Machine Learning for Predicting Parliamentary Elections Using Sentiment Analysis

Doaa Alkhiary[1], Samir Abu El Fotuoh Saleh[2], Mohamed Ebrahim Marie[3]

Faculty of commerce and Business Administration, Business Information System Department, Helwan University, Egypt[1],

Doaa.Mohamed021@commerce.helwan.edu.eg

Faculty of commerce, Information System, Mansoura University, Egypt.[2]

Faculty of Computers and Artificial Intelligence, Information System Department, Helwan University, Egypt[3]

ABSTRACT

During the rapid growth of digital technology, the huge increase in digital text data and diverse opinions on social media have created valuable opportunities for innovative sentiment analysis research. This research aims to assess public opinion across various life areas, political polarization, and its role in election campaigns. Twitter, in particular, acts as a rich source of data and public sentiment. This study analyzes a dataset gathered from Twitter's API, focusing on political opinions related to the 2020 parliamentary elections during both the pre-election period and election day. The dataset includes lists of political parties and independent candidates with Twitter accounts used for campaign promotion. The sample consists of 2,600 tweets, and techniques such as SVM, Naive Bayes (NB), Random Forest (RF), and Decision Trees (DT) were applied along with TF-IDF and weighted averaging to evaluate the results of each method and identify the most accurate one. The comparison revealed that Naive Bayes provided the highest accuracy.

Keywords: Sentiment Analysis, Machine Learning, Egyptian Parliament Election, algorithms

1.INTRODUCTION

A. BACKGROUND ON SENTIMENT ANALYSIS

Sentiment analysis is a crucial component of natural language processing that classifies text into positive, negative, or neutral sentiments. As online platforms allow individuals to share their opinions freely, it becomes essential for organizations to understand these sentiments to make informed decisions. Recognizing customer emotions regarding products and services can enhance satisfaction, improve brand reputation, and drive revenue growth.[11] Additionally, sentiment analysis is important in political contexts, helping assess public sentiment toward parties, candidates, and policies.

The sentiment analysis process involves several steps: preprocessing, feature extraction, and classification. Preprocessing refines raw text by removing irrelevant details, such as stop words, and converting the data into usable features through techniques like TF-IDF[13]. Feature extraction categorizes the processed text into sentiments using machine learning algorithms.

Researchers have concentrated on enhancing the precision of sentiment analysis methods for predicting election outcomes. For instance, Livne et al. (2011) [8] examined how candidates used Twitter during the U.S. midterm elections and discovered that incorporating both the network structure and user-generated content can lead to more accurate predictions. While traditional lexicon-based techniques have been widely applied in this domain, machine learning approaches are increasingly being recognized as effective alternatives. Numerous studies have demonstrated successful predictions of election results using social media data combined with computational models. Machine learning has also been utilized to forecast general election outcomes across different nations. Twitter, in particular, has emerged as a rich data source for sentiment analysis during elections due to its extensive user engagement [1].

B. Research Objective:

The main objective of the research was to develop a reliable forecasting model by implementing sentiment analysis within a computerized system for predicting election outcomes.

The Objectives of this research are given below:

- Examine whether sentiment analysis on Twitter can serve as a predictor for the popularity of specific parties or candidates.
- Explore the model of sentiment analysis and their impact on election campaigns.
- Construct a classifier using machine learning to conduct real-time sentiment analysis on gathered election-related data.

C. Research Questions:

How can models of Predictive Analytics assist policymakers in forecasting election results?.

- How can a sentiment analysis model help predict citizens› participation in elections?.
- Can sentiment analysis be beneficial in electoral campaigns?.
- What is the behavior of tweet sentiment related to a specific candidate or party over a certain period and how does it reflect the public›s attitudes towards political figures?.
- To what extent does Twitter accurately reflect the political sentiment of the electorate?.

II. Literature Review

A. Sentiment Analysis Algorithms

Sentiment analysis algorithms include a range of algorithms for classifying sentiments within text data[9]. Among these, Support Vector Machines (SVM) are widely used due to their robustness in managing high-dimensional spaces, ease of training, and efficient memory usage through kernel functions. SVM is especially effective for sentiment analysis involving large, sparse datasets[14]. Another popular model is Naive Bayes (NB), which is appreciated for its straightforward implementation and understanding, as well as its low computational demands and quick training times[2]. However, NB relies on the assumption that features are independent, which may not always be valid. Random Forest (RF) is a notable ensemble learning model consisting of multiple decision trees and is praised for its accuracy and interpretability. Decision Trees (DT) are relatively easy to comprehend and produce accurate results by incrementally learning from the data[21]. They are also memory-efficient and handle noisy data well, though they can require considerable training time[1].

In sentiment analysis, these machine learning model are essential for categorizing text into positive, negative, or neutral sentiments. Utilizing models like SVM, NB, RF, and DT, researchers can extract meaningful insights from textual data to better understand emotions and opinions. As shown in figure (1).

The employee performance is measured based on various factors, such as commitment level, responsibility degree, discipline level, flexibility Level, teamwork capability level, time schedule commitment degree, and stress control degree,

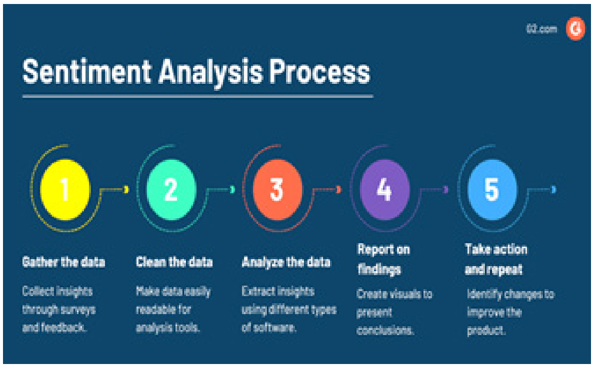


Figure 1: Sentiment analysis process

B. Related Work

Many studies indicate that analyzing sentiments and patterns offers valuable insights into public opinion regarding elections and government policies. For example, Opeoluwa et al. [12] performed sentiment and emotion analysis on prominent candidates, measuring the closeness between political parties based on sentiment and emotion distances, where smaller distances suggest stronger political ties. Similarly, research by Xie et al. (2018) [13] and Jaidka et al. (2019) [14] demonstrated the effectiveness of Twitter data in predicting election outcomes and understanding public sentiment.

Previous work [15] has addressed the challenges of analyzing political tweets, particularly focusing on how sarcasm affects classifier accuracy. For instance, [16] explored strategies to handle sarcastic tweets where positive sentiments appear in negative contexts. For more granular sentence-level analysis, tools like the Stanford Dependency Parser [17] are frequently employed.

Furthermore, the research in [6] applied machine learning classifiers such as Support Vector Machines (SVM) and Naive Bayes (NB) to both Arabic and English datasets. Their findings revealed that SVMs outperform NB classifiers, with only slight differences when using term frequency (TF) versus term frequency-inverse document frequency (TF-IDF) weighting methods.

III. Research Methodology

Research methodologies include:

- The use of a traditional descriptive and analysis approach.
- A case study focused on the Egyptian Parliament Election of 2020.

- Preprocessing of data.
- Sentiment classification through the application of SVM, Naive Bayes, Decision Tree, Random Forest for tweet classification.
- Development of a Model based on sentiment analysis to predict election results.
- Evaluation: Assessment of classification effectiveness through measures of accuracy, precision, recall and F-measure.

Figure (2) illustrates the phases of the Model according to the selected Research Methodology

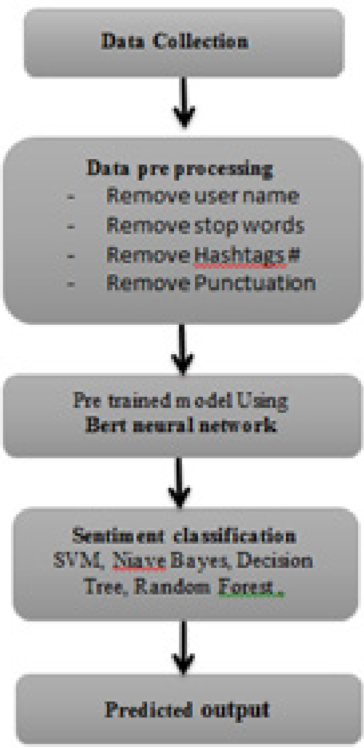


Figure 2 : Methodology Steps

In the following subsection the detail description research model phase and each phase steps is presented .

A. Data collection

The dataset obtained from Twitter using various data about political opinions related to the 2020 Parliamentary Elections, both before and during the election days.

B. Data pre-processing

- Removing @user :

The tweets include multiple Twitter handles (@ users) that identify individual users on the platform. All of these handles are subsequently removed from the dataset. To simplify the workflow, a combined training and testing set is utilized, which reduces time and effort by preventing redundant processing steps

.- Remove Hashtags :

This refers to the procedure of detecting and eliminating hashtags (words or phrases preceded by the «#» symbol) from text data. Removing these hashtags can enhance the accuracy and consistency of the data utilized for training models.

- Remove the URLs from the Text :

A typical step in text data preprocessing involves detecting and removing web addresses or URLs from the text. Eliminating URLs helps to clean the dataset, enhancing its relevance and boosting the effectiveness of sentiment analysis and machine learning models.

- Remove Stop-Words

A stop-word is a term in English that provides minimal or no significant information. Removing these words is essential in text preparation. Each language has its own list of stop-words available in the NLTK library.

شاء » «مصر», «انت»,«مش»,«اللى»»
»,»كلها»,»دايما»,»لازم», »لله»,»ياه», »ليس», »مافيش

- Remove punctuations

This text preprocessing step involves removing

punctuation marks–such as periods, commas, exclamation points, and question marks–from the text data. Eliminating punctuation results in cleaner, more standardized text, which can enhance the accuracy and efficiency of analytical models. [«يا» «,» و «,«.«,«،«,«.«,«/«,«-«,«_«,«ة«,«آ«,«إ«,«أ
ى«,«««,«ـ«,‹««,»»»,»\\»,›\n›, ‹\t›,›"›,›?›,›؟›,›!›].

C. Term Frequency(TF) and Inverse Document Frequency (IDF) :

The TF-IDF score is a commonly employed technique in information retrieval and text summarization. It was developed to emphasize the significance of a term within a specific document. TF-IDF combines two components: Term Frequency (TF) and Inverse Document Frequency (IDF) to extract meaningful features[2]. IDF assigns higher weights to tokens that occur infrequently across the dataset. For instance, if a unique phrase appears twice, it greatly affects the interpretation of the related sentences. The TF-IDF value is obtained by multiplying the TF and IDF scores.

Term Frequency (TF) measures how often a term appears in a document. It is calculated as the ratio of the number of times a term appears in a document to the total number of terms in that document.[12]

Inverse Document Frequency (IDF) measures the rarity of a term across the entire corpus. It is calculated as the logarithm of the ratio of the total number of documents in the corpus to the number of documents containing the term.

For each word TF- IDF is equal to the log (N/n), where, N is the number of documents, and n is how many times it has been in each document.

$$TF\_IDF = TF*IDF \quad (1)$$

Finally, Term Frequency-Inverse Document Frequency (TF-IDF) is used to represent each document as a weighted vector based on the terms present in the document collection. Each term in a document is assigned a weight according to various TF-IDF weighting schemes. Specifically, the TF-IDF weight of a term(ti) depends on the number of documents in the corpus where appears at least once, and the inverse document frequency (IDF) of the term(ti) , as defined in the previous paragraph[22].

$$(ti) = \mathrm{Log}\frac{D}{DF(ti)} \quad (2)$$

Where D is the total number of documents in the dataset.

The weight of term (ti) in a document (di) TF.IdF is defined below.

TF.IDF (ti,di) = TF*IDF(ti) (3)

Consider a document with 100 words where the term «هنتخب» (I will vote) appears 3 times. According to the formulas, the term frequency (TF) for «هنتخب» is 0.03 (3100/). Suppose there are 18,000 documents in total, and «هنتخب» appears in 1,000 of these. The inverse document frequency (IDF) is calculated as log (18000 / 1000) = 1.25.

The TF-IDF score is the product of these quantities: 0.03 × 1.25 = 0.037.

In general, it serves as a metric for assessing the significance of a word within a document relative to a collection, thereby enhancing the recall and precision of the retrieved documents[24].

Furthermore, integrating Weighted Average into sentiment analysis algorithms aids in calculating the average weight of words based on their relevance to sentiment. By assigning different weights to words based on their importance in conveying sentiment, Weighted Average enhances the precision of sentiment classification models. The combination of TF-IDF and Weighted Average techniques significantly boosts the accuracy and effectiveness of sentiment analysis in Arabic, especially when identifying sentiments expressed on platforms like Twitter during events such as elections[13].

D. Sentiment Classification

The obtained data in Figure 3 is divided into training and testing dataset. Training dataset used to build the classification models based on SVM, Naive Bayes, Random Forest and Decision Tree classifiers. The data classified based on its polarity to positive and negative classes .

Testing dataset is used to predict the polarity of the tweets.

The data has been divided into 20% testing and 80% training.

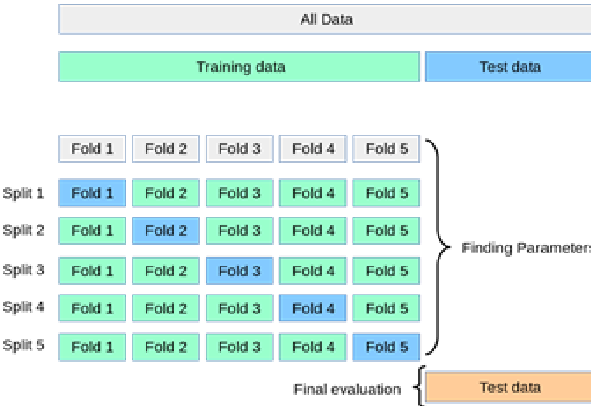


Figure 3. Cross validation in Machine learning

Figure 3 show that Cross-validation involves partitioning a dataset into subsets, training the model on some subsets (training set), and validating it on the remaining subsets (validation set). This process is repeated multiple times to ensure that every data point has a chance to be in both the training and validation sets.

- Support Vector Machine (SVM):

Support Vector Machines (SVM) are a set of supervised learning algorithms applicable to both classification and regression problems. In basic terms, given a collection of labeled training examples divided into two categories, the SVM algorithm builds a model capable of predicting the category of new instances. Fundamentally, the SVM represents data points in a multidimensional space, optimized to separate the two classes by the widest possible margin[24].

$F(x) = \text{sign}(wx+b)$ (4)

Where w is a weighted vector in Rn and b is known as the bias.

SVMs find the hyper plane $y=wx+b$ by separating the space Rn into two half spaces with the maximum-margin[25].

- Result of Classification Using Support Vector Machine (SVM):

Table 1 shows that the result of SVM Algorithm, the Accuracy, Precision, Recall and F measure for Every Candidate.

Table 1: The result of SVM Algorithm for Every Candidate

Candidate	Precision	Recall	F measure	Accuracy
محمود بدر	0.880	0.880	0.875	0.875
قائمة تحالف المستقلين	0.830	0.830	0.833	0.833
قائمة نداء مصر	0.550	0.550	0.545	0.545
مرتضى منصور	0.780	0.780	0.777	0.777
هيثم الحريرى	0.890	0.890	0.888	0.888
احمد طنطاوى	0.620	0.620	0.625	0.625
القائمة الوطنية	0.600	0.600	0.600	0.601
حزب مستقبل وطن	0.730	0.730	0.727	0.727
ضياء الدين داوود	1.000	1.000	1.000	1.000
قائمة ابناء مصر	1.000	1.000	1.000	1.000
اشرف رشاد عثمان	0.658	0.658	0.667	0.667

The obtained data in Table (1) observed that, the higher precision, recall, F measure and accuracy were in قائمة أبناء مصر .ضياء الدين داوود. While the lower precision, recall, F measure and accuracy were in قائمة نداء مصر.

Table 2 : Accuracy and Weighted Average of Support vector machine (SVM) for every Candidate.

candidate	*accuracy*
محمود بدر	0.875
قائمة تحالف المستقلين	0.833
قائمة نداء مصر	0.545
مرتضى منصور	0.777
هيثم الحريرى	0.888
احمد طنطاوى	0.625
القائمة الوطنية	0.601
حزب مستقبل وطن	0.727
ضياء الدين داوود	1.000
قائمة ابناء مصر	1.000
اشرف رشاد عثمان	0.667
Weighted Average	**0.7723**

- Naïve Bayes (NB) :

The Naïve Bayes classifier is a robust classification algorithm frequently used in sentiment analysis and document classification. As a probabilistic model, it calculates the likelihood of categories for given test data by leveraging the joint probabilities of terms and their associated categories $P(c|x) = P(c)P(x|c) / P(x)$ (5)

Where:

$P(c|x)$ is the posterior probability of class c given feature vector x

$P(c)$ is the prior probability of class c.

$P(x|c)$ is the likelihood which is the probability of feature vector x given class c.

$P(x)$ is the prior probability of data x .[25].

Result of Classification Using Naive Bayes (NB):

The result of NB Algorithm, the Accuracy, Precision, Recall and F measure for Every Candidate. As shown in Table 3.

Table 3:The Result of Niave Bayes Algorithm for Every Candidate

Candidate	Precision	Recall	F measure	Accuracy
محمود بدر	0.880	0.880	0.875	0.875
قائمة تحالف المستقلين	0.830	0.830	0.833	0.833
قائمة نداء مصر	0.730	0.730	0.727	0.727
مرتضى منصور	0.780	0.780	0.777	0.777
هيثم الحريرى	0.890	0.890	0.889	0.889
احمد طنطاوى	0.620	0.620	0.625	0.625
القائمة الوطنية	0.600	0.600	0.600	0.600
حزب مستقبل وطن	0.730	0.730	0.727	0.727
ضياء الدين داوود	1.000	1.000	1.000	1.000
قائمة ابناء مصر	1.000	1.000	1.000	1.000
اشرف رشاد عثمان	0.670	0.670	0.666	0.666

The obtained data in Table (3) observed that, the higher precision, recall, F measure and accuracy were in ضياء الدين داوود. قائمة أبناء مصر. While the lower precision, recall, F measure and accuracy were in القائمة الوطنية.

Table 4: Accuracy and Weighted Average of Naive Bayes (NB) for every Candidate

Candidate	*accuracy*
محمود بدر	0.875
قائمة تحالف المستقلين	0.833
قائمة نداء مصر	0.727
مرتضى منصور	0.777
هيثم الحريري	0.889
احمد طنطاوى	0.625
القائمة الوطنية	0.600
حزب مستقبل وطن	0.727
ضياء الدين داوود	1.000
قائمة ابناء مصر	1.000
اشرف رشاد عثمان	0.666
Weighted Average	**0.8285**

- Decision Tree (DT) :

A decision tree algorithm is a supervised learning method commonly applied in data mining and machine learning. It operates by recursively dividing a dataset into smaller subsets, each linked to a target value [23]. This division is guided by posing questions about the data›s attributes.

$$\text{Entropy}(S) = -\sum_{-CEC}⟦P(C)\text{Log}_2 p(c)⟧ \quad (6)$$

- A represents a specific attribute or class label
- Entropy(S) is the entropy of dataset, S.
- Sv|/ |S| represents the proportion of the values in Sv to the number of values in dataset, S.
- Entropy(Sv) is the entropy of dataset, Sv .[25].

- Result of Classification using Decision Tree:

Table 5 shows that the result of DT Algorithm, the Accuracy, Precision, Recall and F measure for Every Candidate.

Table 5: The Result of Decision Tree Algorithm for Every Candidate

Candidate	Precision	Recall	F measure
محمود بدر	0.880	0.880	0.875
قائمة تحالف المستقلين	0.670	0.670	0.667
قائمة نداء مصر	0.550	0.550	0.545
مرتضى منصور	0.780	0.780	0.777
هيثم الحريرى	0.890	0.890	0.888
احمد طنطاوى	0.380	0.380	0.375
القائمة الوطنية	0.400	0.400	0.400
حزب مستقبل وطن	0.450	0.450	0.454
ضياء الدين داوود	1.000	1.000	1.000
قائمة ابناء مصر	1.000	1.000	1.000
اشرف رشاد عثمان	0.670	0.670	0.666

The obtained data in Table (5) observed that, the higher precision, recall, F measure and accuracy were in ضياء الدين داوود. قائمة أبناء مصر. While the lower precision, recall, F measure and accuracy were in أحمد طنطاوي

Table 6: Accuracy and Weighted Average of Decision Tree for every Candidate.

candidate	*accuracy*
محمود بدر	0.875
قائمة تحالف المستقلين	0.667
قائمة نداء مصر	0.545
مرتضى منصور	0.777
هيثم الحريرى	0.888
احمد طنطاوى	0.375
القائمة الوطنية	0.400
حزب مستقبل وطن	0.454
ضياء الدين داوود	1.000
قائمة ابناء مصر	1.000

اشرف رشاد عثمان	0.666
Weighted Average	**0.703**

- Random Forest (RF) :

Random forest builds multiple decision trees using various subsets of the original dataset and determines the final class by selecting the mode of the classes predicted by the individual trees [25].

$$\text{Gini Index} = 1 - \sum_{i=1}^{n}(Pi)^2 \quad (7)$$

$$= 1 - [(P_{-})^2 + (P_{-})^2] \quad (8)$$

- Result of Classification Using Random Forest:

Table 7 shows that the result of RF Algorithm, the Accuracy, Precision, Recall and F measure for Every Candidate.

Candidate	Precision	Recall	F measure	Accuracy
محمود بدر	0.880	0.880	0.875	0.875
قائمة تحالف المستقلين	0.830	0.830	0.833	0.833
قائمة نداء مصر	0.550	0.550	0.545	0.545
مرتضى منصور	0.780	0.780	0.777	0.777
هيثم الحريرى	0.890	0.890	0.888	0.888
احمد طنطاوى	0.620	0.620	0.625	0.625
القائمة الوطنية	0.600	0.600	0.600	0.600
حزب مستقبل وطن	0.640	0.640	0.636	0.636
ضياء الدين داوود	1.000	1.000	1.000	1.000
قائمة ابناء مصر	1.000	1.000	1.000	1.000
اشرف رشاد عثمان	0.670	0.670	0.666	0.666

The obtained data in Table (7) observed that, the higher precision, recall, F measure and Accuracy were in قائمة أبناء مصر .ضياء الدين داوود While the lower precision, recall, F measure and accuracy were in قائمة نداء مصر.

Table 8: Accuracy and Weighted Average of Random Forest for every Candidate

candidate	*accuracy*
محمود بدر	0.875
قائمة تحالف المستقلين	0.833
قائمة نداء مصر	0.545
مرتضى منصور	0.777
هيثم الحريرى	0.888
احمد طنطاوى	0.625
القائمة الوطنية	0.600
حزب مستقبل وطن	0.636
ضياء الدين داوود	1.000
قائمة ابناء مصر	1.000
اشرف رشاد عثمان	0.666
Weighted Average	**0.752**

Table 9: Comparison of various Classification Accuracy

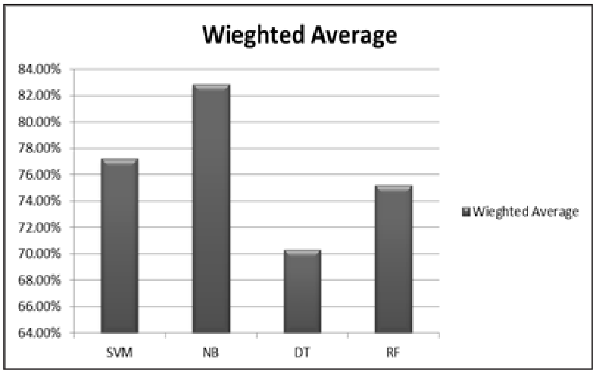
Classification Type	SVM	NB	DT	RF
Weighted Average	0.7723	0.8285	0.703	0.752

The results showed the accuracy and Weighted Average of the Naive Bayes algorithm, which is considered the most famous, accurate, and fastest algorithm compared to other algorithms, where it achieved the highest weighted average of 82.85%, which is the highest possible accuracy.

IV . Experimental Results and Evaluation:

The results indicate that tweets were analyzed using several algorithms, including SVM, Naive Bayes (NB), Decision Tree (DT), and Random Forest (RF). Specifically, SVM achieved an accuracy and weighted average of 77.23%, while NB achieved 0.82.85% , Decision Tree recorded 70.3%. Random

Forest showed a weighted average of 75.2%. Among these, the Naive Bayes algorithm–known for being popular, accurate, and fast–attained the highest weighted average of 82.85%, representing the best overall accuracy. Conversely, the Decision Tree algorithm exhibited the lowest accuracy and weighted average, as illustrated in Figure 4



V. CONCLUSION AND FUTURE WORK :

The study highlights the importance of sentiment analysis in understanding public opinion, particularly on social media platforms like Twitter. By utilizing machine learning algorithms such as Support Vector Machines, Naive Bayes, Random Forests, and Decision Trees, researchers can gain valuable insights from the sentiments expressed about the Egyptian Parliament elections. The Naive Bayes algorithm, recognized for its popularity, accuracy, and speed, demonstrated the highest weighted average accuracy . Many of the results align closely with actual outcomes.

Overall, sentiment analysis is essential for gauging public sentiment during Egyptian elections. By utilizing advanced analytical techniques and developing specialized resources, such as a colloquial electoral dictionary, The parliamentary election outcome prediction model can be a powerful tool for decision-makers in Egypt, contributing to the development of the democratic process and the achievement of political stability.

More research is needed to improve the accuracy of sentiment prediction by considering sarcasm and emojis, by addressing more complex areas such as spam and negation using advanced tools.

References :

[1] Khalaf Salman Al-Tameemi, I., Feizi-Derakhshi, M. R., Pashazadeh, S., & Asadpour, M. (2022). A Comprehensive Review of Visual-Textual Sentiment Analysis from Social Media Networks. arXiv e-prints, arXiv-2207.

[2] Brahimi, B., Touahria, M., & Tari, A. (2021). Improving sentiment analysis in Arabic: A combined approach. Journal of King Saud University-Computer and Information Sciences, 33(10), 1242.1250-

[3] Oladele, T. M., & Ayetiran, E. F. (2023). Social unrest prediction through sentiment analysis on Twitter using support vector machine: experimental study on Nigeria's# EndSARS. Open Information Science, 7(1), 20220141.

[4] Ghallab, A., Mohsen, A., & Ali, Y. (2020). Arabic sentiment analysis: A systematic literature review. Applied Computational Intelligence and Soft Computing, 2020(1), 7403128.

[5] Shevtsov, A., Oikonomidou, M., Antonakaki, D., Pratikakis, P., & Ioannidis, S. (2023). What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020. Plos one, 18(1), e0270542.

[6] Elshakankery, K., & Ahmed, M. F. (2019). HILATSA: A hybrid incremental learning approach for Arabic tweets sentiment analysis, Egyptian Inform.

[7] Omara, E., Mousa, M., & Ismail, N. (2022). Character gated recurrent neural networks for Arabic sentiment analysis. Scientific Reports, 12(1), 9779.

[8] Livne, A., Simmons, M., Adar, E., & Adamic, L. (2011)

The party is over here: Structure and content in the 2010 election.In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 5, No. 1, pp. 201208-).

[9] Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. Applied Sciences, 13(7), 4550.

[10] Mataoui, M. H., Zelmati, O., & Boumechache, M. (2016). A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic. Research in Computing Science, 110(1), 55.70-

[11] Wang, H., Hanafy, A., Bahgat, M., Noeman, S., Emam, O. S., & Bommireddipalli, V. R. (2015, April). A system for extracting sentiment from large-scale Arabic social data. In 2015 First International Conference on Arabic Computational Linguistics (ACLing) (pp. 7177-). IEEE.

[12] Chaudhary, L., Girdhar, N., Sharma, D., Andreu-Perez, J., Doucet, A., & Renz, M. (2023). A Review of Deep Learning Models for Twitter Sentiment Analysis: Challenges and Opportunities. IEEE Transactions on Computational Social Systems.

[13] Alanazi, S. A., Khaliq, A., Ahmad, F., Alshammari, N., Hussain, I., Zia, M. A., & Afsar, S. (2022). Public's mental health monitoring via sentimental analysis of financial text using machine learning techniques. International Journal of Environmental Research and Public Health, 19(15), 9695.

[14 Alfreihat, M., Almousa, O., Tashtoush, Y., AlSobeh, A., Mansour, K., & Migdady, H. (2024). Emo-SL Framework: Emoji Sentiment Lexicon Using Text-Based Features and Machine Learning for Sentiment Analysis. IEEE Access.

[15] Saldivar, J., Parra, C., Laconich, M., & Cernuzzi, L. (2022). The electoral success of social media losers: a study on the usage and influence of Twitter in times of elections in Paraguay. SN Social Sciences, 2(7), 98.

[16] Alqarni, A., & Rahman, A. (2023). Arabic Tweets-based Sentiment Analysis to investigate the impact of COVID-19 in KSA: A deep learning approach. Big Data and Cognitive Computing, 7 (1): 16.

[17] Mataoui, M. H., Zelmati, O., & Boumechache, M. (2016). A proposed lexicon-based sentiment analysis approach

for the vernacular Algerian Arabic. Research in Computing Science, 110(1), 5570-.

[18] Aljameel, S. S., Alabbad, D. A., Alzahrani, N. A., Alqarni, S. M., Alamoudi, F. A., Babili, L. M., ... & Alshamrani, F. M. (2021). A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia. International journal of environmental research and public health, 18(1), 218.

[19] Alotaibi, T., & Al-Dossari, H. (2024). A Review of Fake News Detection Techniques for Arabic Language. International Journal of Advanced Computer Science & Applications, 15(1).

[20] Alowaidi, S., Saleh, M., & Abulnaja, O. (2017). Semantic sentiment analysis of Arabic texts. International Journal of Advanced Computer Science and Applications, 8(2).

[21] Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach. Social Network Analysis and Mining, 13(1), 31.

[22] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-.1113

[23] Mushtaq, M. F., Fareed, M. M. S., Almutairi, M., Ullah, S., Ahmed, G., & Munir, K. (2022). Analyses of public attention and sentiments towards different COVID-19 vaccines using data mining techniques. Vaccines, 10(5), 661.

[24] Elghazaly, T., Mahmoud, A., & Hefny, H. A. (2016, March). Political sentiment analysis using twitter data. In Proceedings of the International Conference on Internet of things and Cloud Computing (pp. 1.(5-

[25] Doaa,A., Saleh, S., & Marie, M. (2025). a Proposed Machine learning model for predicting Egyptian Parliament Election Results. International Journal of Advanced Science and Computer Applications, 4(1).