第 49 卷第 2 期

七月 2024

Diqiu Kexue - Zhongguo Dizhi Daxue Xuebao/Earth Science Journal of China University of Geosciences

https://dqkx-periodicals.com

Vol. 49 No. 2

Oct - Dec 2 0 2 4

DOI: 10.5281/zenodo.14655667

Sentiment Analysis Model for Parliamentary Elections Combining Electoral Dictionary Using Machine Learning

Doaa Mohamed Alkhiary[a], Samir Abu El Fotuoh Saleh[b], Mohamed Ebrahim Marie[c]

Faculty of commerce and Business Administration, Business Information System Department,

Helwan University, Egypt

Corresponding author at: doaa.alkhiary@yahoo.com

Received: 10 November2024
Accepted: 17 December 2024
Published: 15 January 2025

Abstract: Amid the massive digital boom, the vast expansion of digital textual data, and the variety of opinions on social media, significant opportunities have emerged for innovative research in sentiment analysis to gauge public opinion across various life domains, political polarization, and its use in election campaigns. Twitter, in particular, serves as a repository of data and opinions. This study focuses on a dataset collected from Twitter via the API, related to political opinions on the 2020 parliamentary elections, both in the pre-election period and on election days. The data includes lists of parties and independent candidates with Twitter accounts used to promote their campaigns. The sample contains 2600 tweets, and techniques such as Support vector machine (SVM), Naive Bayes (NB),Random Forest (RF), and Decision Tree (DT) were applied, along with TF-IDF and weight average, to obtain results for each technique and determine which is more accurate and reflective of reality. The comparison shows that NB is the most accurate technique. This study also found that having a specialized dictionary for electoral terminology is essential, as no researchers have yet developed a dictionary specifically for electoral terms in Egyptian colloquial Arabic.

Keywords: *sentiment analysis, Election, parliament, techniques, dictionary*.

1. Introduction

1.1 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, including support vector machines .

Applications of sentiment analysis are diverse, spanning various industries.

In finance, for example, it helps analyze news articles and social media posts to forecast stock prices and identify investment opportunities [9]. It also plays a vital role in predicting election outcomes and gauging public opinions on different political policies. Although much research has focused on English text, there is increasing demand for sentiment analysis studies in Arabic due to its limited availability[12].

By leveraging sentiment analysis methodologies, organizations can gain valuable insights from user-generated content on platforms like Twitter[16]. This data enables companies to monitor user sentiments towards products, services, or events, ultimately improving business strategies and enhancing customer satisfaction. Understanding consumer attitudes through sentiment analysis can lead to increased sales and better alignment of offerings with customer expectations[10].

1.2. Literature Review

Sentiment analysis methods encompass a variety of strategies for categorizing sentiments within text data[9]. One commonly used method is Support Vector Machines (SVM), which is renowned for its robustness in high-dimensional spaces, ease of training, and efficiency in memory usage through kernel mapping. SVM proves to be particularly effective in sentiment analysis when handling extensive, sparse sample collections[14]. Another technique, Naive Bayes (NB), stands out for its simplicity in implementation and comprehension, along with its minimal computational requirements and training duration[2]. However, it is important to note that NB assumes independence among features, which may not always hold true.. Random Forest (RF) stands as a noteworthy ensemble learning technique comprised of multiple decision trees, recognized for its accuracy and interpretability. Decision Trees (DT) are relatively straightforward to grasp and yield precise outcomes by progressively learning from the data[21]. They excel in memory efficiency and adeptly handle noisy data, though they may necessitate substantial training[1].

In the realm of sentiment analysis, these machine learning techniques play a pivotal role in categorizing sentiments into positive, negative, or neutral classifications. By leveraging algorithms like SVM, NB, RF, and DT, researchers can glean valuable insights from textual data to gain a deeper understanding of people's emotions and viewpoints. As shown in figure (1).

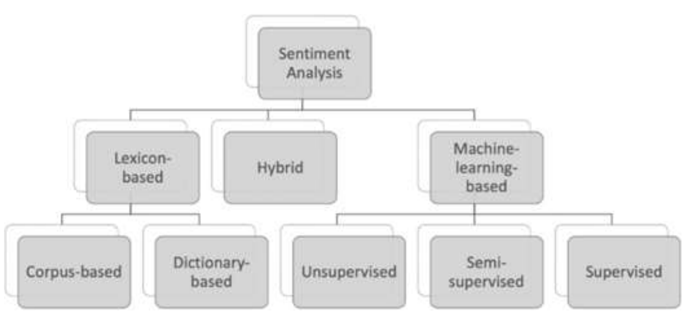

Figure 1.2. Sentiment analysis approaches

1.3 Related Work

Numerous studies suggest that analyzing sentiments and patterns can provide valuable insights, helping to understand public opinion on elections and government policies. In study Opeoluwa et al. [12], researchers conducted sentiment and emotion analysis for prominent candidates, assessing the closeness of political parties through sentiment and emotion data, where a smaller distance implies stronger political connections. Studies, Xie et al.(2018) [13] and Jaidka et al. (2019) [14] also highlighted the effectiveness of Twitter data for predicting election results and gaining insights into public opinion.

Earlier studies [15] have tackled the challenges involved in analyzing political tweets, with a focus on the effects of sarcasm on classifier accuracy. For instance, [16] investigated methods for managing sarcastic tweets in which positive sentiments are contrasted with negative contexts. For more detailed sentence analysis, tools like the Stanford Dependency Parser[17]are commonly used.

Moreover, [18], [19], and [20] introduced deep learning techniques, including the bidirectional encoder from transformers (BERT), as emerging approaches in sentiment analysis for election prediction [6].

The study in [6] applies machine learning classifiers to both Arabic and English corpora, specifically utilizing Support Vector Machines (SVM) and Naive Bayes (NB) classifiers. The results indicate that SVMs outperform the NB classifier, with minimal differences observed between using term frequency (TF) and term frequency-inverse document frequency (TF-IDF) as weighting methods.

Additionally, Heredia et al.(2018) [18], Bose et al.(2019) [19], and Bilal et al.(2018) [20] introduced deep learning methods, such as the bidirectional encoder representations from transformers (BERT), as promising approaches for sentiment analysis in election forecasting [6].

2. Method

To achieve the objective of this research, the following research methods, techniques and steps will be applied:

- Traditional Descriptive and analysis method.
- Case study: Egyptian parliamentary elections 2020 .
- Proposed Model building: build system based on sentiment analysis for predicting the outcome of elections .
- Classification for tweets (Positive , Negative , Neutral).
- Proposed Dictionary of Political Terminology for Elections.

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

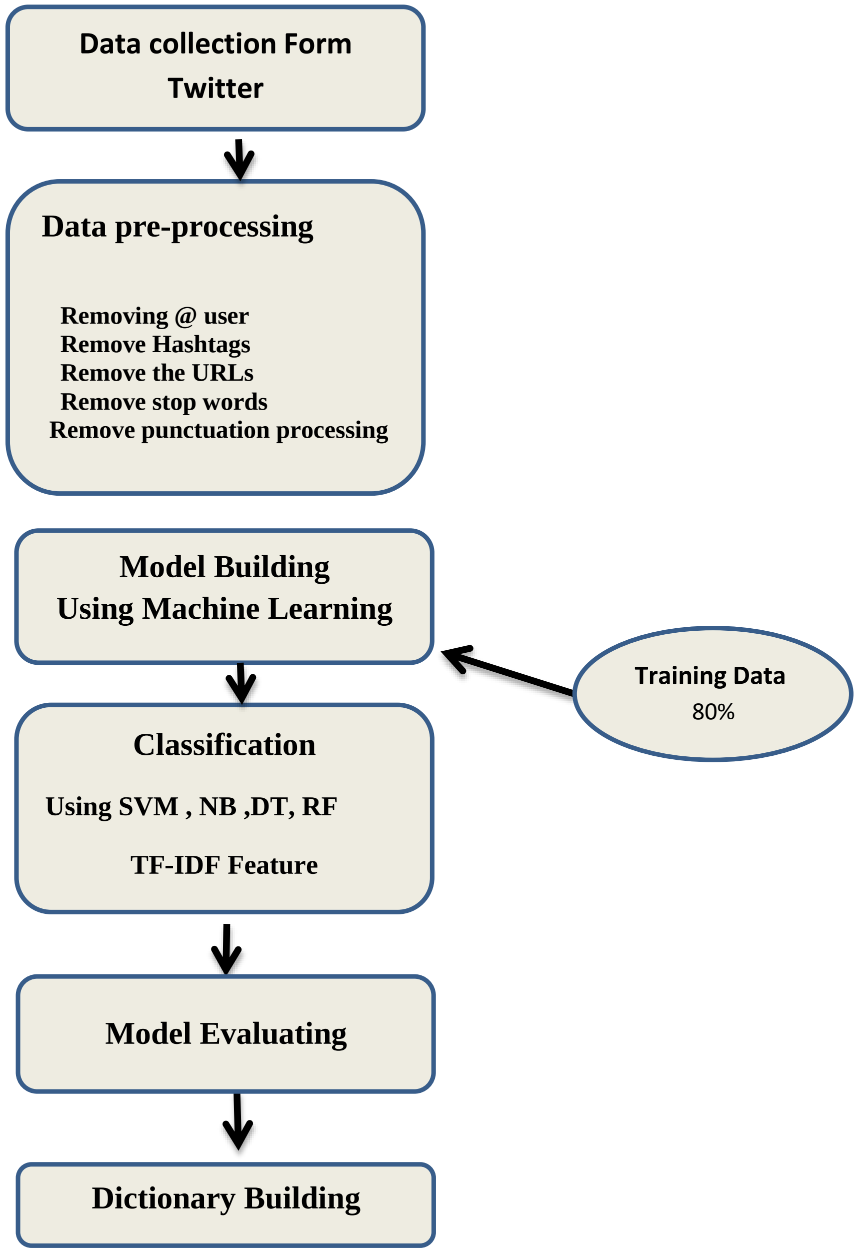

Figure 2 . The proposed Model for Parliamentary Elections

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

2.1. Data collection

The dataset gathered from Twitter via the Twitter API contains a variety of data regarding political opinions about the 2020 parliamentary elections, both in the lead-up to and during the election days. The data includes party lists and independent candidates who have Twitter accounts and have promoted their campaigns on Twitter. This dataset consists of thousands of tweets. However, during the collection process, the API connection was cut off by Twitter, halting further data collection using the same keywords and leading to increased repetition among the gathered tweets.

After cleaning and filtering the tweets, the total volume of data was reduced to 2600 was divided into training and testing dataset.

2. 2. Data pre-processing

- Removing @user :

The tweets contain various Twitter handles (@users), which identify Twitter users. Additionally, all these handles are removed from the database. To streamline the process, we use a combined train and test set, saving time and effort by avoiding repeated actions.

- Remove Hashtags :

Refers to the process of identifying and deleting hashtags (words or phrases preceded by the "#" symbol) from text data.

Removing them can help improve the accuracy and consistency of the data used for training models.

- Remove the URLs from the Text :

Is a common step in text data pre-processing where web addresses or links (URLs) are identified and deleted from the text. Removing URLs helps to clean the text, making it more relevant and improving the performance 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.

"God willing," "Egypt," "you," "not," "the one," "always," "must," "for God," "oh," "not," "there is no."

- Remove punctuations

Is a text pre-processing step where punctuation marks—such as periods, commas, exclamation points, and question marks—are removed from text data. By removing punctuation, the text becomes cleaner and more uniform, which can improve the accuracy and efficiency of analytical models ["A","I","AA","h","_","-","/",".",","," and "," O ","'","-","'","y","\",'\n', '\t','"','?','؟','!'].

2.3. Utilization of TF-IDF and Weight Average in Sentiment Analysis

In sentiment analysis, key components such as TF-IDF (Term Frequency-Inverse Document Frequency) and Weighted Average play a vital role, especially when dealing with Arabic content. The integration of TF-IDF helps determine the significance of words within a document in relation to a corpus, which aids in identifying crucial terms that convey sentiment. This method allows for the recognition of essential words that contribute to the overall sentiment expressed [2].

TF-IDF is calculated by multiplying two values: Term Frequency (TF) and Inverse Document Frequency (IDF)[12]. 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.

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

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].

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

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**.**

$$\text{TF.IdF (ti,di)} = \text{TF*IdF(ti)} \quad (3)$$

Consider a document with 100 words where the term " We will elect" (I will vote) appears 3 times. According to the formulas, the term frequency (TF) for " We will elect" is 0.03 (3/100). Suppose there are 18,000 documents in total, and " We will elect" 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].

2.4. Sentiment Classification

The data 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 .
This section covers four existing approaches to text classification : SVM, NB,DT,RF .

- Using pre-trained model to label the data:

Pre-training models have emerged as powerful techniques for effectively stratifying and labeling data in various natural language processing tasks, including sentiment analysis.

Table 1 . True positive and true Negative of each candidate

Candidate	Positive	Negative	Neural
Mahmoud Bader	81%	8%	11%
Independent Alliance List	54%	8%	38%
Call of Egypt	22%	49%	29%
Mourtda mansour	32%	64%	4%
Haithem Alhariri	72%	24%	4%
Ahmed tantawy	39%	42%	19%
National List for Egypt	32%	48%	20%
Future of a Nation Party	13%	54%	33%
Diaa Eldin Dawod	95%	0%	5%
Sons of Egypt List	90%	5%	5%
Ashraf Rashad Othman	9%	79%	12%

The highest percentage of positive tweets goes to Diaa Eldin Dawod, followed by Sons of Egypt List , then the candidate Mahmoud Bader.
However, both the candidate Ashraf Rashad Othman and Future of a Nation Party received the fewest and lowest positive tweets, due to voters' dissatisfaction with their performance.

- Support Vector Machine (SVM): is a group of supervised learning techniques used for both classification and regression tasks. Simply put, given a set of training examples, each labeled as belonging to one of two categories, the SVM training algorithm constructs a model that can predict the category of new examples. Essentially, the SVM model represents examples as points in space, mapped in a way that separates the two categories with the widest possible margin between them [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].

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

Candidate	Accuracy
Mahmoud Bader	0.875
Independent Alliance List	0.833
Call of Egypt	0.545
Mourtda mansour	0.777
Haithem Alhariri	0.888
Ahmed tantawy	0.625
National List for Egypt	0.601
Future of a Nation Party	0.727
Diaa Eldin Dawod	1.000
Sons of Egypt List	1.000
Ashraf Rashad Othman	0.667
Accuracy Weighted Average	**0.7723**

- Naïve Bayes (NB) : is a powerful classification algorithm commonly employed for sentiment analysis and document classification. As a probabilistic model, the Naïve Bayes classifier utilizes the joint probabilities of terms and their respective categories to determine the likelihood of categories based on given test data [24].

$$P(c|x) = P(c)P(x|c) / P(x) \quad (5)$$

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].

Table 3: Accuracy and Weighted Average of Naive Bayes (NB) for every Candidate

Candidate	Accuracy
Mahmoud Bader	0.875
Independent Alliance List	0.833
Call of Egypt	0.727
Mourtda mansour	0.777
Haithem Alhariri	0.889
Ahmed tantawy	0.625
National List for Egypt	0.600
Future of a Nation Party	0.727
Diaa Eldin Dawod	1.000
Sons of Egypt List	1.000
Ashraf Rashad Othman	0.666
Accuracy Weighted Average	**0.8285**

- Decision Tree (DT) :

A decision tree algorithm is a supervised learning technique used in data mining and machine learning. They work by splitting a dataset into smaller and smaller subsets while associating each subset with a target value [23]. This splitting is done by asking questions about the attributes of the data.

$$\text{Entropy } (S) = -\sum_{CEC} P(C) Log_2 p(c) \quad (6)$$

- 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].

Table 4. Accuracy and Weighted Average of Decision Tree for every Candidate

Candidate	Accuracy
Mahmoud Bader	0.875
Independent Alliance List	0.667
Call of Egypt	0.545
Mourtda mansour	0.777
Haithem Alhariri	0.888
Ahmed tantawy	0.375
National List for Egypt	0.400
Future of a Nation Party	0.454
Diaa Eldin Dawod	1.000
Sons of Egypt List	1.000
Ashraf Rashad Othman	0.666
AccuracyWeighted Average	**0.703**

- Random Forest (RF) :

Random forest trains multiple decision trees on different subsets of the original dataset and outputs the class that is the mode of the classes output by individual trees [25].

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

= 1- [(P+)2 + (P-)2] (8)

Table 5. Accuracy and Weighted Average of Random Forest for every Candidate.

Candidate	Accuracy
Mahmoud Bader	0.875
Independent Alliance List	0.833
Call of Egypt	0.545
Mourtda mansour	0.777
Haithem Alhariri	0.888
Ahmed tantawy	0.625
National List for Egypt	0.600
Future of a Nation Party	0.636
Diaa Eldin Dawod	1.000
Sons of Egypt List	1.000
Ashraf Rashad Othman	0.666
AccuracyWeighted Average	**0.752**

Table 6. Comparison of various Classification Accuracy

Classification Type	**SVM**	**NB**	**DT**	**RF**
Accuracy 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 technique compared to other algorithms, where it achieved the highest

weighted average of 82.85% , which is the highest possible accuracy.

2.5 Importance of Specialized Dictionaries in Sentiment Analysis :

Specialized lexicons are crucial for sentiment analysis, particularly within the context of the Egyptian electoral landscape. These lexicons provide a structured collection of sentiment-related vocabulary and their orientations, facilitating the categorization of tweets by sentiment. While dictionary-based methods are computationally efficient and allow for quick vocabulary expansion, they often fall short in capturing domain-specific sentiments that rely heavily on context. In contrast, corpus-based approaches utilize large datasets to determine word polarity, offering greater resilience but demanding more computational resources [6].

In analyzing sentiments in Arabic texts, lexicons are created by combining training datasets from various sources and applying preprocessing techniques to derive distinct stemmed words. Each word is assigned positive, negative, or neutral labels based on its occurrences in tweets, both with and without negations [1]. By calculating scores for each sentiment category and establishing inclusion thresholds, irrelevant words can be effectively filtered out. Additionally, emotion lexicons are developed to represent common emotions along with their corresponding polarities.

Moreover, specialized vocabularies and idiomatic expressions are crafted to capture nuanced sentiments expressed in Arabic tweets about elections. These tailored lexicons enhance the sentiment analysis process by providing context-specific terms that may not appear in standard dictionaries[4].

Overall, the importance of specialized lexicons in sentiment analysis is significant as they serve as essential tools for accurately categorizing sentiments expressed in social media. The creation of colloquial electoral lexicons tailored to specific scenarios, such as the Egyptian elections, can greatly improve the accuracy and reliability of sentiment analysis outcomes.

- The following phases of Dictionary :

1 . Sampling sentences from the relevant field.

2 . Classifying the sentences based on the strength of their sentiment.

3 . Estimating the sentiment degree of the sentences.

4 . Differentiating between important and unimportant words.

5 . Estimating the sentiment degree of each word.

```
df["word_tokens"]=df["clean_text"].apply(lambda x: nltk.word_tokenize(x))
df.head()
```

Table 7 . clean text and word tokens

Tweet	label	Score	Word tokens	Clean text
It is sad that a candidate in what is called the National List for Egypt does not know the history of Egypt and puts a picture of Nefertiti instead of Cleopatra.۱	negative	0.979237437	Something sad, a candidate, called the national list, is known.	A sad thing is being nominated called the national list known as history.
A former parliamentarian paid a bribe of 1.25 million pounds in exchange for candidacy on the national list for Egypt, but he did not get what he wanted.	neutral	0.627615213	Former parliamentarian, bribery, candidacy, the list,	Former parliament member bribed to run for the national list.
The Egypt Calling list surpasses the National List for Egypt.	neutral	0.573243439	[Excellence, list, call, the list, national]	The supremacy of the list of the national call.

The National List for Egypt sweeps with an absolute majority in Cairo and Giza.	neutral	0.8762447237968445	The national list sweeps with an absolute majority.	The national list sweeps with an absolute majority in Cairo.
Candidates for the House of Representatives from the National List for Egypt.	positive	0.985940337	Filters, national, yes, Egypt	National list candidates for Egypt .
Your voice is a trust for our future -------- Choose for yourself -------- Choose for tomorrow	positive	0.851560116	Your voice is a trust for our future; choose for yourself, for tomorrow.	Your voice is the trust of our future.

The table (7) explains the process of sentiment analysis of tweets in four main steps:

1- Select a sample of tweets to analyze.

2 - Clean the tweets by removing unimportant words to focus on meaningful words.

3 - Break down each tweet into individual words.

4- Analyze each word separately to measure the sentiment orientation (positive or negative) in it.

- Word Tokens:

Table 8. Word tokens

	Word- tokens	label	score
0	**Thing**	**negative**	**0.626930**
1	**Sad**	**negative**	**0.984947**
2	**candidate**	**neutral**	**0.845157**
3	**called**	**neutral**	**0.506489**
4	**the list**	**neutral**	**0.949459**

3. Experimental Results and Evaluation :

The obtained results show that , The tweets were evaluated using a set of algorithms such as; SVM, NB, Decision tree and Random forest.

The result showed the accuracy and weighted Average of SVM 77.23% and DT 70.3% but FR weighted average 75.2% .

The results showed the accuracy and Weighted Average of the Naive Bayes algorithm, which is considered the most famous, accurate, and fastest technique compared to other algorithms, where it achieved the highest weighted average of 82.85% , which is the highest possible accuracy but Decision tree it is considered the lowest rate in weighted average and accuracy as shown in figure (3) .

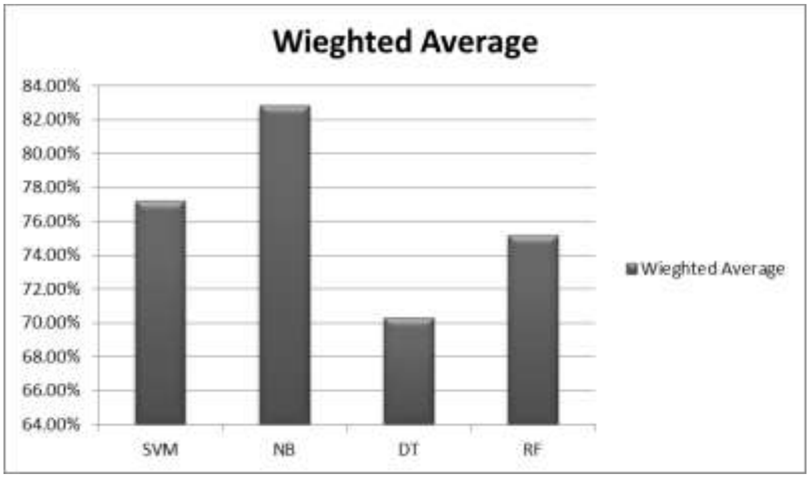

Figure 3. The result of weighted Average for various Classification

Most of the results match the actual reality, as the candidate ""Diaa Eldin Dawod" achieved a high accuracy reaching 100% in reality, where he garnered the majority of the votes in parliament, It is also the " Sons of Egypt List " list achieved high accuracy reaching 100%, which is consistent with reality, as it dominated the votes in many areas.

The integration of specialized dictionaries was found to boost the precision of sentiment analysis outcomes. Recognizing the necessity for a specialized electoral lexicon, efforts were directed towards creating a dictionary that could enhance sentiment analysis within the context of Egyptian elections. The development of such a resource is expected to refine sentiment classification and capture the subtleties inherent in public opinion. In summary, the findings and discussions underscored the importance of machine learning approaches in sentiment analysis and emphasized the potential for future exploration in crafting tailored dictionaries for electoral sentiment analysis as shown in tables (9,10) for positive and negative dictionary for words related to the Egyptian elections.

Table 9. Positive dictionary

Word - tokens	**label**	**score**
Excellence	positive	0.954451978
You will be honored	positive	0.959191263
Your voice	positive	0.901723921
Trust	positive	0.76950103
For our future	positive	0.814585745
They love	positive	0.936541855
Homeland	positive	0.700829148
Success	positive	0.941472232
Success	positive	0.969972908

Table 10 . Negative dictionary

Word - tokens	label	score
Something	negative	0.626930475
Sad	negative	0.984946609
Substitute	negative	0.927833021
Bribe	negative	0.923000991

Failure	negative	0.947804451
Terrible	negative	0.732258618
Their money	negative	0.98791182
And I objected	negative	0.56041944
They excluded me	negative	0.859924316
And I will expose them	negative	0.988466501
Scandal	negative	0.994906843
Bribes	negative	0.975153387
And vouchers	negative	0.658040881
And corruption	negative	0.981457531

4. CONCLUSION AND FUTURE WORK :

The study underscores the significance of sentiment analysis in interpreting public opinion, especially on social media platforms like Twitter. By employing machine learning techniques such as Support Vector Machines, Naive Bayes, Random Forests, and Decision Trees, researchers can extract valuable insights from the sentiments expressed in tweets concerning the Egyptian Parliament elections.

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

Most of the results match the actual reality, as the candidate "Diaa Eldin Dawod" achieved a high accuracy reaching 100% in reality, where he garnered the majority of the votes in parliament, It is also the " Sons of Egypt List " list achieved high accuracy reaching 100%, which is consistent with reality, as it dominated the votes in many areas.

Creating a colloquial electoral dictionary specifically designed for the Egyptian context could greatly improve the accuracy of sentiment analysis by capturing the unique nuances and contexts associated with local elections. The development of this dictionary would involve compiling terms, phrases, and expressions commonly used by Egyptians during election periods.

This study, after the model achieved results consistent with reality, helps guide government efforts. Predicting outcomes enhances integrity and increases transparency by providing accurate and objective predictions, which limits the spread of rumors or media manipulation, thereby ensuring the smooth conduct of the electoral process. The model assists decision-makers in formulating long-term strategies to understand voting patterns and voter tendencies, which helps in shaping public policies aligned with citizens' aspirations. This, in turn, fosters political participation as citizens' awareness of their role improves, encouraging them to engage actively.

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.

5. 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] Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial intelligence review, 1-57.
[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. 71-77). 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), 55-70
[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).