

Election Voting Trend Prediction System

Abstract— This paper presents the use of Bayesian networks and K-Nearest Neighbor algorithms for predicting election results. Our motivation stemmed from the complexities of the election data available, which spans over 120,000 voting locations across 36 states in Nigeria, and the requirement to develop a procedure that takes into consideration voter trends that are influenced by political parties seeking to win. The system architecture's translation was utilised, and the prototyping methodology was adopted. In order to realize the requirements, the system was designed and implemented using Java and MySQL in accordance with specifications. Since the outcome is positive, it can serve as a benchmark for further study in this field, particularly when it comes to using data mining tools to analyze election results.

Keywords—K-Nearest Neighbour, Bayesian Network, Election, Voters, Prototyping methodology.

I. INTRODUCTION

Elections are a crucial component in choosing leaders in a democratic society. Elections give citizens a great chance to choose capable leaders whose main responsibility will be to advance the country. Voting trends in Nigeria from 1999 to the present seem to indicate that citizens are not actively choosing the politicians they want to lead them. It has been estimated that during an election, 60–65% of voters cast ballots. This indicates that the great majority of people's preferences are not fully expressed. Voting is how the people assert their sovereignty [1], this indicates that the cornerstone of a sound democratic ideal is the integrity of the electoral process. For the electorate, who are stakeholders in democracy, to develop confidence in the electoral process, it is vital that our governments function transparently. It should come as no surprise that the historical narrative surrounding elections in Nigeria is heavily skewed in the wrong ways. This situation is marked by varied degrees of election manipulation that fail to accurately reflect both the real election outcomes and the voting behavior of the populace. Proper documentation may only seem elusive and unachievable with such a poorly designed system in existence. This can only imply that a thorough investigation into voter behavior during the election process and the resulting election outcomes is highly necessary in order to enhance voting procedures, voter turnout, and the selection of credible candidates.

Before the 2023 general election in Nigeria, the conventional election system has been largely manual and human manner in nature. This approach entails gathering, tallying, and organizing paper ballots during elections. This process has shown to be vulnerable to human errors (omission) and manipulation, the majority of which are deliberate, this perennial approach calls to question, the credibility of the entire electoral process itself.

Given this, instituting a trustworthy electoral process and system that will ensure that voters can easily cast their ballots securely and that their votes are devoid of any form of manipulations. This emphasizes the credibility of implementing a one-person, one-vote system serving as a foundation of any credible election [11]. This is achievable and sustainable through the deployment of data-driven technology that will create data warehouse that will include datasets from every election in time past and in the future. This in view, further drives the quest for innovations on the means of election data acquisition, analysis and interpretation as well as the prediction of voter's behavior for future election forecast. Predicting patterns and trends on the basis of the driven data from election datasets could favour the direction of selecting the desired credible leaders. Additionally, by extracting and displaying interesting trends from the election datasets stored in the data warehouse, voters' interest and awareness can be piqued as well as forecast how they will behave in future elections.

There are numerous political organizations in Nigeria, each with its own structures and ideas that are employed to sway voters' decisions and voting habits. The nation is divided into electoral constituencies, which are defined as component parts or areas that are constitutionally represented in a state or nation. The Single Member Plurality (SMP), also referred to as "First-Past-The-Post" (FPTP), voting system used in Nigeria is seen to be the most basic [2]. This division or political zones also makes it necessary to investigate voter behavior in more detail. Factional principles, ethnic distribution, and regional disparities are a few of the variables influencing voters' attitudes regarding the election process. The political system's divisions, entities, and structure have made it possible to efficiently create a data warehouse from which patterns in future elections can be predicted. This can be powered up by the application of Artificial Intelligence (AI) in elections (Townsend, 2024).

The 2023 general elections in Nigeria deployed a version of electronic procedure atleast as a preliminary to future electronic voting system. The system was identified as the Bimodal Voter's Accreditation System (BVAS). The BVAS was deployed as a way of capturing eligible voter's data that will enable them to partake in the actual voting. The BVAS included features that capture distinguishing human qualities such one's biometrics in a dual mode and are used as a means of authenticating voters [12]. The BVAS possesses that capacity to retain voter's data such as voters' identification numbers (VINs) and was designed to transmit election data to a dedicated server in real time [13], however, flaws were witnessed and the claims for the BVAS became a doubt for future election result prediction [12].

Better models that have wide and proven application in many decision-making areas have become necessary, hence, this paper uses the KNN and Bayesian classification algorithm to forecast election results to illustrate Nigerian election trends. The usage of KNN and Bayesian networks as used in this paper has been used in both computer science and non-computer science fields alike for decision-making.

II. Related Work

[10] Reviewed over eighty three articles on the use of social media as a predicting data acquisition model for determining future election forecast. They suggested that most of the social media networks adopt machine learning algorithms that can support decision-making. That most of the model utilized are Support vector machine, (SVM), Artificial Neural Network (ANNs) and that these machine learning algorithm lack detail data clustering features that are cogent for detail classification patterns sufficient prediction. They proposed the addition of regression analysers to the state –of –the –art machine learning techniques in order to enhance better election prediction outcomes. Social media appeals to the sentiment of the voters which do not always reflect in the actual election. Hence they call for a future work that will adopt machine learning techniques that will be detailed in data clustering and classification that can sufficiently be used to predicting future election outcomes.

This is in itself poses a limitation that this research paper seeks to solve.

[3] Discussed the research topic "Shopping for Voters: Using Association Rules to Discover Relationships in Election Survey Data". Their work described a case study in which the Association node in SAS Enterprise Miner was used to examine relationships between the demographic characteristics, political party affiliation and media influences of respondents in a survey of U. S. voters. In using the Enterprise Miner software, a team made up of five students in a data mining course at Cleveland State University chose to

pose as political consultants analyzing the 2000 National Election Study in order to

- 1) Discover something interesting about the data and
- 2) Gain experience with the data mining process.

The study methodology is well documented, and the data are available in the form of a SAS data set. Each survey question is represented by a numeric variable and a corresponding format, which provides descriptions of the coded response values. Invalid or non-applicable responses are coded with special values such as 99. The results obtained shows groups of clusters related to voter's information and relevance to the electioneering process. However the research had limitation of Poor data sets and the results were not in plain English for easy readability. [4] In their work Data Mining Application into Potential Voters Trends in USA Elections with Regression Analysis. Data mining technique was very useful in bringing out the hidden information which was to provide solution to the problem. The research was to provide a basic model which relates potential voters in USA elections with periods of registration. Their results indicate that there is relationship with potential voters or participant and years of registrations. [5] Worked on Participation Anticipating in Elections Using Data Mining Methods. In their work, they described a general schematic of the architecture of participation anticipating system in presidential election by using KNN, Classification Tree and Naïve Bayes and tools orange based on crisp which had hopeful output.

They provided a method to anticipate public participation in presidential election by introducing data mining and its practical methodologies. The results obtained show that anticipating the political behaviour of candidates will be considerable helpful for election candidates to assess the possibility of their success and to be acknowledged about the public motivations to select the best candidates suitable for the job.

In consonance with [10], [15] adopted the ARIMAX model to predict the election result in Turkey 2023 election using the data from the traditional 2019 election and opinion poll shares from social media platforms of parties in May 2023. ARIMAX model is a time series forecasting model that uses exogenous variables to analyze and predict time-dependent data that may be influenced by external factors. Linear regression analysis were performed on the data from social media and was used to make prediction. Social media data was volume-based and relied only on the shares and post by various political parties that participated in the election. The result was remarkable but was open to future work. They suggested that social media data lack structure and completeness and are open to confounding algorithms. This poses a challenge and may introduce potential biases into the predictions. Hence the need for a reliable machine learning algorithm for future accurate prediction of election results.

[6] Implemented a Visual Data Mining in Indian Election System. For this purpose a Data Warehouse was built containing all the information related to election to increase awareness of voting. Interesting patterns that are extracted and represented using Visual Data mining to arrange awareness program. The approach was divided into 5 phases. The output generated using visual data mining techniques facilitates decision makers to make decisions like fluctuation in voting; voting ratio of male to female; voting rating of different party in a city/state. This type of analysis helps to arrange appropriate awareness program in different regions.

[7] Worked on the Analysis of Parliamentary Election Results and Socio-Economic Situation using Self-Organizing Map by presenting a complementary method based on data mining and specifically on the use of the self-organizing map.

[16] Employed a sentiment analysis data mining technique on social media data by using Naïve Bayes, Support Vector Machine and Decision Tree algorithm to build classifiers. This is used to anticipate participation in the 2019 general elections in India. The study however could not determine the total population of vote prediction on the basis of social media use because of misrepresentation of semantic in the classifier as it is almost impossible to analyse social media data on the basis of sentiment. This gives room for predictive models that are more direct.

The listed works demonstrate the hard work and creative ideas of numerous scholars toward the creation of a more advanced prediction system. More work needs to be done, which is why this paper has tried to predict election results using a different methodology. The works made use of information obtained in 2015 from Nigeria's electoral organization, the Independent National Electoral Commission (INEC).

II. PROPOSED WORK

Prediction systems entail the likelihood of events happening in the future which largely depends on trained data or evidence used to calculate the likelihood. The outcomes of the systems must be understood in terms of making declarative probability statements [8]. The additional material one can use to disorder the likelihood statement, the more precise the prediction. In this research work, the concept of K-nearest Neighbours algorithm is used for predicting the result of an election; the use of machine learning and probability theory applied on historical datasets in the prediction of Presidential Election Results by state distribution using Voters' trend. The proposed model framework is the K-Nearest Neighbour Technique, which employs an objective supervised learning approach. A major consideration in the use of the system is that there is no linear mapping method in which the outcome can be chosen based solely on a group of statistics. The relative probabilities and the distances of features will be used to obtain the classified data which in turn will be used for the prediction, i.e

the data set generated will assume an automatic classification and clustering.

The system could further be used to perform non-linear mapping based on a variety of relevant statistics, hence for this research work the dataset to be considered will be available from the central source aggregated over a period of time. The weight of each statistic must be determined prior to making a prediction.

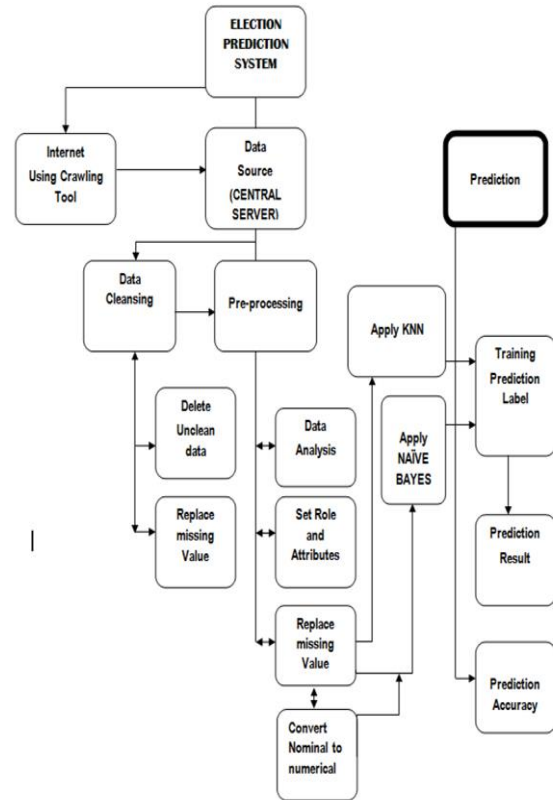


Figure 1: ELECTION TREND PREDICTION SYSTEM

III. METHODOLOGY

The system has the following sections Problem definition, Data collection and pre-processing, Modelling, Training and Performance evaluation. The system will provide a correct understanding of the existing problem. Here the understanding is broken into the project objectives and the requirements, data collection and pre-processing is handled by downloading the dataset on the server using URL crawling tool and Microsoft Excel Web Extraction Feature then pre-processed with excel to acquire the right features. The purpose of data collection is to determine the related features used in the prediction.

Pre-processing of data involved critically analyzing the datasets by replacing missing values, selecting roles and

attributes for the purpose of transformation and use by the model. The processed data is feed into the k-nearest network model and naïve Bayesian network. Bayesian networks have a causal semantics that makes the encoding of casual prior knowledge particularly straightforward, and also encode the strength of casual relationships with probabilities as consequence; prior knowledge and data can be efficiently combined. Bayesian networks offer an efficient and principled approach for avoiding the over fitting of data. BN was used to smoothen the data in a way that all available data can be used for training. The k-Nearest Neighbour technique is based on learning by analogy, that is by comparing a given test data with training data that are similar to it [14]. The training examples are described by the attributes. State, voter's reg. announced, post, business rules, deviation, total votes, candidate, and party. Training is done by classified votes of majority of its neighbours, with the dataset being assigned to the class most common amongst its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the dataset is simply assigned to the class of its nearest neighbour. The neighbours are taken from a set of examples for which the correct classification (the value of the label) is known. The basic k-Nearest Neighbour algorithm is composed of two steps: Find the k training examples that are closest to the unseen example. Take the most commonly occurring classification for these k examples (or, in the case of regression, take the average of these k label values). If the model does not fit the original expectations, they go back to the modelling phase and rebuild the model by changing its parameters until optimal values are achieved. The system was able to reduce computation time, Determine probability of interest from model and Infer Structure and Probabilities from Data.

IV. RESULTS AND DISCUSSION

The experiment was conducted by implementing a data mining (clustering) technique on the dataset that was used in this research work. Analyzing data, anticipating and assessing via patterns, classifying, categorizing and establishing association can be done by data mining.

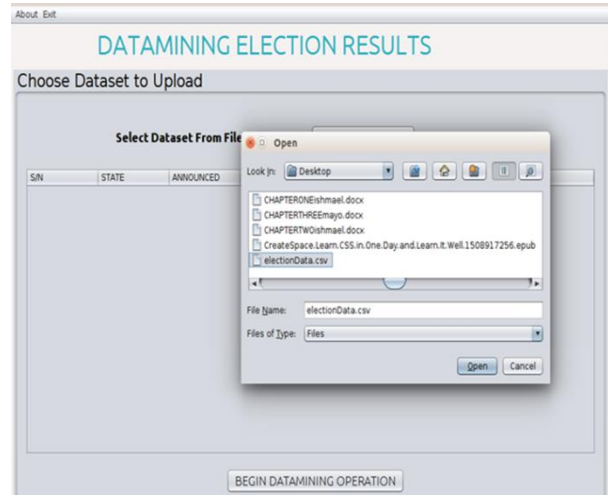


Figure 2: Dataset upload interface

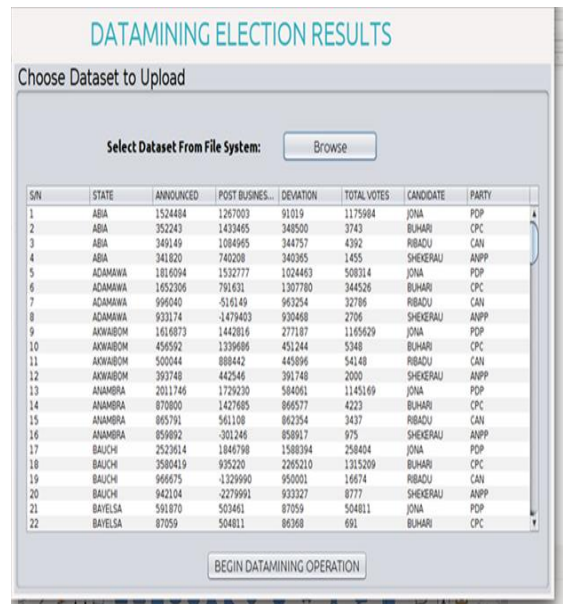


Figure 3: Dataset uploaded

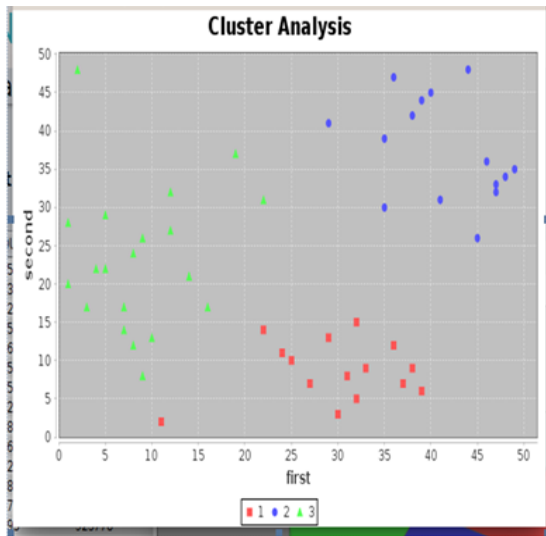


Figure 4: Cluster Analysis

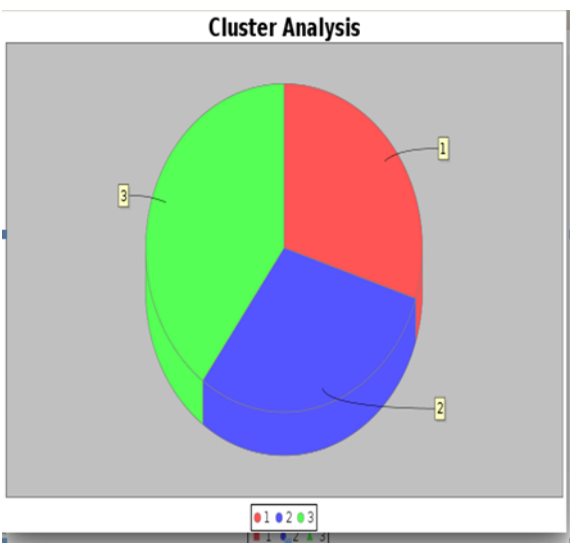


Figure 5: The Election Result Output Display

The result obtained are shown in figures 2 through figure 5. Figures 2 and 3 are the interface for uploading and displaying the datasets. Figures 4 and 5 provide a process that accounts for voters' trend that is geared towards winning the elections by a political party. KNN and BN were used to achieve these, identifying potential knowledge that lies in these data. The result shows that voters voted along party affiliation, political behaviour of candidates and ethnic divide. The application developed for this research is a high performance software that can run on a client/server or parallel processing computers and helps to analyze massive dataset in past elections and electioneering procedures to deliver answers to questions such as, "Which political party(ies) is/are most likely to win the present election or future elections, and why?"

V. CONCLUSION AND FUTURE SCOPE

In recent years, data mining algorithms and approaches have proven useful in identifying possible links between hidden patterns in a pool of massive datasets, owing to the growing volume of databases and the inadequacy of classic statistics methods to extract knowledge from data. Given its ability to forecast, classify, and generate predictions that are helpful for election dispositions, data mining techniques must be adopted in order to predict election results, a problem for which this work provides a solution. The development of real-time systems with better machine learning algorithms will be a focus of future study.

REFERENCES

- [1] G. Christian, *American Sovereigns. The People and America's Constitutional Tradition before the Civil War*. Cambridge University Press. pp 7 ISBN 978-0-521-88188-3, 2008.
- [2] B. Mashood, "Law and development in Africa: Towards a new approach". NIALS Journal of Law and Development, Vol. 1, issue. 1, pp 4-5 2011.
- [3] M. Mary, "Shopping for Voters: Using Association Rules to Discover Relationships Election Survey Data. pp 122-128. 2002.
- [4] M. Olagunju and A.R Tomori , "Data Mining Application Into Potential Voters Trends In USA Elections With Regression Analysis". Journal of Asian Scientific Research. Vol 2(12): pp 893-899, 2002.
- [5] B.S, Amin S. K Reza. and E. Laya "Participation Anticipating In Elections Using Data Mining Methods." International Journal on Cybernetics & Informatics (IJCI) Vol.2. pp 9-10 2013.
- [6] T.M. Kodinariya and R. Seta "visual data mining in indian election system" international journal of Computer Science and Engineering, IJCSE Vol. 4, No. 7; pp 1323-1329.
- [7] N. Pyry and H. Timo. "Analysis of Parliamentary Election Results and Socio-Economic Situation using Self-Organizing Map". Ics online journal pp 1-5, 2015
- [8] C.C., William Prediction Theory for Control Systems. Spring Lake, USA. Pp 50 2014.
- [9] S. Townsend. Predicting Election Results: The Power of Artificial Intelligence. *AI for the next century* pp 1-2, 2024
- [10] Kellyton dos Santos Brito, Member, IEEE, Rogério Luiz Cardoso Silva Filho, and Paulo Jorge Leitão Adeodato.. ASystematic Review of Predicting Elections Based on Social Media Data: Research Challengesand Future Directions, 2021.
- [11] Appel, A.W., DeMillo, R.A., and Stark, P.B. (2020). Ballot-marking devices cannot ensure the will of the voters. *Election Law Journal: Rules, Politics, and Policy*, 19(3), 432-450. <https://doi.org/10.1089/elj.2019.0619>.
- [12] T. P. Ayeni & O. M. Aweh. Examining the impact of bimodal voter Accreditation system (BVAS) as a game-changer In Nigeria 2023 elections, 2023, *The Information Technologist: An International Journal of Information and Communication Technology (ICT) Vol. 20 No. 2, December, 2023*
- [13] L. Bayode, Overview: Bimodal Voter Accreditation System (BVAS), *ToT for TEI Trainers for Anambra Governorship Election 2021*. pp 13-14, 2022.

-
- [14] A. Singh. *Analyzing Election Results and Predictions Using Machine Learning*. LAP Lambert Academic publishing, pp 21-23, 2023.
- [15] A. Bozanta, F. Bayrak & A. Basar. Prediction of the 2023 Turkish Presidential Election Results Using Social Media Data. *Rapor-85-TR-Mayis, Vol 1, pp 11-15, 2023*.
- [16] A.A. Kulkarni, M.G. Dhopeswarkar, P.S. Jaiswal & B. A. Marathwada. Analyzing Political Opinions and Prediction of Election Result of the Indian Election Using Data Mining Approach. *International Journal of Creative Research Thoughts (IJCRT)*, 9(12), 2021, ISSN: 2320-2882.