International Journal of Advanced Science and Computer Applications (2023); 2(2):
DOI: 10.47679/ijasca.v2i2.61
https://ijasca.org/index.php/ijasca

RESEARCH ARTICLE

Proposed Machine learning model for predicting Egyptian Parliament Election Results

Doaa Alkhiary, Samir Abu El Fotoh Saleh , Mohamd Ebrahim Marie

Published online: 14 September 2023

Abstract

Political life and election have become one of the most important comments on social media sites. Governments have shown a keen interest in predicting the results of elections, whether presidential or parliamentary. The purpose of this study is to predict the results of the Egyptian Parliament elections using sentiment analysis, specifically Support Vector Machines (SVM), Naive Bayes, Decision Trees, and Random Forests in the context of machine learning. In this study, a sentiment analysis approach is employed to analyze public sentiment towards political parties and candidates leading up to Parliament elections. The sentiment analysis techniques are utilized to classify sentiment from textual data collected from Tweeter; Data were obtained in November 2020 before and during election days. The study utilizes a machine learning framework to train and test the models using a labeled dataset of sentiment-labeled political texts. The findings of this study reveal that sentiment analysis using machine learning can effectively predict the results of Parliament elections. The accuracy and performance of each technique are evaluated and compared to determine the most accurate and reliable predictor of election outcomes. This study contributes to the existing literature by applying sentiment analysis techniques to predict Parliament election results. The use of machine learning algorithms in combination with sentiment analysis, offers a novel approach to election forecasting. The findings of this study can be valuable for political analysts, election strategists, and policymakers seeking to understand public sentiment and predict election outcomes accurately.

Keyword: Sentiment analysis; Twitter; machine learning techniques

Introduction

Background on political life and elections on social media sites

Machine learning plays a significant role in sentiment analysis, particularly in the context of elections. By analyzing social media data, machine learning algorithms can determine the sentiment of the public towards political candidates and predict election outcomes. Social media platforms such, as Twitter, Instagram and Facebook have gained popularity as channels for expressing opinions and discussing events. They have also played a role in shaping campaigns and elections. We witnessed the impact of media in politics during the 2008 US election when Barack Obama effectively utilized Twitter for his campaign. This influence was further highlighted during the victory of Donald Trump in the 2016 US election leaving astonished. In one study, CNN was used to analyze the sentiments of the public towards candidates at both national and state levels. The study found that while tweet volume was not a good predictor of election outcomes at the national level, sentiment analysis provided accurate results. At the state level, neither tweet volume nor sentiment analysis predicted positive election results [1], [2].

Researchers have also focused on improving the accuracy of sentiment analysis techniques for election prediction. For example, Livne et al. (2011) [3] studied the usage pattern of Twitter by candidates in US midterm elections and found that analyzing the graph structure and content produced by users can improve the accuracy of election prediction.

Traditional lexicon-based approaches have been commonly used in election prediction, but machine learning approaches are emerging as a promising alternative. Deep learning techniques have shown promising results in predicting election outcomes using social media data [4], [5], [2].

The power of social media data and intelligent computational approaches is evident in predicting positive election results. Many studies have successfully predicted positive outcomes using social media data and computational models. Machine learning algorithms have also been applied to predict general election results in various countries. Twitter has proven to be a valuable source of data for sentiment analysis during elections due to its vast amount of user-generated content [1]. However, there are still challenges that need to be addressed in this field. The literature study reveals that there is a need for valid and acceptable predictive models for election result

Helwan university, Egypt

*) corresponding author

Email: Doaa.Mohamed021@commerce.helwan.edu.eg

Received 11 May 2023 | Received in revised form 10 Juli 2023 | Accepted 12 August 2023
Available online 14 September 2023

prediction. While sentiment analysis has shown promise, more work needs to be done to improve its accuracy and reliability [6].

Various studies have been conducted to analyze sentiments expressed on media platforms during elections offering insights into their correlation with results. Similar analyses have been performed for elections held in countries like India, Iran, Singapore and Colombia [7]. The art of sentiment analysis holds potential in uncovering sentiments regarding candidates in the 2020 US election amidst concerns surrounding election manipulation and postal voting [2]. It has proven valuable in monitoring campaigns in both Italy and the United States by providing researchers with insights into public attitudes and behaviors [8]. Machine learning models like support vector machines and Naive Bayes have demonstrated their effectiveness when it comes to sentiment analysis tasks by utilizing Twitter data to accurately predict election outcomes [9]. Social media platforms provide real time data that can be analyzed to comprehend opinion trends and even predict election results. By harnessing machine learning techniques researchers can gain insights, into sentiments and behaviors exhibited during elections [6], [10].

Overall, machine learning plays a crucial role in sentiment analysis for election prediction. By analyzing social media data and applying machine learning algorithms, researchers can gain insights into public sentiment and make accurate predictions about election outcomes. The emerging power of social media data and intelligent computational approaches highlights the significance of machine learning in sentiment analysis for elections [10], [11].

Importance of predicting election results

Predicting election results is of utmost importance in understanding the sentiments and preferences of the public. With the rise of social media platforms like Twitter, Instagram, and Facebook, individuals now have a powerful way to express their opinions and sentiments about political events, campaigns, and elections. This wealth of real-time data provides a valuable resource for conducting sentiment analysis and predicting election outcomes [4], [12].

One notable example is the U.S. presidential elections of 2016 and 2020. These elections witnessed the emergence of controversies, delays in announcing results, and allegations of rigging. Social media sentiment analysis played a crucial role in capturing public views before, during, and after these elections. By analyzing Twitter data using techniques like TF-IDF and Naive Bayes Classifier, researchers were able to extract meaningful insights from the masses' opinions [2].

The results of these sentiment analyses revealed interesting patterns. In most cases, there was a strong correlation between the sentiment expressed on social media platforms and the actual election outcomes. This suggests that studying social media sentiment can provide valuable insights into public opinion and preferences. Additionally, analyzing pre- and post-election sentiments shed light on any shifts or changes in public sentiment over time [7], [13]. Moreover, sentiment analysis has proven useful in predicting general election results in various countries around the world. For example, studies conducted on elections in Turkey demonstrated that Twitter data could be used to predict election outcomes with a high degree of accuracy. By applying machine learning techniques such as sentiment analysis and natural language processing to analyze Twitter data related to political candidates and parties, researchers were able to make accurate predictions about election results [6].

Machine learning models have been particularly effective in sentiment analysis tasks. Models like support vector machines (SVM), Naive Bayes classifiers, random forests, and logistic regression have been widely used to analyze sentiments expressed on social media platforms during elections. These models leverage large amounts of data to make predictions about public opinion. However, challenges remain in optimizing sentiment analysis for election prediction. Researchers continue to explore the best methods and techniques to ensure accurate predictions. Deep learning models, such as residual long short-term memory (LSTM), have shown promise in improving sentiment analysis accuracy but require larger datasets for training [14].

The emergence of social media data and sentiment analysis has provided researchers with a new avenue for predicting election outcomes. Numerous studies have been conducted to explore the potential of using social media metrics, particularly on platforms like Twitter, to gauge public sentiment and forecast election results. However, despite the positive results reported in some studies, there are several challenges and limitations that need to be considered when using sentiment analysis for election predictions [6].

One of the main challenges is the lack of consistency in findings across different studies. While some research has shown a direct correlation between the volume and sentiment of Twitter chatter and future electoral results, other studies have contradicted these findings. For example, a study on the 2010 US congressional elections found no correlation between the analysis results based on Twitter data and the actual electoral outcomes. This inconsistency raises questions about the reliability and accuracy of using social media data as a predictive tool for elections [15].

Another limitation is the difficulty in determining what principle enables positive election predictions based on social media metrics. Despite positive results being reported in some studies, there is still a lack of understanding about why these predictions are accurate or how social media data can effectively capture public opinion. Without a clear explanation or model explaining the predictive power of social media, it becomes challenging to accept predictions based solely on social media data. Additionally, there are limitations related to data collection and analysis techniques. Many studies rely on collecting Twitter data as a primary source for sentiment analysis. However, this approach may not provide a comprehensive view of public sentiment due to potential biases in social media usage patterns or limited representation of certain demographics. Moreover, determining sentiment from text data can be challenging due to linguistic nuances, sarcasm, irony, and context-dependent interpretations [4].

Furthermore, while machine learning techniques have been widely used for sentiment analysis, they also come with their own set of limitations. The effectiveness of machine learning models depends on the quality and size of the training data, as well as the chosen features and algorithms. Additionally, these models may struggle to handle evolving political landscapes, as public sentiment can change rapidly during election campaigns. Ethical concerns also need to be addressed when using sentiment analysis for election predictions. It is crucial to ensure that data collection methods are transparent and ethical, with safeguards in place to protect against manipulation or misinformation campaigns that can distort public sentiment. Moreover, while sentiment analysis using social media data holds promise for predicting election outcomes, there are significant challenges and limitations that need to be considered. Inconsistent findings, a lack of understanding about the underlying principles of accurate predictions, data collection biases, linguistic complexities, limitations of machine learning models, and ethical concerns all contribute to the complexity of using sentiment analysis for election predictions. Future research should focus on addressing these challenges and developing robust methodologies that can provide more reliable and accurate predictions in this field [10].

In conclusion, sentiment analysis is a powerful tool for predicting election results. The abundance of real-time data on social media platforms allows researchers to analyze public sentiments and preferences towards political candidates and parties. By leveraging machine learning models and techniques, accurate predictions about election outcomes can be made. As social media continues to play an influential role in political campaigns and elections, sentiment analysis will remain a valuable resource for understanding public opinion and making informed predictions [10], [8] as shown in figure 1.

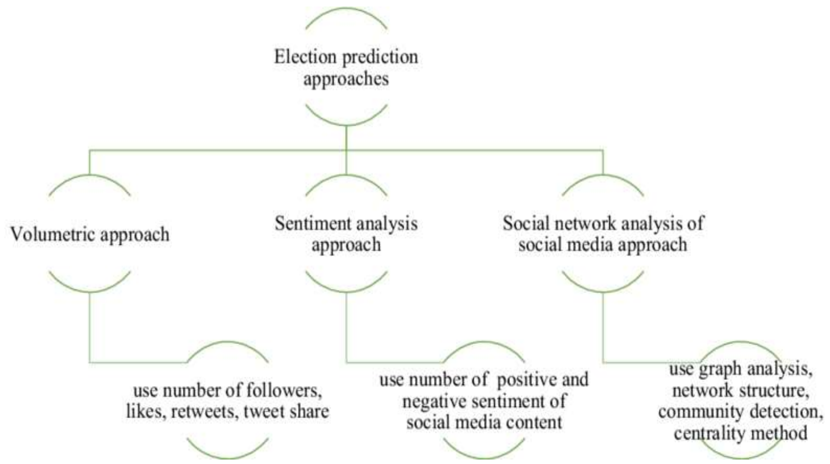

Fig 1. Election prediction approaches

Related work

Several studies have explored the use of sentiment analysis in predicting election outcomes through the analysis of social media data. Jaidka et al. (2019) [16] used volumetric, sentiment, and social network analysis approaches to predict general election outcomes in Malaysia, Pakistan, and India. Their study accurately predicted the outcomes for India and Pakistan but was incorrect for Malaysia. The study also highlighted the importance of recent Twitter posts and the combination of multiple approaches for accurate predictions [15]. In addition, Xie et al. (2018) [17] employed data from various online sources to predict the Taiwan general election in 2016. They utilized a combination of Kalman filter and event study methods for their prediction task, suggesting that this approach could be a fundamental tool for political vote analytics. The study also emphasized that events occurring during election time can significantly affect results and found that Twitter plays a crucial role in political communication [8]. Moreover, Heredia et al. (2018) [18], Bilal et al. (2018) [19], and Bose et al. (2019) [20] introduced deep learning techniques, including the bidirectional encoder from transformers (BERT), as emerging approaches in sentiment analysis for election prediction [6].

Livne et al. (2011) [3] focused on analyzing Twitter usage patterns by candidates during the 2010 US midterm elections. Their research demonstrated that analyzing both content and graph structure produced by users could improve the accuracy of election prediction models [7], [2]. In addition, in another study, authors used supervised machine learning algorithms to classify tweets into positive, negative, and neutral sentiments to study and forecast Pakistan's general election. The findings indicated that social media content could serve as a useful indicator for identifying political behavior [21].

The literature review highlights that most studies utilize Twitter as a corpus for predicting election results, emphasizing sentiments-oriented content from social media platforms like Twitter for analysis purposes [1]. Although traditional lexicon-based approaches have been widely used, recent studies have shown an increasing interest in machine learning approaches, especially deep learning techniques. However, there is still a need for a valid and acceptable predictive model in this field.

Method

The methodology using for Building Machine Learning Classifies Consists of four steps (Fig 2):

1- Data collection from twitter.
2- Data pre- processing.
3- Building Model.
4- Classification (Positive, Negative, Neutral).

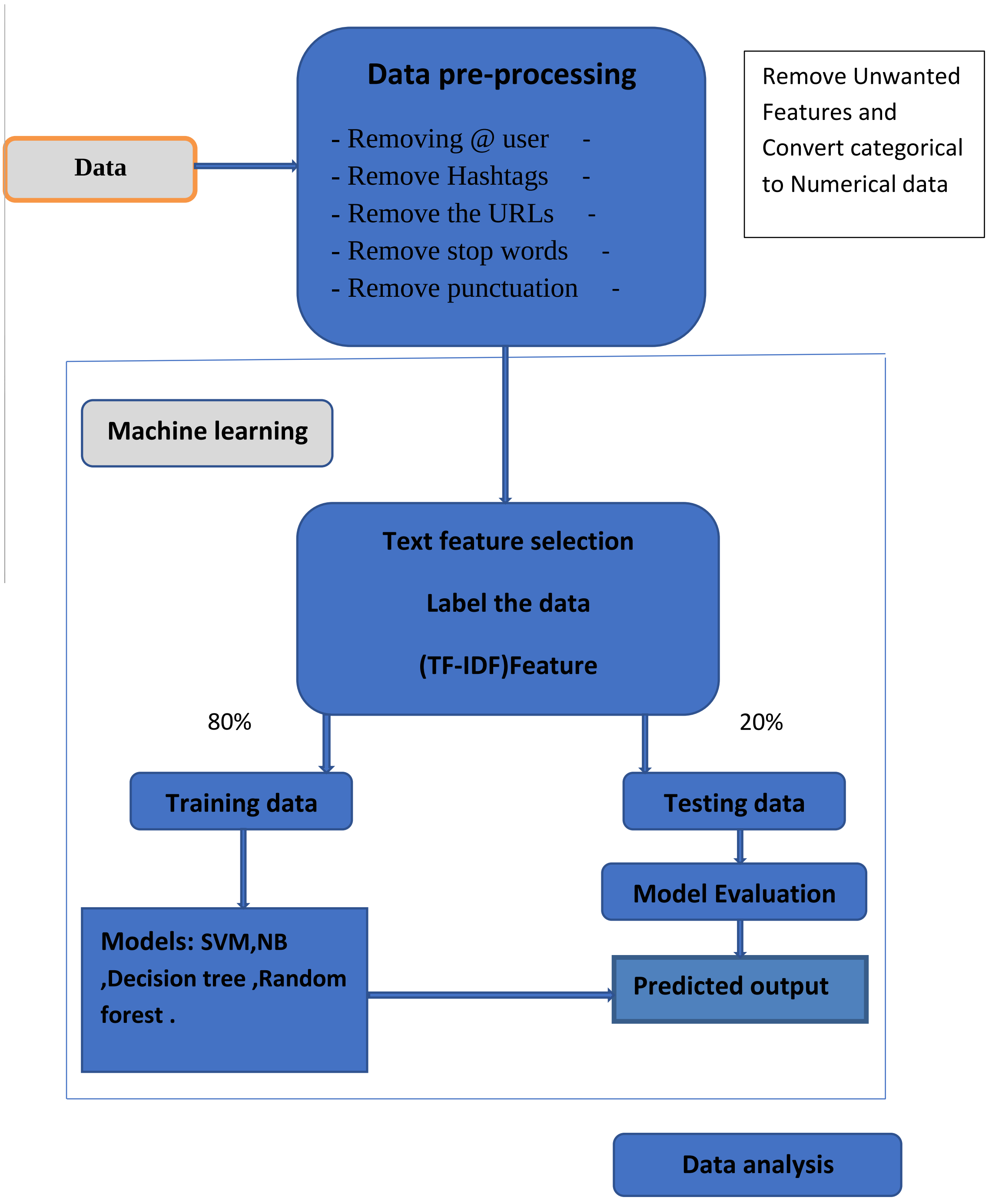

Fig 2. Model steps

Data collection

Data about political opinions in the 2020 parliamentary elections in Egypt before and during the election days. In addition, this data consists of thousands of tweets. This data consists of thousands of tweets. While collecting data, Twitter disconnected the interface connection API, meaning no more data was obtained by searching for the same keywords and there was more repetition in the tweets therefore the researcher resorted beside those data to manual research, where all the hashtags that were created before and during the elections were covered for that after cleaning and filtering the tweets the data volume was smaller.

The data set was divided into training and testing dataset: General data set consists of 2600 tweets, 2080 training data and 520 testing data.

Data pre-processing

(1). Removing Twitter Handles (@user)

There are various Twitter handles (@users) in the tweets, which is how Twitter users are recognized. Moreover, we delete all these Twitter handles from the database. The first is a combined train and test set for simplicity. This saves time and effort by not having to repeat the same actions.

(2). Remove Hashtags

Text = text. Replace ("#", "") #removing hashtags.

(3). Remove the URLs from the Text

A python code like given below is used to delete the URLs from a text, as displayed.

```
def remove_url(row):
    txt = (['tidy_text']). ('https')[0]
return txt
['tidy_text'] = data. (remove_url, axis = 1).
```

(4). Stop-Words Removal

A stop-word is a term in English that conveys little or no meaningful information. Text preparation necessitates their removal. Every language's stop-words are listed on nltk.

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

(5). Remove punctuations:

["ى","""","-","''"," يا "," و ","،",".","/","-","_","ة","آ","إ","أ","\\",'\n', '\t','"','?','؟','!'].

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. As shown in Table 1 and Table 2.

Table 1. labeled output data for negative

	Tweet	Label	Score
0	...شئ محزن ان مرشحا ضمن ما تسمى القائمة الوطنية من اجل مصر لا يعرف تاريخ مصر	Negative	0.979237
18	...يبقى غيرهم غير وطنى لما قايمة انتخابية يبقى اسمها القائمة الوطنية	Negative	0.906505
21	القائمة الوطنية فشلت فشل زريع فى محافظة البحيرة	Negative	0.968898
22	... حد يفهمنا ايه القائمة الوطنية من اجل مصر دى وايه الحزب اللى مكان الحزب الوطنى	Negative	0.539732
25	... بعد ما عرفت مصدر أموالهم واعترضت ف استبعدونى من الترشح	Negative	0.989320
...	...	...	...
534	اكبر تجار مخدرات فى جمهورية مصر العربية	Negative	0.991804
536	الصوت وصل 400 جنيه والحسابة بتحسب	Negative	0.434879
537	بيقولك بيدى الناس 100 جنيه علشان ينتخبوه	Negative	0.868884
538	حد يروح يساومهم وياخد سعر احسن	Negative	0.706155
539	بخيل اوى انت يا اشرف	Negative	0.987943

Table 2. Labeled output data for positive

	Tweet	Label	Score
5	... مرشحات لمجلس النواب عن القايمة الوطنية من اجل مصر حتشرفوا المجلس يا عظيمات	Positive	0.985940
6	لقايمة الوطنية من اجل مصر غرب الدلتا الاسكندرية تكتسح	Positive	0.571617
7	---- صوتك امانة لمستقبلنا -------- اختر لنفسك	Positive	0.851560
9	كل الدعم والتأييد للقائمة الوطنية من اجل مصر	Positive	0.979276
13	صوتك بيكتب مستقبل مصر - انزل - شارك	Positive	0.760466
524	احمد طنطاوى المحترم	Positive	0.831132
526	انت انسان محترم واحسن واحد فى كفر الشيخ	Positive	0.990431
527	قمة الأدب والإحترام والاخلاص	Positive	0.988700
540	حملة تستحق التضامن وبقوة انا متضامنة شكلا وموضوعا	Positive	0.918345

Sentiment classification

The data is divided into training and testing dataset. The training data set used to build the classification models based on SVM, Naive Bayes, Random Forest and decision tree classifiers. Moreover, the data classified based on its polarity to positive and negative classes .The testing data set is used to predict the polarity of the tweets.

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

This section covers existing approaches to text classification:

SVM, NB, Diction tree and Random Forest.

Support vector machine (SVM): is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyper plane that best separates the data into different classes. The algorithm aims to maximize the margin between the hyper plane and the closest data points, known as support vectors, in order to achieve the best generalization performance.

Mathematically, SVMs learn the sign function: F(x) =sign (wx+b) [22].

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.

Table 3. The result of SVM Algorithm for every candidate

Class	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

Niave Bayes (NB) is: a simple and widely used supervised machine learning algorithm based on applying Bayes' theorem with strong independence assumptions.

The function of the Naive Bayes algorithm can be defined through Bayes' theorem:

$$P(c|x) = P(c)P(x|c) / P(x)$$

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 [23].

Table 4. The result of Niave Bayes Algorithm for every candidate

Class	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

Decision Tree: is a supervised learning technique used in data mining and machine learning. It uses a decision tree as a predictive model to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves).

They work by splitting a dataset into smaller and smaller subsets while associating each subset with a target value. This splitting is done by asking questions about the attributes of the data.

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

- 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 [24].

Table 5. The result of Decision Tree Algorithm for every candidate

Class	Precision	Recall	F measure	Accuracy
محمود بدر	0.880	0.880	0.875	0.875
قائمة تحالف المستقلين	0.670	0.670	0.667	0.667
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	0.380	0.380	0.375	0.375

القائمة الوطنية	0.400	0.400	0.400	0.400
حزب مستقبل وطن	0.450	0.450	0.454	0.454
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

Random Forest: is an ensemble learning method that combines multiple decision trees to make more accurate predictions. It utilizes the concept of randomization to create a diverse set of decision trees and aggregates their outputs to reach a final prediction. The equation for the random forest algorithm can be expressed as follows:

$$\begin{aligned} Gini\ Index &= 1 - \sum_{i=1}^{n}\left(P_i\right)^2 \\ &= 1 - [(P_+)^2 + (P_-)^2] \end{aligned}$$

This algorithm was introduced by Leo Breiman in 2001 [25], and it has been widely used in various domains such as classification, regression, and feature selection due to its robustness and versatility [26].

Table 6. The result of Random Forest Algorithm for every candidate

Class	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

Results and Discussion

There are different measures that can used to measure classification accuracy. The basic measures are accuracy, precision, recall and F-measure.

For the evaluation of classification results, well-known measures were addressed. The basic measurements are the counts of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) with respect to each class of each instance. These depend on whether the class predicted by the classifier matches the expected prediction [5].

True positive (TP): refers to positive instances that are correctly labeled.

False Negative (FN): are the positive instances that are incorrectly labeled.

False Positive (FP): are the negative instances that are incorrectly labeled.

True negative (TN): refers to negative instances that are correctly labeled.

The most basic measure is accuracy (Acc). The accuracy can be calculated by a simplified equation below [27].

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

Accuracy is a good measure when classes are distributed uniformly in the collection. However, as class imbalances grow more pronounced, high accuracy might be attained by a classifier that has a bias towards the majority class. In addition, precision and recall are often used as an alternative, providing a more detailed analysis of the classifier's behavior with respect to each class. The precision measures the relative frequency of correctly classified examples that were predicted to belong to c as the equation below [11].

$$\text{Precision} = \frac{number\ of\ true\ positives}{number\ of\ true\ positive + false\ positive}$$

Recall: is the percentage of the total sentences for the given topic that are correctly classified. It can calculate as follows:

$$\text{Recall} = \frac{true\ positives}{true\ positives + false\ negatives}$$

The harmonic mean of precision and recall is called the F-measure.

It is calculated as the equation below [28].

$$F = 2 \times \frac{precision\ x\ recall}{precision + recall}$$

In this study, the four measures were calculated to evaluate the correctness of classifying tweets as positive or negative class.

Here are the results of the model evaluation as shown in the following Table:

Table 7. The result of the model evaluation

Precision	Recall	Accuracy	F-Measure
0.937037037037	0.9504195270785	0.94545454545454	0.9417560417560417

From table 7 above, good results for Recall, Accuracy and F-measure, which equal to 95.0%, 94.5%, 94.2%; respectively.

The confusion matrix

A confusion matrix provides a detailed breakdown of the performance of a classification model, especially regarding the types of errors it makes. Therefore, it allows quick assessment of model accuracy.

Fig 3. The confusion matrix

In the above Figure (Fig 3), to evaluate the model on random sample of the data has been generated 23 negative tweets; the model classified 22 tweets as negative and only one tweet incorrectly as neutral. Also has been generated 19 positive tweets, the model classified 17 tweets as positive and 2 tweets incorrectly. As for neutral tweets, and generate 13 neutral tweets and the model classified all 13 as neutral. Furthermore, the obtained results showed an error factor of only three mistakes. Thus, the model confirmed minimizing errors and achieving the highest accuracy results. In addition, the obtained results showed that, the accuracy of the Naive Bayes algorithm, which is considered the most famous, accurate, and fastest technique compared to other algorithms. Most of the results match the actual reality, as the candidate "ضياء الدين داوود" achieved a high accuracy reaching 100% in reality, where he garnered the majority of the votes in parliament. It is also the " قائمة أبناء مصر" list achieved high accuracy reaching 100%, which is consistent with reality, as it dominated the votes in many areas.

Conclusion and future work

Sentiment analysis in election prediction using machine learning has shown great potential for understanding public opinion and predicting election outcomes. This research uses Twitter to gather data to analyze the citizens' sentiments towards the candidates in the 2020 Egyptian parliamentary elections, whether they are positive or negative, and the study used machine learning techniques SVM, NB, decision tree, Random Forest. Moreover, the obtained results showed that, the accuracy of the model that was built, reaching a model evaluation of 94.5%, and the accuracy of the Naive Bayes algorithm results, which is the best in terms of accuracy and speed, and its conformity to reality compared to other algorithms. In addition, the results of both the candidate" ضياء الدين داوود" and "قائمة ابناء مصر" were among the best results, reaching 100% accuracy and conformity to reality. However, there are several areas that require further research to enhance the accuracy, reliability, and ethical considerations of this approach. By integrating multiple data sources, addressing causality and influence, improving machine learning techniques, considering cross-cultural differences, and conducting longitudinal analysis, researchers can contribute to the advancement of this field and provide more robust predictions for future elections.

References:

[1] Ali, H., Farman, H., Yar, H., Khan, Z., Habib, S., & Ammar, A. (2022) Deep learning-based election results prediction using Twitter activity. Soft Computing, 26(16), 7535-7543.

[2] Alvi, Q., Ali, S.F., Ahmed, S.B., Khan, N.A., Javed, M., & Nobanee, H. (2023) On the frontiers of Twitter data and sentiment analysis in election prediction: a review. Peer J. Computer Science, 9, e1517.

[3] 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. 201-208).

[4] Chauhan, P., Sharma, N., & Sikka, G. (2021) The emergence of social media data and sentiment analysis in election prediction. Journal of Ambient Intelligence and Humanized Computing, 12, 2601-2627.

[5] Qorib, M., Gizaw, R.S., & Kim, J. (2023). Impact of Sentiment Analysis for the 2020 US Presidential Election on Social Media Data. In Proceedings of the 2023 8th. International Conference on Machine Learning Technologies (pp. 28-34).

[6] Baker Al Barghuthi, N., & E. Said, H. (2020) Sentiment analysis on predicting presidential election: Twitter used case. In Intelligent Computing Systems: Third International Symposium, ISICS 2020, Sharjah, United Arab Emirates, March 18–19, 2020, Proceedings 3 (pp. 105-117). Springer International Publishing.

[7] Chaudhry, H.N., Javed, Y., Kulsoom, F., Mehmood, Z., Khan, Z.I., Shoaib, U., & Janjua, S.H. (2021) Sentiment analysis of before and after elections: Twitter data of us election 2020. Electronics, 10(17), 2082.

[8] Nausheen, F., & Begum, S.H. (2018) Sentiment analysis to predict election results using Python. In 2018 2nd international conference on inventive systems and control (ICISC) (pp. 1259-1262). IEEE.

[9] Khan, A., Zhang, H., Boudjellal, N., Ahmad, A., Shang, J., Dai, L., & Hayat, B. (2021) Election prediction on twitter: a systematic mapping study. Complexity, 1-27.

[10] Rizk, R., Rizk, D., Rizk, F., & Hsu, S. (2023) 280 characters to the White House: predicting 2020 US presidential elections from twitter data. Computational and Mathematical Organization Theory, 1-28.

Olabanjo, O., Wusu, A., Asokere, M., Padonu, R., Olabanjo, O., Ojo, O., & Aribisala, B. (2022). From Twitter to Aso-Rock: A natural language processing spotlight for understanding Nigeria 2023 presidential election.

[11] Budiharto, W., & Meiliana, M. (2018. Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. Journal of Big data, 5(1), 1-10.

[12] Oyewola, D.O., Oladimeji, L.A., Julius, S.O., Kachalla, L.B., & Dada, E.G. (2023) Optimizing sentiment analysis of Nigerian 2023 presidential election using two-stage residual long short-term memory. Heliyon, 9(4) - PMC.

[13] Guellil, I., Saâdane, H., Azouaou, F., Gueni, B., & Nouvel, D. . (2019). Arabic natural language processing: an overview. . Journal of King Saud University-Computer and Information Sciences.

[14] Ceron, A., Curini, L., & Iacus, S. M. (2015) Using sentiment analysis to monitor electoral campaigns: Method matters—evidence from the United States and Italy. Social Science Computer Review, 33(1), 3-20.

[15] Jaidka, K., Ahmed, S., Skoric, M., & Hilbert, M. (2019) Predicting elections from social media: a three-country, three-method comparative study. Asian Journal of Communication, 29(3), 252-273. Qorib, M., Gizaw, R. S., & Kim, J. (2023), March.

[16] Xie, Z., Liu, G., Wu, J., & Tan, Y. (2018) Social media would not lie: Prediction of the 2016 Taiwan election via online heterogeneous data. arXiv preprint arXiv:1803.08010.

[17] Heredia, B., Prusa, J.D., & Khoshgoftaar, T.M. (2018) Location-based twitter sentiment analysis for predicting the US 2016 presidential election. In The Thirty-First International Flairs Conference.

[18] Bilal, U., Knapp, E.A., & Cooper, R.S. (2018) Swing voting in the 2016 presidential election in counties where midlife mortality has been rising in white non-Hispanic Americans. Social Science & Medicine, 197, 33-38.

[19] Bose, R., Dey, R.K., Roy, S., & Sarddar, D. (2019) Analyzing political sentiment using Twitter data. In Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2018, Volume 2 (pp. 427-436). Springer Singapore.

[20] Gorodnichenko, Y., Pham, T., & Talavera, O. (2018). Social media, sentiment and public opinions: Evidence from #Brexit and #USElection: National Bureau of Economic Research.

[21] B. Liu. (2012) “Sentiment analysis and opinion mining”,Synth. Lect. Hum. Lang. Technol., vol. 5, no. 1, pp. 1-167.

[22] Elhawary, M., Elfeky, M. (2010) Mining Arabic Business Review. In Proceedings of the 20 10. IEEE International Conference on Data Mining Workshops. 1108-1113.

[23] Kotsiantis, S.B. (2013) Decision trees: A recent overview. Artificial Intelligence Review, 39(4), 261-283.

[24] Breiman, L. (2001) Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical Science, 16(3), 199-231.

[25] Liaw, A., & Wiener, M. (2002) Classification and regression by random Forest. R news, 2(3), 18-22.

[26] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 system demonstrations (pp. 115-120).

[27] Kanaan, G., Al-Shalabi, R., Ghawanmeh, S., and Al Ma’adeed, H. 2009.A Comparison of Text-Classification Techniques Applied to Arabic Text, Journal of the American society for information science and technology, 60 (9) 1836-1844.