2026 Volume: 02 Issue: 11 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-16

AN ENHANCED REMOTE VOTING SYSTEM FOR SMART CITIES USING K-MEANS CLUSTERING FOR VOTER BEHAVIOUR ANALYSIS

Chioma Jessica Onwudiwe* and Moses Okechukwu Onyesolu

Department of Computer Science, Nnamdi Azikiwe University, Awka – Nigeria.

Article Received: 24 September 2026
Article Revised: 14 October 2026
Published on: 03 November 2026

***Corresponding Author: Chioma Jessica Onwudiwe**
Department of Computer Science, Nnamdi Azikiwe University, Awka – Nigeria.

DOI: https://doi-doi.org/101555/ijrpa.2679

ABSTRACT

The increasing application of digital technologies in public services has created opportunities to modernize electoral processes through remote voting. However, the effective use of voter-related data to identify patterns in remote voting systems remains an important area for further investigation. This study developed an enhanced remote voting system for smart cities and incorporated K-Means clustering to analyze voter-related characteristics. A system-development approach was adopted, using a simulated dataset of 500 voters. Age, voting duration, and login attempts were selected as clustering variables. K-Means clustering using Euclidean distance was configured with four clusters and a maximum of 100 iterations, converging after eight iterations. The resulting clusters contained 60 (12.0%), 84 (16.8%), 186 (37.2%), and 170 (34.0%) voter records, respectively, with differences in their mean characteristics. System functionality evaluation recorded an Authentication Accuracy (AA) of 98%, Vote Recording Accuracy (VRA) of 83.7%, and Voting Success Rate (VSR) of 83.7%. The findings demonstrate the feasibility of integrating K-Means clustering into a remote voting platform to identify patterns in voter-related system data. The developed system combines remote voting functionality with an analytical component that may support system evaluation and improvement.

KEYWORDS: Remote voting; Electronic voting; Smart cities; K-Means clustering; Voter behaviour; Voting system.

1.0 INTRODUCTION

The development of smart cities has increased the use of digital technologies to provide efficient, accessible, and data-driven public services. As governments continue to adopt digital platforms for civic participation and public administration, voting systems can also benefit from technologies that support remote access and improved management of voter-related information. A remote voting system enables eligible voters to participate in elections without being physically present at a designated polling station. Within a smart-city environment, such a system can contribute to the delivery of accessible and technology-enabled electoral services. Despite the potential benefits of remote voting, the effectiveness of such systems depends not only on the ability to allow voters to cast their ballots remotely but also on how voter-related system data are utilized to understand system usage patterns. Conventional remote voting platforms generally concentrate on core functions such as voter registration, authentication, ballot access, vote submission, and result processing. However, limited attention may be given to the analysis of data generated during voter interaction with the system. Such data can provide useful information about observable characteristics of system usage, including voter age, voting duration, and login attempts.

The integration of data-analysis techniques into remote voting systems provides an opportunity to identify meaningful patterns within voter-related system data. K-Means clustering, an unsupervised machine-learning technique, can group data records according to similarities among selected variables without requiring predefined class labels. Its application in a remote voting system can therefore provide an analytical component for examining patterns in voter-system interactions and supporting better understanding of system usage.

This work therefore developed an enhanced remote voting system for smart cities and incorporated K-Means clustering to identify patterns in voter-related system data. The clustering analysis was based on age, voting duration, and login attempts from a simulated dataset of 500 voters. The study demonstrates how K-Means clustering can be integrated into a remote voting platform as a data-analysis component for examining observable voter-system interaction characteristics.

1.1 Aim and Objectives of the Study

Aim: To develop and evaluate an enhanced remote voting system for smart cities incorporating a K-Means clustering component for exploratory analysis of voter-system interaction data.

The objectives are to:

1. To design and implement a prototype remote voting system with voter registration, authentication, and voting modules.
2. To generate a simulated dataset of voter-system interaction attributes.
3. To apply K-Means clustering to identify groups based on age, voting duration, and login attempts.
4. To evaluate clustering quality and discuss implications for system usability and support.

1.2 Remote Voting

Remote voting refers to the ability of electorates to cast their votes from locations other than their designated polling stations, whether they are located abroad or within the country. This democratized approach allows voters to participate in elections without the constraints of physical presence, thereby fostering inclusivity and engagement. Merino (2023) defines remote voting systems as electronic voting platforms that allow eligible voters to participate in elections from locations other than designated physical polling stations by using Internet-connected electronic devices. Such systems are designed to facilitate convenient voter participation while maintaining essential electoral requirements such as voter authentication, vote integrity, privacy, security, and system availability. Remote voting systems offer numerous advantages, including increased participation, cost efficiency, timely reporting.

1.3 Smart Cities

Smart cities are urban areas that use advanced technology, data and innovation to enhance the quality of life of citizens, improve the efficiency of city services, and promote sustainable development. Intelligent sensors, particularly Internet of Things (IoT) devices, along with 5G technology, are becoming increasingly pervasive, effectively meeting users' needs. They rely heavily on these technologies to collect and track data related to vehicular traffic, waste management, water systems, drainage, potholes, smart buildings, smart grids, and environmental monitoring (Ullah *et al.,* 2023). Smart cities leverage technology and data to enhance the quality of life for citizens and improve urban services. IoT and machine learning have emerged as key technologies enabling smart city solutions that depend on large-scale data collection, analysis, and decision-making. Smart cities are urban areas that incorporate privately owned technologies and software into public spaces while also privatizing aspects of public service provision (Alahi *et al*., 2023).

1.4 K-Means Clustering

K-Means clustering is an unsupervised machine learning algorithm used to partition numerical observations into a predetermined number of clusters based on similarity (Figure 1). The algorithm requires the researcher to specify the number of clusters, represented by K**,** before the clustering process begins. Géron (2022) identifies K-Means as one of the principal clustering algorithms used in unsupervised learning, while Han *et al.* (2022) discuss clustering as a major data mining task for discovering meaningful groupings within datasets. K represents the predetermined number of clusters to be formed, while means refers to the calculation of the arithmetic mean of the observations assigned to each cluster. This mean serves as the cluster's centroid**,** representing the centre of the observations belonging to that cluster. The fundamental objective of K-Means is to assign observations to clusters in such a manner that observations within the same cluster are relatively similar and observations belonging to different clusters are comparatively dissimilar. The algorithm achieves this by minimizing the within-cluster variation, commonly expressed through the Within-Cluster Sum of Squares (WCSS) or the Sum of Squared Errors (SSE).

1.5 Working Principle of K-Means Clustering

The K-Means algorithm begins by selecting the desired number of clusters (K), followed by random initialization of centroids. Each observation is assigned to the nearest centroid based on Euclidean Distance, after which new centroids are calculated as the mean of the assigned observations. The Euclidean Distance is the most commonly used distance measure in K-Means because it computes the straight-line distance between two points in multidimensional space. This iterative process continues until convergence is achieved (Géron, 2022). An iteration consists of assigning observations to clusters and recalculating centroids, whereas convergence occurs when centroid positions or cluster memberships no longer change significantly. The objective of K-Means is to minimize the Within-Cluster Sum of Squares (WCSS), thereby reducing the total squared distance between observations and their corresponding centroids. It divides data into clusters so that data inside the same cluster are very similar and data in different clusters are very different. The process of K-Means clustering involves several key steps - **choosing the number of clusters (K)** (This involves the user defining the number of clusters to form; **initializing centroids** (this involves randomly selecting K data points as initial centroids; a**ssigning data points** (each data point is assigned to the nearest centroid based on a distance metric, typically Euclidean distance; **recalculating centroids** (after all points are assigned, the centroids are recalculated as the

mean of all points in each cluster and i**terating** (a**ssigning data points** and r**ecalculating centroids** are repeated until the centroids stabilize and do not change significantly, indicating convergence (Ikotun *et al*., 2023).

2.0 RELATED LITERATURE

Noriega Zapata *et al.* (2026) investigated the accessibility of electronic voting for persons with disabilities in developing countries. Their review focused specifically on electronic voting technologies associated with accessibility and disability inclusion. The findings identified persistent barriers relating to digital accessibility, inclusive design, technological infrastructure, and institutional arrangements, and emphasized that accessibility should be incorporated into the design of electronic voting systems from the beginning rather than added subsequently. The study revealed technological and structural barriers that can prevent persons with disabilities from benefiting equally from electronic voting, particularly in developing countries. Fantozzi *et al.* (2025) conducted a worldwide review of electronic and Internet voting and examined socioeconomic and political factors associated with adoption. Using data covering countries worldwide, the authors investigated factors including population, geographical area, GDP, Internet usage, and the Democracy Index. Their results indicated that adoption varies considerably across regions and that Internet voting remains limited in Africa.

Mohamed and Ng Sok (2025) conducted a comprehensive review of recent developments in electronic voting from 2020 to 2025**.** The researchers used a literature-review methodology to examine recent e-voting research and organized the findings around three major themes: security enhancement, accessibility and inclusivity, and trust, verifiability and governance**.** Their review found substantial advances in electronic voting technologies but also identified continuing challenges involving scalability, unequal socio-technical adoption, regulatory differences, and the need to balance technological development with human and institutional considerations. The study synthesized recent research and demonstrated that improving electronic voting requires more than technological functionality alone.

Adida *et al.* (2025) proposed MERGE, a remote-voting protocol designed to address the integrity and privacy problems associated with Internet-based voting for overseas military personnel. The methodology involved the design and formal analysis of a voting protocol**,** supported by a prototype implementation. MERGE combines rapid electronic submission of a vote record with a paper ballot that can subsequently be used in a risk-limiting audit**.** The

study demonstrated that the approach can provide authentication, privacy, paper-based cast-as-intended verification, public tally verification, and probabilistic recorded-as-cast verification. Its contribution is the development of a remote-voting protocol that combines electronic speed with physical evidence for auditing.

Barbierato and Gatti (2024), in Decoding Urban Intelligence: Clustering and Feature Importance in Smart Cities, investigated the use of unsupervised machine learning to examine smart-city intelligence using data associated with the IMD Smart City Index. Their study explored patterns in smart-city indicators and the importance of different features in characterising urban intelligence. The research illustrates how clustering can support the analysis of complex smart-city data. However, its focus was urban intelligence rather than remote voting or voter-system interactions. Consequently, it does not directly address the integration of clustering into a remote voting platform.

Ikotun et al. (2023) presented a comprehensive review of K-Means clustering algorithms, their variants, applications, and continuing research challenges. The authors identified the algorithm's simplicity and relatively low computational complexity as important reasons for its widespread use. They discussed limitations associated with selecting the number of clusters, initializing centroids, convergence to local solutions, and the use of Euclidean distance. The review provides a methodological foundation for applying K-Means to voter-related system data. However, it addresses the algorithm broadly rather than its integration into an operational remote voting system. The reviewed literature reveals several relevant but distinct research directions. Studies by Fantozzi et al. (2025) examine the adoption and electoral implications of Internet voting. Barbierato and Gatti (2024) demonstrate the application of clustering to smart-city data, and Ikotun et al. (2023) review the K-Means algorithm and its methodological challenges. Although these studies contribute to the understanding of remote voting, digital trust, smart-city analytics, and clustering methods, they address these subjects from different perspectives. The research gap addressed by this study is the integration of K-Means clustering into a developed remote voting system for analyzing voter-level system data using age, voting duration, and login attempts**.**

To address this gap, the work developed an enhanced remote voting system for smart cities and incorporated K-Means clustering to identify patterns within 500 simulated voter records. Four clusters were generated to demonstrate how the algorithm can group records according to similarities in the selected variables.

3.0 MATERIALS AND METHODS

The study adopted a design and development research approach to develop an enhanced remote voting system for smart-city environments. The system was implemented as a functional prototype incorporating voter registration, authentication, remote voting, and a K-Means clustering component for analysing voter-related system data. An experimental computational approach was used to apply and evaluate the clustering process.

3.1 Study Population and Data Source

The study used a simulated dataset representing 500 remote voters. The dataset was generated to provide the voter-related information required for system testing and clustering analysis. Although the simulated dataset contained several voter attributes, age, voting duration, and login attempts were selected as the clustering variables because they represent observable characteristics of voter interaction with the remote voting system. Therefore, the study population for the clustering analysis comprised 500 simulated voter records rather than actual Nigerian voters.

3.2 Data Simulation Procedure

A simulated dataset of 500 voter records was generated to evaluate the application of the K-Means clustering algorithm in the proposed enhanced remote voting system. Each record contained three variables: voter age, voting duration, and the number of login attempts. The data were synthetically generated to represent variations in voter characteristics, the time taken to complete the voting process, and the number of attempts required to access the system. The simulated records were organised into a structured dataset and prepared for analysis. Voter age, voting duration, and login attempts were used as input features for the K-Means clustering algorithm, with the number of clusters set to ($K = 4$). The algorithm grouped the voter records into four clusters based on similarities in these features, enabling the examination of distinct patterns in voter-related system activities. The resulting clusters were analyzed using visualisations and summary statistics to demonstrate the potential of K-Means clustering for voter behaviour analysis. Since the dataset was simulated rather than collected from actual voters, the results serve as a proof of concept and do not represent verified real-world voter behaviour.

3.3 Technology Stack: The enhanced remote voting system was developed using a web-based technology stack comprising HTML, CSS, JavaScript, PHP, and MySQL. HTML and CSS were used to structure and style the user interface, while PHP handled server-side

processing and application logic. MySQL served as the database management system, with PHP Data Objects (PDO) providing database connectivity. The application was developed and tested in a local WAMP Server environment. K-Means clustering was incorporated as the data-analysis component to group 500 simulated voter records using age, voting duration, and login attempts, with Euclidean distance and four predefined clusters. Scatter plots and bar charts were used to present the clustering results.

3.4 System Architecture

System architecture diagram is a visual representation of the major components of a software system and how they interact with one another. It shows the structure of the system, how data flows between components, and how the components work together to achieve the system's objectives.

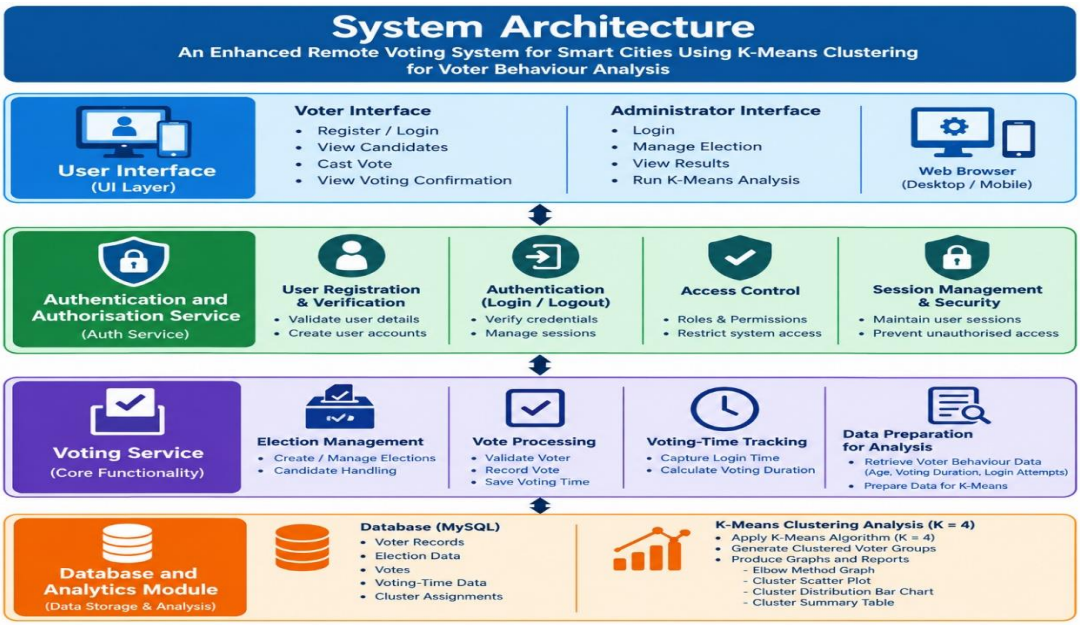


Figure 1: System Architecture.

3.5 Data Analysis

K-Means clustering was applied as the principal data-analysis technique. The selected variables were standardized using z-score normalization before clustering, and Euclidean distance was used to measure the similarity between observations and cluster centroids. The number of clusters was set to four (K = 4), with a maximum of 100 iterations. The algorithm converged after eight iterations, producing four clusters of 60, 84, 186, and 170 simulated voter records, respectively.

3.5.1 Elbow Method

The Elbow Method was employed to assist in determining an appropriate number of clusters for the K-Means algorithm. The method evaluates the Within-Cluster Sum of Squares (WCSS), also referred to as inertia, for different values of (K). The Elbow Method graph generated by the developed system was used to examine the reduction in WCSS as the number of clusters increased. The curve provides empirical support for selecting the number of clusters used in the final analysis.

The final K-Means implementation used four clusters. The selected value of (K=4) was subsequently used to generate the voter cluster assignments and the corresponding visual analytics. The Elbow Method is as an aid to determining the appropriate number of clusters rather than an absolute proof that one particular value of (K) is universally optimal. This is because WCSS will normally continue to decline as additional clusters are introduced.

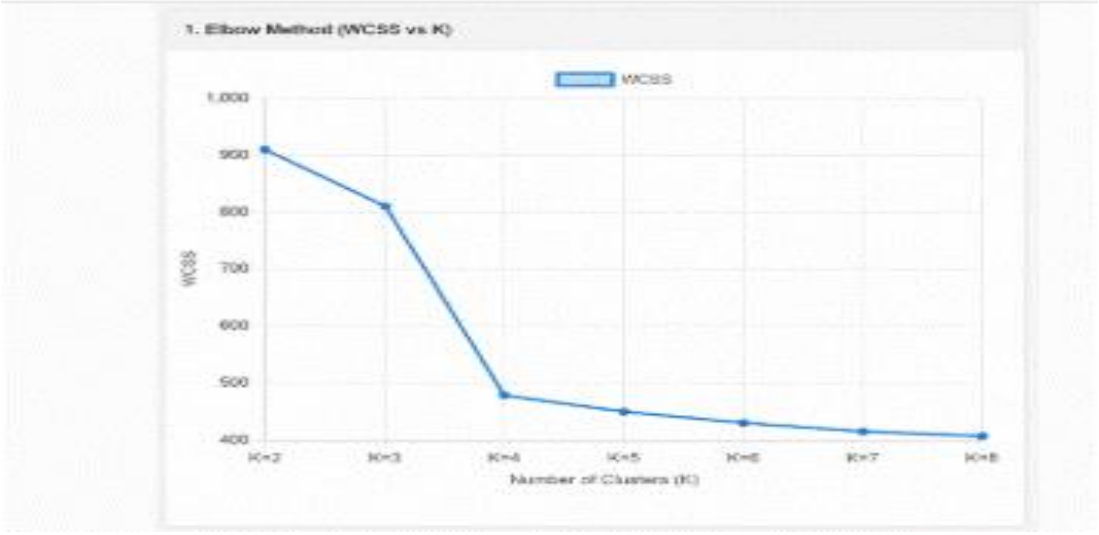


Figure 2: Elbow Method Graph.

4.0 RESULTS

The developed remote voting system successfully incorporated K-Means clustering as a data-analysis component for identifying patterns in voter-related system data. The analysis was performed on a simulated dataset of 500 voter records using age, voting duration, and login attempts as the clustering variables. The K-Means algorithm used Euclidean distance and was configured to produce four clusters (K = 4), with a maximum of 100 iterations. The algorithm converged after eight iterations, indicating that stable cluster groupings were obtained within the specified iteration limit.

4.1 K-Means Cluster Scatter Plot

A two-dimensional scatter plot was generated to visualize the distribution of the 500 simulated voters across the four clusters. The visualization represents:

- **X-axis:** Age
- **Y-axis:** Voting Duration
- **Cluster membership:** represented using different marker groups
- **Centroids:** represented using distinct centroid markers

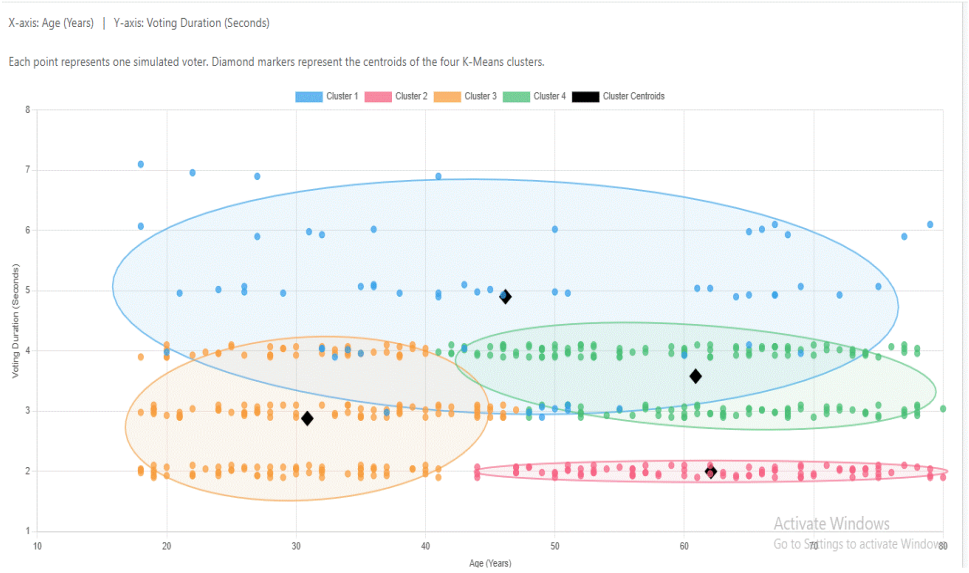


Figure 3: K-Means Cluster Scatter Plot.

4.2 K-Means Cluster Distribution and Bar Chart

Table 1 shows the cluster distribution used to compare the number of voters assigned to each cluster.

Table 1: K-Means Cluster Distribution.

Cluster	Number of Voters	Percentage
Cluster 1	60	12.0%
Cluster 2	84	16.8%
Cluster 3	186	37.2%
Cluster 4	170	34.0%
Total	**500**	**100%**

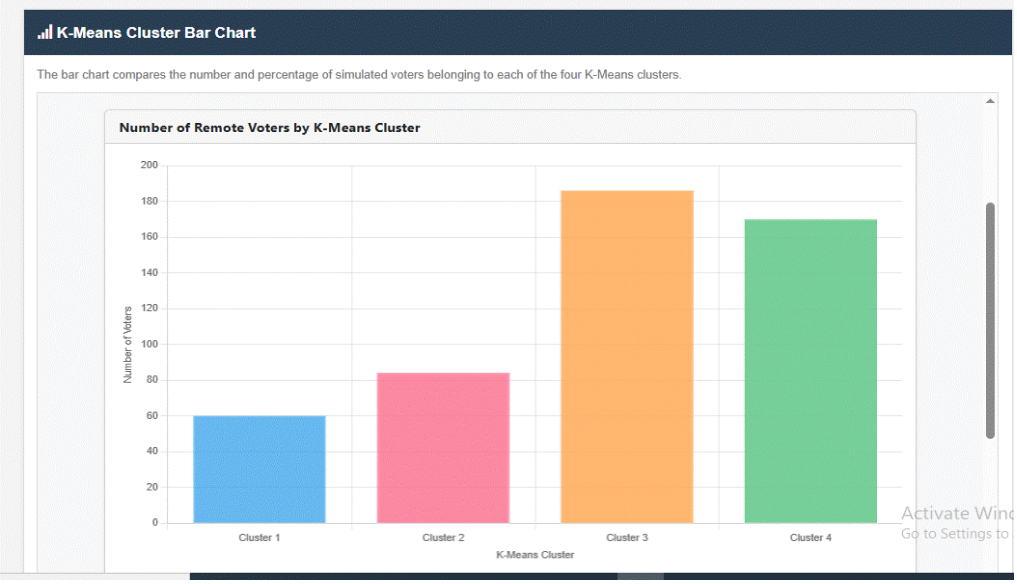


Figure 4: K-Means Cluster Bar Chart.

4.3 Interpretation of Cluster Distribution

The cluster distribution indicates that Cluster 3 is the largest group, containing 186 voters (37.2%), while Cluster 1 is the smallest, containing 60 voters (12.0%). Cluster 4 contains 170 voters (34.0%), while Cluster 2 contains 84 voters (16.8%).

The distribution demonstrates that the K-Means algorithm did not divide the simulated voters into four equally sized groups. Rather, it identified groups according to similarities in the selected feature space.

Cluster 1

Cluster 1 contains 60 voters (12.0%). The average age is approximately 46.18 years, the average voting duration is approximately 4.90 seconds, and the average number of login attempts is approximately 2.33. The cluster is distinguished by its relatively high voting duration and login-attempt level. It can therefore be interpreted as a group of simulated voters whose interaction with the remote voting process generally involved longer voting activity and more authentication attempts.

Cluster 2

Cluster 2 contains 84 voters (16.8%). Its average age is approximately 62.06 years, average voting duration is 2.00 seconds, and average login attempts is approximately 1.00. This cluster represents an older voter group characterized by relatively short voting duration and very few repeated login attempts.

Cluster 3

Cluster 3 is the largest cluster, containing 186 voters (37.2%). Its average age is approximately 30.87 years, average voting duration is approximately 2.88 seconds, and average login attempts is approximately 1.02. The cluster therefore represents the dominant younger voter group within the simulated dataset, with relatively short voting activity and approximately one login attempt.

Cluster 4

Cluster 4 contains 170 voters (34.0%). Its average age is approximately 60.88 years, average voting duration is approximately 3.58 seconds, and average login attempts is approximately 1.01. This cluster represents another relatively older voter group. However, compared with Cluster 2, the voters in Cluster 4 exhibit a longer average voting duration.

4.4 Silhouette Score

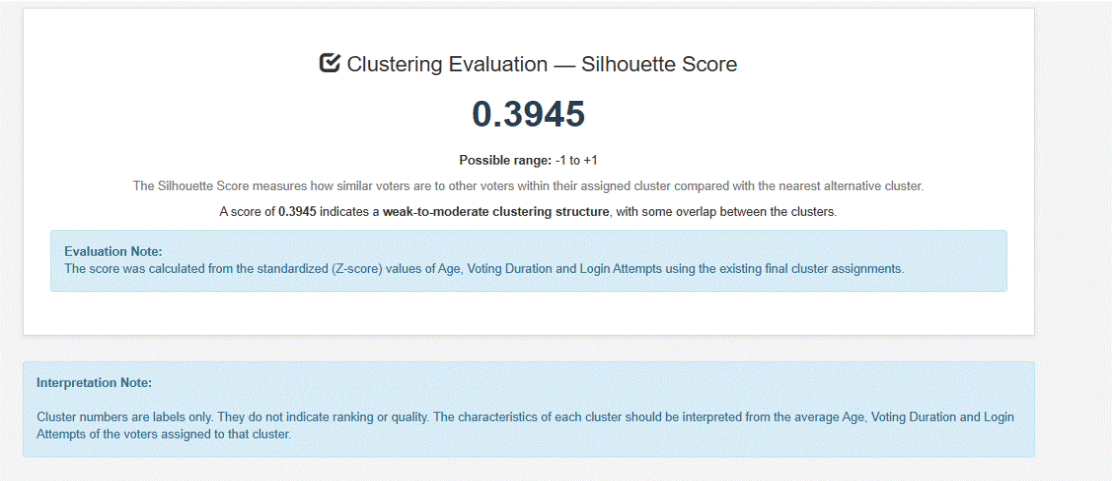


Figure 5: Silhouette Score Clustering Evaluation Result.

Interpretation

The obtained Silhouette Score of 0.3945 indicates a moderate but not strongly separated clustering structure. The result suggests that the clusters possess some identifiable internal structure, but there is also a degree of overlap between some voter groups.

4.5 System Functionality Metrics

The system functionality evaluation considered four principal metrics: authentication accuracy, vote recording accuracy and voting success rate. These metrics provide an indication of the reliability, correctness, usability, and operational efficiency of the developed voting system.

1. Authentication Accuracy (AA):

Authentication Accuracy measures the percentage of registered voters who were correctly authenticated before gaining access to cast their votes.

$$AA = \frac{\text{Correct Authentications}}{\text{Total Authentication Attempts}} \; x \; 100$$

$$AA = \frac{49}{50} \; x \; 100 \qquad = 98\%$$

Interpretation: The authentication module achieved an accuracy of 98%, indicating that almost all legitimate voters were successfully verified before voting.

2. Vote Recording Accuracy (VRA):

Vote Recording Accuracy measures the percentage of votes that were successfully stored in the database without omission, duplication, corruption or unintended modification.

$$VRA = \frac{\text{Correct Recorded}}{\text{Total Vote Attempts}} \; x \; 100$$

$$VRA = \frac{41}{49} \; x \; 100 \qquad = 83.7\%$$

Interpretation: The Vote Recording Accuracy of 83.7% demonstrates that almost every vote submitted by authenticated voters was successfully stored in the database.

3. Voting Success Rate (VSR)

Voting Success Rate measures the percentage of authenticated voters who successfully completed the voting process.

$$VSR = \frac{\text{Successful Votes}}{\text{Authenticated Voters}} \; x \; 100$$

$$VSR = \frac{41}{49} \; x \; 100 \qquad = 83.7\%$$

Interpretation: The result shows that about 83.7% of authenticated voters successfully completed the voting exercise.

5.0 DISCUSSION

The findings demonstrate the feasibility of integrating K-Means clustering into a remote voting system to analyse patterns in voter-related system data. Using a simulated dataset of 500 voter records, the algorithm grouped the records into four clusters based on age, voting duration, and login attempts. The resulting clusters differed in size and average characteristics, demonstrating the potential of unsupervised machine learning to identify

similarities among records within a digital voting platform. This approach provides an analytical component that complements the functional aspects of remote voting systems and supports exploratory data analysis in smart-city environments.

The cluster distribution revealed distinct patterns within the simulated dataset. Cluster 3 was the largest, comprising 186 records (37.2%), with an average age of 30.87 years, a mean voting duration of 2.88 seconds, and approximately 1.02 login attempts. Cluster 1 was the smallest, containing 60 records (12.0%), with an average voting duration of 4.90 seconds and approximately 2.33 login attempts. Cluster 2 contained 84 records (16.8%) and had the highest average age of 62.06 years, while Cluster 4 comprised 170 records (34.0%), with an average age of 60.88 years. These results indicate that K-Means identified groups with different characteristics based on the selected variables. However, because the dataset was simulated, these patterns should not be interpreted as evidence of actual behavioural differences among Nigerian voters.

The study has practical implications for developing data-driven voting services in smart-city environments. By grouping records according to similarities in age, voting duration, and login attempts, the clustering component provides a basis for exploring patterns in system usage. If evaluated using suitable real-world data, similar analysis could help administrators investigate usability concerns, identify areas requiring technical support, and inform improvements to the voting interface. For example, unusually long voting durations or repeated login attempts could justify further investigation. The study demonstrates the feasibility of integrating K-Means clustering into a remote voting system for exploratory analysis of voter-related system data. Its principal contribution is the incorporation of a clustering-based analytical component into a developed remote voting platform within a smart-city context.

5.1 Limitations of the System

Although the enhanced remote voting system was designed to improve voting efficiency and security, it has some limitations. The system depends on a stable internet or mobile network connection. Users in areas with poor or no network coverage may be unable to access the platform or cast their votes successfully. Secondly, the K-Means component was evaluated using a simulated dataset of 500 voters. Although the simulated dataset provided a controlled and consistent basis for developing, testing, and demonstrating the clustering functionality, it may not fully represent the demographic and behavioural characteristics of actual voters. The

use of simulated data was necessary within the scope of the study because access to authentic voter records involves privacy, security, ethical, and administrative considerations.

6.0 CONCLUSION

This study developed an enhanced remote voting system for smart cities and integrated K-Means clustering for voter-related data analysis. The system provides functions for voter registration, authentication, remote voting, data storage, and election-result presentation. Using a simulated dataset of 500 voters and three variables—age, voting duration, and login attempts—the K-Means algorithm produced four clusters. The algorithm converged after eight iterations, producing clusters of 60, 84, 186, and 170 voters.

The results demonstrate that K-Means can identify distinguishable groups within voter-related system data. The contribution of the proposed system is therefore the integration of remote voting functionality and voter-data clustering within a single system. Unlike prior systems that only record votes, this work shows how unsupervised learning can provide an early warning layer for usability - e.g., Cluster 1 (high duration + 2.33 login attempts) could trigger a help prompt in future smart-city deployment.

However, the findings are descriptive and based on simulated data. They should not be interpreted as evidence of political preferences, actual voter behaviour in a population, or election outcomes. Future research can evaluate the system with appropriately governed real-world datasets and assess additional dimensions such as usability, efficiency, accessibility, and clustering quality.

REFERENCES

1. Adida, B., Caron, J., Mirzaei, A., & Teague, V. (2025). MERGE: Matching electronic results with genuine evidence, for verifiable voting in person at remote locations. In *Electronic Voting (E-Vote-ID 2025)* (pp. 1–16). Springer. https://doi.org/10.1007/978-3-032-05036-6_1
2. Alahi, M. E. E., Sukkuea, A., Tina, F. W., Nag, A., Kurdthongmee, W., Suwannarat, K., & Mukhopadhyay, S. C. (2023). Integration of IoT-Enabled Technologies and Artificial Intelligence (AI) for Smart City Scenario: Recent Advancements and Future Trends. *Sensors*, *23*(11), 5206. https://doi.org/10.3390/s23115206
3. Barbierato, E., & Gatti, A. (2024). Decoding urban intelligence: Clustering and feature importance in smart cities. *Future Internet, 16*(10), 362. https://doi.org/10.3390/fi16100362

4. Fantozzi, P., Iecher, M., Laura, L., Naldi, M., & Rughetti, V. (2025). Electronic voting worldwide: The state of the art. *Information, 16*(8), 650. https://doi.org/10.3390/info16080650
5. Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.
6. Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann. https://doi.org/10.1016/C2013-0-18660-6
7. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences, 622*, 178–210. https://doi.org/10.1016/j.ins.2022.11.139
8. Merino, L.-H. (2023). Electronic voting. In Trends in data protection and encryption technologies (pp. 129–133). Springer.
9. Mohamed Moubark, M. F., & Ng Sok Choo, A. (2025). A comprehensive review of recent advances, persistent challenges and future directions in electronic voting systems (2020–2025). *Journal of Engineering & Technological Advances, 10*(2), 106–121. https://doi.org/10.35934/segi.v10i2.168
10. Noriega Zapata, J. P., Vargas, A., Castañeda, J., & Bernuy, A. (2026). Towards more inclusive elections: State of the art and challenges of electronic voting for people with disabilities in developing countries. *Disability, CBR & Inclusive Development, 37*(1), 82–93. https://doi.org/10.20372/dcidj.859
11. Ullah, A., Anwar, S. M., Li, J., Nadeem, L., Mahmood, T., Rehman, A., & Saba, T. (2024). Smart cities: The role of Internet of Things and machine learning in realizing a data-centric smart environment. *Complex & Intelligent Systems*, *10*(1), 1607-1637.