# A Study on Election Prediction using Machine Learning Techniques

Karthick Myilvahanan
Department of Information Science and Engineering
*New Horizon College of Engineering*
Bengaluru, India
kmvphd@pm.me

Yashas P
Department of Information Science and Engineering
*New Horizon College of Engineering*
Bengaluru, India
yashaspal2562@gmail.com

Sameer Pasha
Department of Information Science and Engineering
*New Horizon College of Engineering*
Bengaluru, India
sameerkhan7452@gmail.com

Mohammed Ismail
Department of Information Science and Engineering
*New Horizon College of Engineering*
Bengaluru, India
ismail.adnan012@gmail.com

Vimjam Tharun
Department of Information Science and Engineering
*New Horizon College of Engineering*
Bengaluru, India
vimjamtharun@gmail.com

*Abstract*—**The article that follows demonstrates how to build new characteristics from pre-existing ones and forecast election outcomes using a popular machine learning method. However, this paper has observed a method of leveraging social media to anticipate results that is not more precise in the real world, due to the manipulation and usage of bots to quickly increase in popularity, so that the accuracy has gone down. This notepaper shows that a machine learning supervised learning technique which uses the most recent Indian electoral datum for forecasting the overall outcomes of national elections as well as many local results. The findings imply that the forecasting outcomes are close to the actual outcome with more accurate results when compared to the other methodologies. Additionally, because of it is simple, reproducibility, so resistance in opposite to volume manipulation, it outperforms several state-of-the-art approaches. As far as this research aware, this is the initial attempt to verify ML prototype for the forecasting of the 2019 Indian elections.**

*Keywords— Election Result Prediction, Prediction of election results, KNN Algorithm, Feature Engineering, Training and Testing models.*

## I. INTRODUCTION

Elections are a crucial part of governance in many nations. It's the most crucial democratic instrument for connecting voters with politicians. India strives to be a leader in politics and governance as the largest democratic nation in the world. A lot of people have always been interested in making election predictions. The main motive of our research is to use the general elections in the greatest democracy in the world as a potential data mine, with over 600 million people casting ballots for 8500+ candidates over 543 constituencies. If there are no pre-poll projections, it is very challenging to forecast the mood of an election in a nation of billions. They assist individuals in determining that their votes are cast for candidates who might truly benefit from them rather than for those who have no chance of defeating the major parties. It substantially reduces vote splitting. The losing side receives a dose of reality and the chance to fight harder and defy the odds. It is understandable why promises increase as pre-poll results are released. The most recent datasets on various factors are chosen to acquire accurate findings because it has been discovered that web information, such as blog posts or social network data, such as tweets, can be altered to create inaccurate predictions. Many strategies are restricted to Twitter, preventing everyone from having access to social media platforms where they could express their opposition to one side and support for the other. When generating a prediction in the end, supervised machine learning algorithms like KNN provide outcomes that are more accurate than those produced by other methods.

## II. BACKGROUND STUDY

### A. KNN in Machine Learning

The KNN Algorithm is supervised machine learning algorithm which does not have parameters . It is summarises things according to the class of the nearest neighbours.

### B. Feature Engineering

Feature Engineering is the process of utilizing statistical or machine learning techniques to transform raw data into desired characteristics. This method's objective is to streamline and accelerate data transformations while maintaining model correctness.



Fig. 1. Working of Feature Engineering

Cleaning non-numerical data and removing helpful numerical embeddings is a key function of feature engineering. The major goal of these features is to substantially improve the model's performance, which also improves the accuracy in this ongoing research

## III. LITERATURE SURVEY

In paper [1] "predict the elections in Brazil and the US via social media data and machine learning."

The author has successfully used neural social networks, Facebook and twitter it focuses on the posts of authenticated candidates and contour in all 3 social media platforms.

In paper [2] "Probabilistic algorithms for election results prediction ".

The author used algorithms for predicting elections, naive stochastic algorithm for predicting the results of elections, and also mean analysis for election prediction.

In paper [3] "A machine learning based strategy for elections results prediction ".

The author used twitter in different way to predict the election results, author also used recursive neural networks and natural language processing to predict the results of the election.

In paper [4] "How efficiently is twitter predicting 2012 U.S. presidential election using support vector machine vi twitter and comparing against an electronic market".

This author used support vector machines (SVMs) to predict the results via twitter the frequency count distribution of the sentiment analysis on all days in a positive correlated with IEM and SVM .and also effective market hypothesis.

In paper [5] "Districts by Demographics: predicting U.S.

House of Representatives Elections using machine learning and Demographic Data ".

The author used naïve bayes, this analysis provides strong evidence that demographics and serve as a meaningful consistent factor in deciding representative elections

In paper [6] "Casing online votes: to predict online results".

In this the author uses the multinominal naive bayes classifier, this research is done to identify the best machine learning to find the polarity of the tweets and predict the sentiment of people in upcoming elections[7-15] .

## IV. PROPOSED SYSTEM

The project's primary objective is to use the features to the fullest of their efficiency. In machine learning the data that is provided plays the most important role in the working of the model to its fullest. Machine learning algorithms use mostly numerical data for the process, numerical data cannot be fully correct and up to the mark there will be some missing values, values which do not play an important role in the prediction etc, such type of data needs to be filtered out. Using the data that is most appropriate and accurate will give an increase in efficiency of the model.

The process of cleaning the dataset, make the dataset, start new characteristics of the existing ones forecasting the results using a popular machine learning classifier. This research uses Univariate choosing algorithm in order to find the important characteristics and delete others. It will be using Select well known from Scikit library using chi-square test to find the featured importance.

The algorithm is a Non parametric Supervised Learning Classifier. It uses proximity to make Classification or Predict the data point to which group it belongs to.
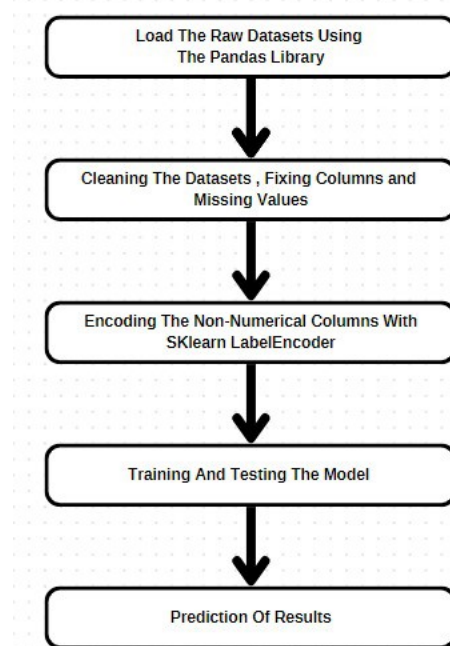


Fig. 2. Steps in Predicting election results

The steps are as follows :-

A. Libraries Imported

This research prefers to use panda's library in order to give the dataset as it does all of the data executing procedures really simple and effortless. The Python libraries used are NumPy, seaborn, pandas, matplotlib. After loading the data, the datasets contain numerical columns mostly of the columns does not contain numerical.

B. Cleaning the Dataset

Research started by correcting the wrong column field and restore of the space to underscore (' '). Although it is not required, find it bothersome when erroneous column names are used.
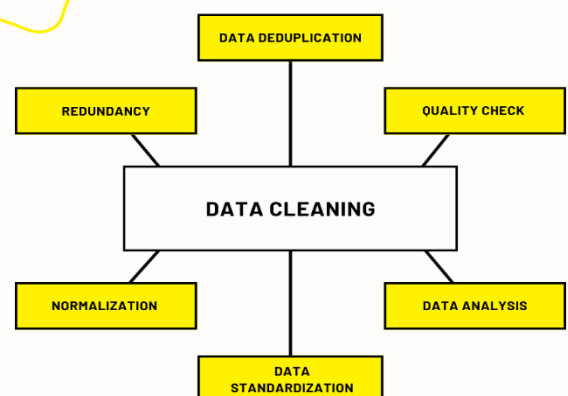


Fig. 3. Data Cleaning

## C. Searching the missing values

This paper shows that few of the row data are missing. Missing values can be handled in a variety of methods among. This is eliminated for simplicity. However, keep in mind that doing so would reduce the predictive model's accuracy.

## D. Normalization

After getting the model accuracy, now it uses normalization of the data and see that how accuracy increases. And have applied MinMaxScaler from the scikit-learn library to bring down the values to the 0–1 range.
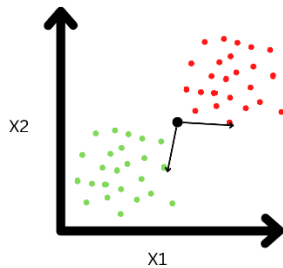


Fig. 5. Normalizing Values

## V. DISCUSSIONS

A feature engineering-based machine learning method that accurately and precisely forecasts the election outcomes. It can foretell election results using real-time datasets since it cannot be altered by bots. And to create a model which is capable of predicting the election results which can be further used just by changing the data set. This improves people's confidence in election predictions before the official results are announced.

## VI. CONCLUSIONS

A model's accuracy will increase if unimportant or unnecessary elements are removed. Indeed, feature engineering efforts will significantly improve the model's performance. The dataset will first be cleaned up and all the values will be converted to a numerical representation. Then, utilizing the already-existing features, the process of building some new features while eliminating all the less crucial ones. Applying cross-validation would allow us to increase accuracy even further and predict precise results.

## REFERENCES

[1] M. -H. Tsai, Y. Wang, M. Kwak and N. Rigole," A Machine Learning Based Strategy for Election Result Prediction," 2019 International Conference on Computational Science and Computational Intelligence (CSCI), 2019, pp. 1408-1410, doi:10.1109/CSCI49370.2019.00263.

[2] K. d. S. Brito and P. J. L. Adeodato," Predicting Brazilian and U.S. Elections with Machine Learning and Social Media Data," 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9207147.

[3] B. Richardson and D. F. Hougen," Districts by Demographics: Predicting U.S. House of Representative Elections using Machine Learning and Demographic Data," 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 2020, pp. 833-838, doi:10.1109/ICMLA51294.2020.00136.

[4] P. Juneja and U. Ojha," Casting online votes: To predict offline results using sentiment analysis by machine learning classifiers," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2017, pp. 1-6, doi: 10.1109/ICCCNT.2017.8203996.

[5] A. Attarwala, S. Dimitrov and A. Obeidi," How efficient is Twitter: Predicting 2012 U.S. presidential elections using Support Vector Machine via Twitter and comparing against Iowa Electronic Markets," 2017 Intelligent Systems Conference (IntelliSys), 2017, pp. 646-652, doi: 10.1109/IntelliSys.2017.8324363.

[6] K. B. Shibu Kumar, V. S. Devi, K. K. Rajeev and A. Bhatia," Probabilistic Algorithms for Election Result Prediction," 2014 InternationalConference on Soft Computing and Machine Intelligence, 2014, pp. 79- 82, doi: 10.1109/ISCMI.2014.12.

[7] Ali, Haider Farman, Haleem Yar, Hikmat Khan, Zahid Habib, Shabana Ammar, Adel. (2021). Deep Learning-Based Election Results Prediction Using Twitter Activity. 10.21203/rs.3.rs-839553/v1.

[8] Shobha, T., Anandhi, R. J. (2020).Robust Classifier Design with Ensemble Neural Network using Differential Evolution. International Journal of Engineering Trends and Technology, ISSN: 2231-5381,174-181.

[9] Prashanth, Paul., Prashanth, V., Prem Kumar, Kalaivanan,S. (2020).A Machine Learning Perspective towards Detecting Fake News.International Journal for Research in Applied Science and Engineering Technology,5(8),1- 5.

[10] Vibha, T. G., and S. Sivaramakrishnan. "A Survey of Deep Learning Region Proposal and Background Recognition Techniques for Moving Object Detection." Computer Networks and Inventive Communication Technologies. Springer, Singapore, 2023. 147-164.

[11] Sivaramakrishnan, S., and T. Kesavamurthy. "Identifying Cluster Head and Data Transmission Through Them for Efficient Communication in Wireless Sensor Network." Journal of Computational and Theoretical Nanoscience 14.8 (2017): 4014-4020.

[12] Sivaramakrishnan, S., et al. "Augmentation of Terahertz Communication in 6G and Its Dependency for Future State-of-the-Art Technology." Challenges and Risks Involved in Deploying 6G and NextGen Networks. IGI Global, 2022. 91-105.

[13] Devi BK, Vijayakumar V, Suseela G, Kavin BP, Sivaramakrishnan S, Rodrigues JJ. AN IMPROVED SECURITY FRAMEWORK IN HEALTH CARE USING HYBRID COMPUTING. Malaysian Journal of Computer Science. 2022 Mar 31:50-61.

[14] Harika, Vangala Ramanuja, and S. Sivaramakrishnan. "Image Overlays on a video frame Using HOG algorithm." 2020 IEEE International Conference on Advances and Developments in Electrical and Electronics Engineering (ICADEE). IEEE, 2020.

[15] Vanitha, U., Sivaramakrishnan, S., Karthika, S., & Babukarthik, R. G. (2020, December). Efficient Communication Routing Through WiMax Network During Disaster. In IOP Conference Series: Materials Science and Engineering (Vol. 994, No. 1, p. 012035). IOP Publishing.