraqis Dialect in Social Media Using Machine Learning Algorithms," *Iraqi Journal of Information and Communication Technology,* vol. 1, no. 2, 2018.
[31] V. D. Derbentsev, V. S. Bezkorovainyi, A. V. Matviychuk, O. M. Pomazun, A. V. Hrabariev, and A. M. Hostryk, "A comparative study of deep learning models for sentiment analysis of social media texts," in *CEUR Workshop Proceedings*, 2023, pp. 168-188.
[32] Z. Lu, "Web Page Classification Using Features from Titles and Snippets," Université d'Ottawa/University of Ottawa, 2015.
[33] S. A. Aljuhani and N. S. 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, 2019.
[34] 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.
[35] L. Auria and R. A. Moro, "Support vector machines (SVM) as a technique for solvency analysis," 2008.
[36] M. A. Khalifa, A. M. Ali, S. Alsadai, N. F. Alwan, and G. S. Mahdi, "A Novel Arabic Words Recognition System Using Hyperplane Classifier," *Wasit Journal of Computer **and** Mathematics Sciences,* vol. 1, no. 2, pp. 12-20, 2022.
[37] R. Rodríguez-Pérez, M. Vogt, and J. r. Bajorath, "Support vector machine classification and regression prioritize different structural features for binary compound activity and potency value prediction," *ACS omega,* vol. 2, no. 10, pp. 6371-6379, 2017.
[38] 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*, 1992, pp. 144-152.
[39] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, pp. 273-297, 1995.
[40] M. A. Fauzi, "Random forest approach fo sentiment analysis in indonesian," *Indones. J. Electr. Eng. Comput. Sci,* vol. 12, pp. 46-50, 2018.
[41] 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)*, 2016, pp. 257-261: IEEE.
[42] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier," in *Journal of Physics: Conference Series*, 2019, vol. 1192, no. 1, p. 012024: IOP Publishing.
[43] 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)*, 2021, pp. 932-936: IEEE.
[44] A. Tyagi and N. Sharma, "Sentiment analysis using logistic regression and effective word score heuristic," *International Journal of Engineering and Technology,* vol. 7, no. 2, pp. 20-23, 2018.
[45] 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)*, 2020, pp. 493-496: IEEE.
[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, 2015.
[47] A. S. J. B. S. J. AL-Jumaili, "A hybrid method of linguistic and statistical features for Arabic sentiment analysis," *Baghdad Science Journal,* vol. 17, no. 1 (Suppl.), pp. 0385-0385, 2020.
[48] S. A. Alsaidi, A. T. Sadeq, and H. S. Abdullah, "English poems categorization using text mining and rough set theory," *Bulletin of Electrical Engineering and Informatics,* vol. 9, no. 4, pp. 1701-1710, 2020.
[49] B. P. Meliani, O. N. Pratiwi, and R. Andreswari, "Comparison of Support Vector Machine and Random Forest Algorithms in Sentiment Analysis on Covid-19 Vaccination on Twitter using Vader and Textblob Labelling," in *International Conference on Applied Science and Technology on Social Science 2022 (iCAST-SS 2022)*, 2022, pp. 620-626: Atlantis Press.