Scholarly Research Journal for Interdisciplinary Studies
Online ISSN 2278-8808, SJIF 2021 = 7.380, www.srjis.com
PEER REVIEWED & REFEREED JOURNAL, JAN-FEB 2023, VOL- 10/75

A MATHEMATICAL APPROACH FOR DATA ANALYSIS IN PREDICTION OF GENERAL ELECTIONS IN INDIA

Giriraj Prashad Kalla

Research Scholar, Department of Mathematics, Pacific University, Udaipur (Raj.)

Dr. Ritu Khanna

Professor, Department of Mathematics, Pacific University, Udaipur (Raj.)

Paper Received On: 5 FEBRUARY 2023

Peer Reviewed On: 28 FEBRUARY 2023

Published On: 01 MARCH 2023

Abstract

The prediction of general elections holds substantial significance in understanding and foreseeing political outcomes. In recent times, mathematical approaches have gained momentum in analyzing vast datasets related to voter behavior, socio-economic indicators, and political landscapes. This review paper aims to explore the mathematical techniques and methodologies employed for data analysis in predicting general elections in India. By examining key research works and their underlying mathematical frameworks, this paper provides insights into the evolution of election prediction models and their implications fordemocratic processes.

1. Introduction: General elections play a pivotal role in shaping the future of a nation. The ability to predict election outcomes accurately contributes to inform decision- making and enhances the democratic process. Mathematical approaches have emerged as valuable tools for analyzing complex data sets, including historical election results, demographic data, and public sentiment indicators. In this review, we delve into various mathematical techniques that researchers have utilized for predicting general elections in India

2. Election Polling and Predictions

The paradox of the variability of American presidential election campaign polls despite the relative predictability of election outcomes. The authors question the methodology behind polling techniques and examine factors contributing to this variation. They emphasize the need for statistical rigor to improve the accuracy of election predictions.**[1]**The lessons learned from the 2000 presidential election to enhance the responsibility and credibility of the party-polling industry in the United States. Through statistical analysis, they examine the accuracy of pre-election polls and suggest improvements for a more reliable polling system**[2].** A work focuses on the relationship between U.S. presidential campaigns and the national vote. Using statistical methods, he investigates how campaign dynamics influence voter behavior and election outcomes, shedding light on the complex interplay between campaigns and public opinion**[3]**. The employs macroeconomic and statistical analysis to develop a model that relates economic indicators to electoral outcomes. His work highlights the influence of economic factors on voter behavior and provides a framework for understanding the connection between political economy and elections**[4]**.

3. Time Series Analysis and Forecasting

The authors introduce time series analysis as a powerful tool for forecasting and control. They present mathematical models to capture underlying patterns in time-dependent data, enabling the prediction of future trends and behaviors**[5].**The provide an accessible introduction to time series analysis and forecasting. Their work lays the foundation for understanding key concepts such as trend, seasonality, and noise, along with methods for modeling and predicting time-dependent data**[6]**.

4. Machine Learning and Predictive Analytics

The present a comprehensive overview of statistical learning techniques, emphasizing their application to data mining, inference, and prediction. The authors discuss algorithms for regression, classification, and unsupervised learning, showcasing the versatility of statistical learning in political analysis**[7]**. An introduces the concept of random forests as an ensemble learning method. By combining multiple decision trees, this approach enhances predictive accuracy and mitigates overfitting. Breiman's work demonstrates the potential of ensemble techniques in election prediction and political analysis**[8]**. The foundation for support-vector networks, a powerful classification technique. Their work highlights the importance of maximizing the margin between classes, leading to robust models capable of handling complex political data**[9]**.A seminal work on pattern recognition and machine learning provides insights into a wide range of algorithms, from neural networks to hidden Markov models. The integration of machine learning in political analysis enables more nuanced understanding of voter behavior and political dynamics**[10]**.

Predictive models are created using machine learning methods, such as logistic regression, decision trees, random forests, and support vector machines, based on past data and pertinent attributes. These algorithms recognise intricate links and patterns in data, enabling precise predictions. Large datasets may be handled by machine learning algorithms, which also automatically learn from data and adjust to shifting electoral dynamics. Machine learning models are very much helpful for different works (Figure 1)

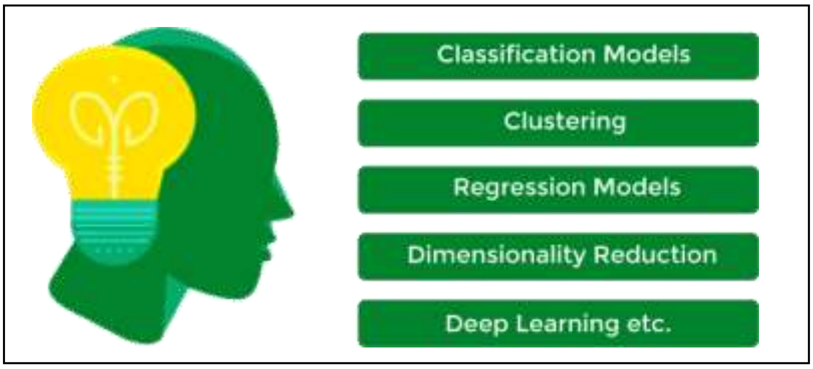

Figure 1: Machine learning model usefulness [15].

5. Social Media and Political Behavior

The application of Twitter data for election prediction. They showcase the potential of social media analytics in capturing real-time sentiment and public opinion, offering a case study of election prediction in the Republic of China**[11]**. A balanced survey of election prediction using Twitter data. The work critically assesses the limitations and possibilities of utilizing social media as a political forecasting tool, highlighting the challenges of bias and representativeness**[12]**.**Bond et al.** conduct a massive experiment on social influence and political mobilization using a dataset of over 61 million individuals. By analyzing the spread of political messages on Facebook, they shed light on the role of social networks in shaping political engagement**[13].**The authors investigate the ideological polarization on Twitter by analyzing the content of political tweets. They demonstrate the potential of social media data to study political communication and the alignment of online discourse with traditional political spectra**[14]**.

6. Mathematical model to data analysis of election prediction in India

Several mathematical models may be used to analyse the data and generate predictions based on it when it comes to data analysis for election prediction in India. In this regard, the following mathematical models are frequently used:

Logistic Regression

A common statistical model for binary classification issues, such as forecasting election results (win or loss), is logistic regression. It uses a logistic function to represent the connection between the binary outcome variable (win or loss) and the predictor factors (such as demographics, campaign expenditure, etc.). Figure 2 depicts the logistic model vs linear model.

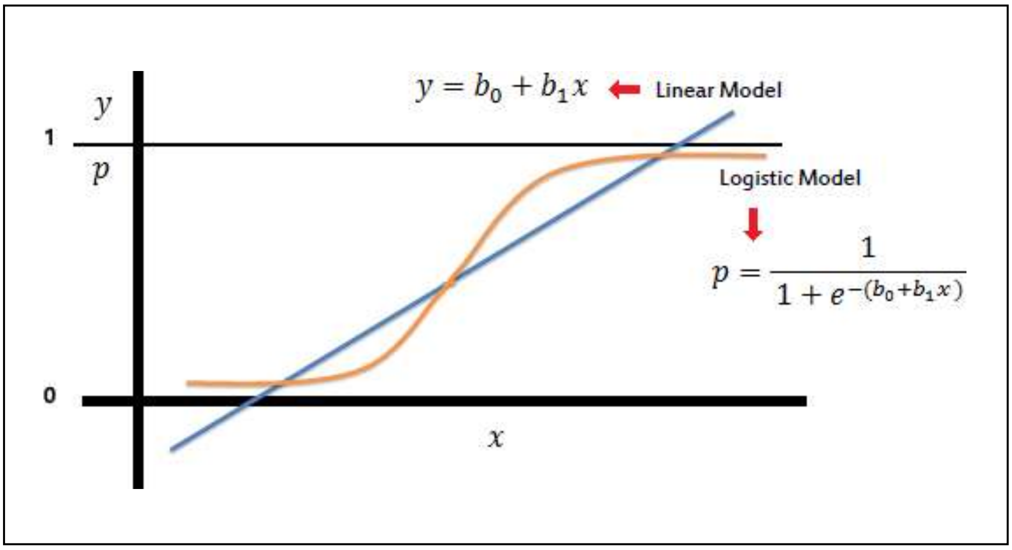

Figure 2 Logistic model vs linear model [7]

Decision Trees

Models like decision trees are flexible and effective for both classification and regression work. Each internal node represents a feature or property, and each leaf node represents a class label or result, resulting in a tree-like structure. Decision trees can give interpretable rules and can capture nonlinear interactions.

Random Forests

Multiple decision trees are combined into random forests, an ensemble model, to increase prediction accuracy and decrease overfitting. The final prediction is based on the combined results of all the trees, each of which is trained on a different random sample of the data. Random forests can handle high-dimensional data and are resistant to noise.

To achieve the optimum result, an ensemble classifier constructs multiple decision trees and combines them. Bootstrap aggregating or bagging is mostly used for tree learning. When given a set of data, $X = \{x_1, x_2, x_3, \ldots, x_n\}$ and the corresponding responses $\{Y = x_1, x_2, x_3, \ldots, x_n\}$ repeat the bagging process from $b = 1$ to B.

The seen sample x' is made by arranging the prediction $\sum_{b=1}^{B} fb(x')$ from every individual tree on x' **[17]**

$$j = \frac{1}{B}\sum_{b=1}^{B} fb(x')$$

(1)

The uncertainty of prediction on this tree is made through its standard deviation:

$$\sigma = \sqrt{\frac{\sum_{b=1}^{B}(fb(x')-f)^2}{B-1}}$$

(2)

Support Vector Machines (SVM)

A machine learning model called SVM seeks the best hyperplane to divide data points into multiple groups. By using kernel functions, it can solve classification issues that are both linear and nonlinear. SVM strives to increase the margin between classes, which leads to effective generalisation to new data.

Regression can be done using the Support Vector Machine method and one of its core features i.e., maximum margin. The Support Vector Regression (SVR) for classification adheres to the same principles as the SVM, with a few minor exceptions. First off, because outcome is a real number, there are an infinite number of alternative outcomes, making it very difficult to predict the information that is now available. Regression involves setting a tolerance margin (epsilon) that is generally in line with the problem's requirements that the SVM would have already specified. The algorithm is more complicated, so there is also a more complicated rationale that needs to be considered. The example below shows how SVR performs its principal function in Eq. (4) **[17]**.

$$Minimise\ SVR = \frac{1}{2}ww^T + c\sum_{i=}^{N}(\alpha^+ + \alpha^-)$$

(3)

Where c and α were the SVR's parameters.

Neural Networks

Election prediction may be done using neural networks, more especially deep learningmodels like Multilayer Perceptrons (MLPs) or Convolutional Neural Networks (CNNs). These models can learn hierarchical representations and capturing complicated connections within the data. For training, they need a lot of data and computer power (Figure 3).

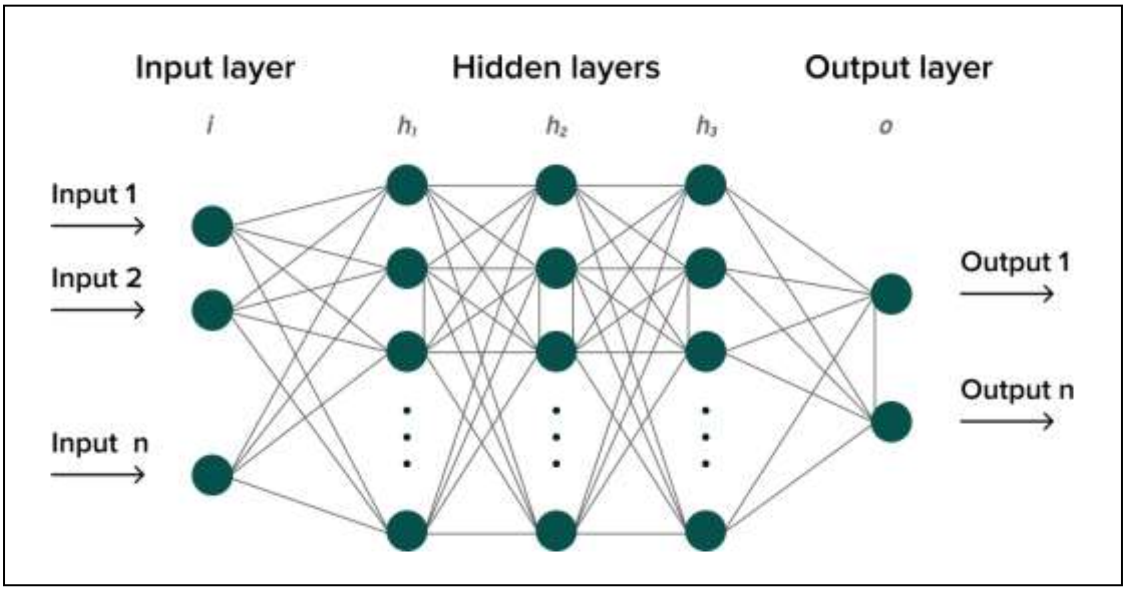

Figure 3. Structure of Neural network

Bayesian Networks

Bayesian networks use probabilistic graphical models to express the connections between variables. They enable the modelling of uncertainties, conditional probabilities, and dependencies. Bayesian networks are helpful for modelling a variety of elements in election prediction because they can handle a mix of discrete and continuous data.

The outcome of a Bayesian analysis of variance is obtained from a distribution of probabilities, as opposed to conventional regression techniques, where the output is simply derived from a single value of each attribute. "Y" results from a normal distribution with normalized mean and variance. Instead of focusing on the model parameters directly,

Bayesian Linear Regression seeks to discover the 'posterior' distribution for the model parameters **[18]**

$$Posterior = \frac{Likelihood+prior}{Normalization}$$

(4)

When an event, such as H, occurs and another event, such as E, also occurs, the probability that H will also occur is said to be posterior, or $(H \mid E)$. (H) tends for priority, which is the probability, that event H occurred before event A. Likelihood is a function in which a marginalized parameter variable is used.

Ensemble Methods

Multiple models are combined using ensemble methods to increase forecast reliability and accuracy. Examples include stacking (using predictions from numerous models as inputto a meta-model), boosting (combining weak classifiers into a strong one), and bagging (bootstrap aggregating). Ensemble approaches improve overall performance by reducing bias and variance problems.

Multi-Layer Perceptron (MLP)

The full form of MLP stands for "multi-Layer Perceptron". This neural network is trained using supervised learning to predict output data points based on input data points by supplying input and output data as Datasets **[19]**.

The activation function of MLP Regressor is provided below:

$(v_i) = (1 + e^{-})^{-1}$ and $y(v_i) = \tanh(v_i)$

(5)

The first is a logistic function that has a similar form to the first but has a range from 0 to 1, while the second is a hyperbolic tangent that goes from -1 to 1. Here, v_i is the weighted sum of the input connections, and y_i is the ith node's (or neuron's) output.

These are only a few illustrations of mathematical models that have been applied to data analysis for Indian election prediction. There are several machine learning methods can be modelled to utilise the nonlinearity nature of the data set (Figure 4). The unique research issue, the data at hand, and the required level of interpretability all influence the model selection. To achieve accurate predictions, it is crucial to evaluate each model's performance and applicability through validation and testing.

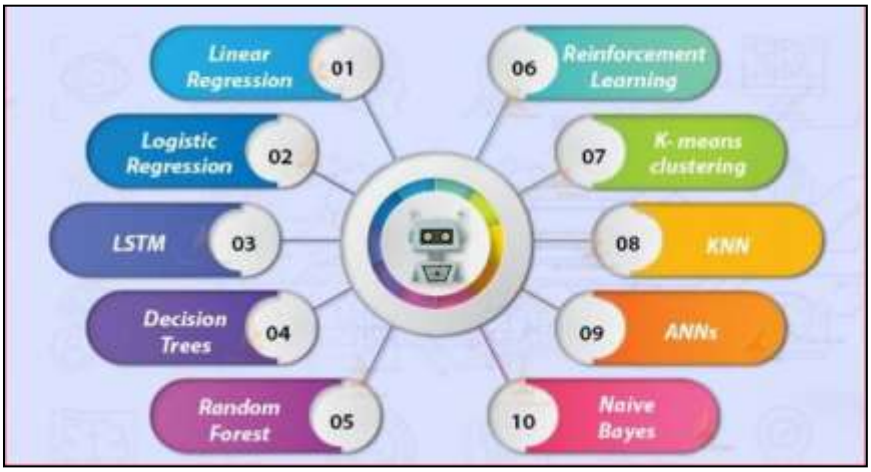

Figure 4. Different Machine learning methods

7. Conclusion

The literature reviewed in this paper exemplifies the dynamic interplay between political science and statistical methods. From election polling and predictive modeling to time series analysis, machine learning, and social media influence, these works collectively contribute to our nuanced understanding of political behavior and processes. As the field continues to evolve, the integration of advanced statistical techniques promises to uncover deeper insights into the intricate world of politics.

References

Gelman, A., & King, G. (1993). Why are American Presidential Election Campaign Polls so Variable when Votes are so Predictable? In Statistical Science.

Nadeau, R., & Lewis-Beck, M. S. (2001). Towards a more responsible party-polling industry in the United States: Lessons from the 2000 presidential election. In International Journal of Public Opinion Research

Campbell, J. E. (2008). The American campaign: U.S. presidential campaigns and the national vote. In Presidential Studies Quarterly.

Hibbs, D. A. (2008). The American Political Economy: Macroeconomics and Electoral Politics. In Annual Review of Political Science.

Tufte, E. R. (1975). Determination of Rotation in Factor Analysis. In British Journal of Mathematical and Statistical Psychology.

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. In Political Analysis.

Lewis-Beck, M. S., & Tien, C. (2018). The Political Economy Model: A Theoretical Update. In Election Sciences.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

Brockwell, P. J., & Davis, R. A. (2016). Introduction to Time Series and Forecasting. Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Breiman, L. (2001). Random Forests. In Machine Learning.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. In Machine Learning.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

https://www.javatpoint.com/machine-learning-models

https://www.saedsayad.com/logistic_regression.htm

A Poojitha Reddy, Anirban Tarafdar, Uttam Kumar Bera, (2023). "Regression based Machine Learning approach to predict Flight Price between Bangalore and Kolkata", IEEE 8th International Conference for Convergence in Technology (I2CT) Pune, India. Apr 7-9, 2023

Saqib, M. (2021). Forecasting COVID-19 outbreak progression using hybrid polynomial-

Taki, M., Rohani, A., Soheili-Fard, F., &Abdeshahi, A. (2018). Assessment of energy consumption and modeling of output energy for wheat production by neural network (MLPand RBF) and Gaussian process regression (GPR) models. Journal of Cleaner Production, 172, 3028-3041