FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 1, January, 2026, pp 321 – 325
DOI: https://doi.org/10.33003/fjs-2026-1001-4402

SURVEY OF DIGITAL TECHNOLOGIES ADOPTION PREDICTION IN GIDAN MADI USING MACHINE LEARNING MODELS

***[1]Usman Musa, [1]Aminu Rabiu Na'iya, [2]Mahmud Malami Shallah and [1]Sagir Musa**

[1]Department of Computer Science Education, Federal College of Education, Gidan Madi, Sokoto State, Nigeria.
[2]Department of Educational Psychology and Counselling, Federal College of Education, Gidan Madi, Sokoto State, Nigeria.

*Corresponding authors' email: usmanmusa@fcegm.edu.ng

ABSTRACT

Digital technologies are widely recognized as catalysts for economic growth, educational access, and social inclusion; however, their adoption remains persistently low in many rural communities in developing regions. In areas such as Gidan Madi, Sokoto State, Nigeria, limited digital literacy, poor infrastructure, and socio-economic constraints continue to widen the digital divide, restricting access to essential digital services. This study aims to develop a machine learning–based predictive framework to identify the key determinants of digital tool adoption in underserved, low-literacy rural communities and to support evidence-based digital inclusion initiatives. A household survey dataset comprising 1,450 records was analyzed using four supervised machine learning algorithms: Logistic Regression, Random Forest, k-Nearest Neighbors, and Gradient Boosting. The models were evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The results indicate that ensemble models significantly outperform single learners, with Random Forest achieving the highest predictive performance. Feature importance and correlation analyses reveal that digital literacy, mobile device ownership, and educational attainment are the strongest predictors of digital adoption, while demographic variables such as age, gender, and household size exert comparatively weaker influence. The findings suggest that low adoption rates in rural communities are driven more by structural and capacity-related barriers than by resistance to technology. In conclusion, this study demonstrates the effectiveness of machine learning in modeling complex adoption behaviors and provides a scalable, data-driven framework for identifying high-impact intervention points. The results offer practical insights for policymakers, development agencies, and stakeholders seeking to design targeted strategies that accelerate digital inclusion in marginalized rural settings.

Keywords: Digital inclusion, Digital literacy, Digital tool adoption, Machine learning, Rural communities

INTRODUCTION

Digital technology has become a fundamental aspect of modern society, influencing economic activities, education, healthcare, and governance. However, rural communities, particularly in developing regions, continue to face significant barriers to digital adoption (Afzal et al., 2023). The integration of mobile technologies, internet access, e-learning platforms, digital financial services, and agricultural innovations remains minimal in these areas due to socio-economic constraints, inadequate infrastructure, and limited digital literacy (Baraka, 2024). The digital divide is particularly evident in Gidan Madi, where a large proportion of the population lacks access to digital tools, further exacerbating economic and social inequalities.

The adoption of digital tools in rural areas is not merely a matter of availability but is deeply influenced by demographic characteristics, digital literacy levels, access to digital infrastructure, and societal attitudes towards technology(Kosasih & Sulaiman, 2024). While some individuals quickly embrace digital tools for communication, education, and financial transactions, others remain hesitant due to technological illiteracy, economic barriers, cultural perceptions, and mistrust of digital systems. Additionally, inadequate infrastructure, such as poor internet connectivity and unstable electricity supply, further hinders widespread digital integration (Heena & Nidhi, 2022).

Digital technology has become an essential driver of economic and social development, yet rural communities, particularly in underserved regions, face significant barriers to digital adoption(Okocha & Dogo, 2024). Access to mobile technologies, internet services, e-learning platforms, digital financial tools, and agricultural innovations remains limited due to socio-economic constraints, poor digital literacy, and infrastructural deficits(Chukwunonso et al., 2024.) . The digital divide is particularly evident in Gidan Madi, where low levels of digital engagement hinder economic opportunities, access to education, and financial inclusion. Understanding the key determinants influencing digital tool adoption in such resource-constrained environments is critical to designing effective interventions that promote digital inclusion (Afzal et al., 2023)

Adopting digital tools has become an essential factor in economic, educational, and social transformation(Gbadebo, 2024). However, in rural communities like Gidan Madi, digital adoption remains significantly low due to infrastructural and socio-economic(Tahmasebi, 2023). Understanding these barriers and predicting digital adoption patterns require a data-driven approach that surpasses traditional technology adoption models. Machine learning (ML) offers a powerful framework to analyze large datasets, identify hidden patterns, and generate actionable insights that can help policymakers and stakeholders drive digital inclusion (Farrokh. Tahmasebi, 2024). Rogers' Diffusion of Innovation (DOI) Theory explains adoption based on perceived advantages, compatibility, complexity, trialability, and observability (García-Avilés, 2020).Technology Acceptance Model (TAM) (Davis, 1989) highlights the importance of perceived usefulness and ease of use (Baraka, 2023).While these models provide theoretical insights, they lack predictive capability, do not account for rural infrastructure challenges, and are not adaptable to dynamic data environments (Gul Mazloum Yar et al., 2024.).

The aim is to develop a machine learning framework for predicting digital tool adoption in underserved rural communities, using Gidan Madi as a case study, to identify key determinants influencing adoption and support targeted digital inclusion initiatives.

Machine Learning Algorithms

Logistic Regression: is a fundamental predictive modeling technique widely used in HR analytics (Ponnuru, 2020). It is especially valuable when the outcome of interest is binary, such as whether an employee will stay or leave. Logistic Regression estimates the probability of the binary outcome based on one or more predictor variables (Shipe et al., 2019). In employee attrition analysis, factors like job satisfaction, compensation, and performance can serve as predictors to estimate the likelihood of an employee leaving the organization (Marín Díaz et al., 2023).

Logistic Regression estimates the probability of a binary outcome (e.g., employee attrition) using the logistic function (also known as the sigmoid function) (Wardhani & Lhaksmana, 2022)

Random forests: are different from standard trees in that for the latter each node is split using the best split among all variables(Liaw & Wiener, 2002). In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node (Kubus, 2019). It overcomes the potential problem of overfitting (where a model is too closely tailored to the training data) by aggregating predictions from multiple trees. In the context of HR analytics, Random Forests can provide more robust attrition risk assessments by reducing the variance and enhancing the generalizability of the model(Punnoose & Xavier, 2016).

k-Nearest Neighbor: is one of the easiest machine learning algorithms, it stores all possible cases and classifies new cases based on a metric of similarity as a distance function (Cunningham & Delany, 2021).The intuition behind Nearest Neighbor Classification is to classify data points based on the class of their nearest neighbors (Flores & Leiva, 2021) It is often useful to take more than one neighbor into account so the technique is more commonly referred to as k-Nearest Neighbor (k-NN) Classification.

Gradient Boosting (GB): is a regression algorithm that shares similarities with boosting (Alsheref et al., 2022). In essence, the core objective of gradient boosting, when applied to a given training dataset, is to approximate the value of a function (Otchere et al., 2022). This approximation is achieved through the minimization of the predicted value derived from a specific loss function, which establishes a connection between input instances represented as 'x' and their corresponding output values 'y,' as defined by the loss function L (y, F(x)) (Li et al., 2020). GB accomplishes this by generating a weighted sum of functions, thereby creating an additive estimate

MATERIALS AND METHODS

Study Area

The study focuses on rural communities in Tangaza Local Government Areas of Sokoto State, Nigeria. A structured survey was administered to households, capturing demographic, educational, economic, and digital access characteristics. The final dataset contained approximately 1,450 records. Dependent Variable: Digital Tool Adoption (Adopter / Non-Adopter) Independent Variables: Age, Gender, Marital Status, Household Size, Education Level, Income Source, Digital Literacy, Mobile Ownership

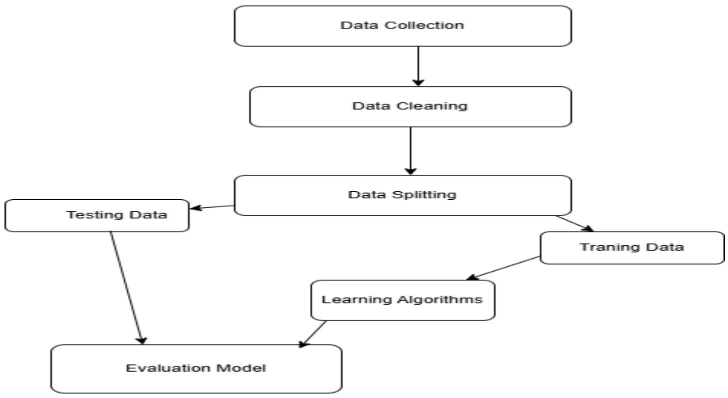

Figure 1: Research Design

This diagram outlines the research process and methodology used in the study. It includes key phases such as data collection, data cleaning and the application of machine learning models for predictive analysis. The diagram also illustrates the performance evaluation phase, highlighting the metrics used to assess the accuracy of the models.

Dataset Description

In this study, data were collected to analyze the adoption of digital tools in rural underserved communities, focusing on Gidan Madi, Sokoto State, Nigeria. The dataset encompasses multiple socio-economic, demographic, and technological variables that influence digital adoption, ensuring a comprehensive understanding of the factors at play. The dataset includes responses from individuals across different backgrounds, covering attributes such as age, gender, education level, occupation, income level, digital literacy, access to digital infrastructure, mobile and internet usage, and engagement with digital services. The expected feature diagnostic will categorize individuals into two classes: "Adopted" and "Not Adopted" digital tools.

The characteristics of the dataset are described as follows:

Table 1: Dataset Attribute

S/N	Feature	Specification	Value
1	RESP_ID	Unique indentifier for each respondent	Alphanumeric
2	AGE	Age (in years)	15-90
3	GENDER	Gender of Respondent	Male, Female

S/N	Feature	Specification	Value
4	MARITAL_STATUS	Marital status	Single, Married, Divorced, Widowed
5	HOUSEHOLD_SIZE	Number of people in household	1-30
6	EDU_LEVEL	Education Level	None, Primary, Secondary, Tertiary
7	OCCUPATION	Occupation type	Farmer, Trader, Student, Civil Servant, Other
8	INCOME	Monthly income level (in Naira)	0-100,000+
9	INCOME_SOURCE	Primary source of income	Farming, Trading, Employment, Remittances, Other
10	DIGITAL_LIT	Digital literacy level	Low, Moderate, High
11	MOBILE_OWN	Owns a mobile device	Yes, No
12	MOBILE_TYPE	Type of mobile device owned	Feature Phone, Smartphone, None
13	INTERNET_ACC	Has access to the internet	Yes, No
14	INTERNET_TYPE	Type of internet connection used	Mobile Data, Wi-Fi, Public Access, None
15	SOCIAL_MEDIA	Uses social media platforms	Yes, No
16	SOCIAL_MEDIA_FREQ	Frequency of social media usage	Daily, Weekly, Monthly, Rarely
17	MESSAGING_APP	Uses messaging apps (e.g., WhatsApp)	Yes, No
18	E_BANKING	Uses electronic banking services	Yes, No
19	E_BANKING_FREQ	Frequency of e-banking usage	Daily, Weekly, Monthly, Rarely
20	ONLINE_SHOPPING	Engages in online shopping	Yes, No
21	ONLINE_SHOPPING_FREQ	Frequency of online shopping	Daily, Weekly, Monthly, Rarely
22	E_LEARNING	Participates in online learning	Yes, No
23	E_LEARNING_FREQ	Frequency of online learning participation	Daily, Weekly, Monthly, Rarely
24	CHALLENGES	Main challenges in digital adoption	Cost, Connectivity, Literacy,
26	CLASS	Digital adoption category	Adopted, Not Adopted

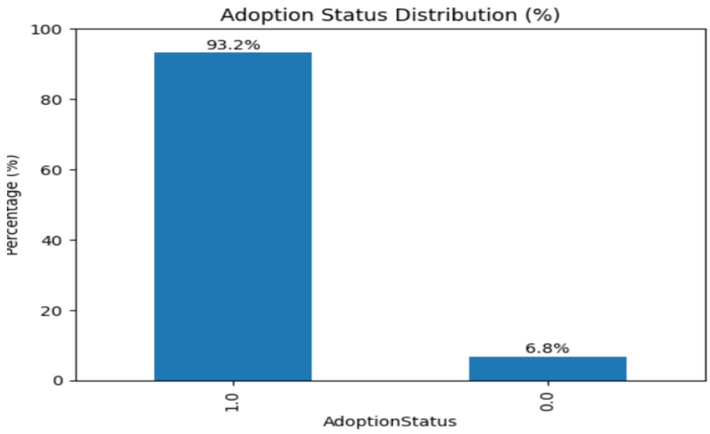

Figure 2: Adoption Status

Percentage-based visualizations revealed that adoption rates were higher among respondents with formal education, mobile ownership, and basic digital literacy. Gender and marital status showed weaker associations

Performance Evaluation Metrics

In assessing the performance of our predictive model, we have carefully considered the unique characteristics of our problem, the nature of our dataset, and our primary objectives. Given the context of our analysis, we have opted to use the following performance evaluation metrics:

Accuracy: Accuracy measures the proportion of correctly classified instances out of all predictions made by the model. It indicates how often the model correctly predicts both employees who leave and those who stay (Ponnuru, 2020).

Precision: Precision measures the proportion of employees predicted to attrite who actually left the organization. A high precision indicates a low false-positive rate (Ahmad,2023).

Recall: Recall measures the proportion of actual attrition cases that are correctly identified by the model. High recall indicates that most employees who left are correctly detected (Fallucchi,2020).

F1-score: The F1-score is the harmonic mean of precision and recall, providing a balance between them. It is especially useful when dealing with imbalanced datasets, such as employee attrition (Ponnuru,2020).

Table 2: Performance Matric

Metric	Description	Formula
Accuracy	Number of correct predictions from all predictions	TP+TN TP+FP+TN+FN
Sensitivity	Proportion of positive predictions that are correctly identified	TPTP+FN
Spacificity	Proportion of negative predictions that are correctly identified	TNFP+TN

F-Measure	Harmonic average of precision and sensitivity	(2*Precision*recall) Precision + recall

RESULTS AND DISCUSSION

This section provides a detailed analysis of the experimental outcomes obtained by applying different models to the dataset. The study utilized four selected machine learning algorithms: Random Forest, k-Nearest Neighbors (k-NN), Logistic Regression, Gradient Boosting Classifier (GBC). These algorithms were chosen for their robust classification capabilities and diverse approaches to handling data structures. The dataset used in this study comprised 1,450 instances and 20 attributes.

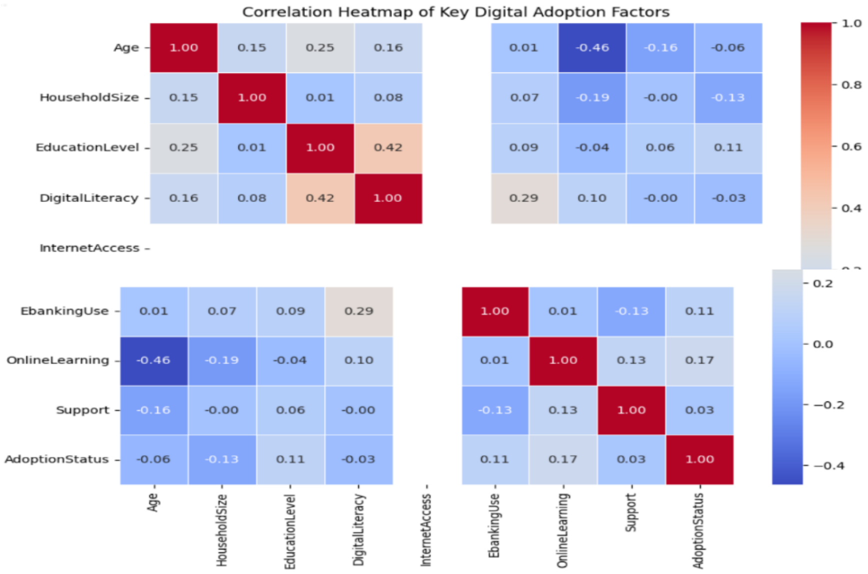

Figure 3: Correlation Heatmap

The correlation heatmap showed strong positive relationships between digital adoption and digital literacy, mobile ownership, and education level. Age exhibited a mild negative correlation, suggesting lower adoption among older respondents. Household size and marital status demonstrated minimal influence, indicating that adoption is primarily driven by individual capacity and access rather than household structure.

Model Performance

Ensemble models (Random Forest and Gradient Boosting) outperformed Logistic Regression and k-NN across most metrics. Random Forest achieved the highest ROC-AUC, indicating strong discriminative ability. Logistic Regression provided interpretability but lower predictive power.

Table 3: Model Results

S/N	Model	Accuracy	Recall	FI	ROC_AUC
1	Logistic Regression	90.09	87.27	89.72	96.82
2	Random Forest	95.40	94.55	95.41	99.81
3	K-nearest Neighbor	80.18	60.00	75.00	92.38
4	Gradient Bosting	93.69	90.90	93.46	98.54

CONCLUSION

This study empirically demonstrates the suitability and robustness of machine learning approaches for predicting digital tool adoption in low-literacy rural communities. By leveraging multiple classification algorithms, the findings reveal that ensemble models consistently outperform single learners in capturing the complex, non-linear relationships between socio-economic, demographic, and infrastructural factors that influence adoption behavior. This underscores the value of advanced data-driven techniques in understanding technology uptake within underserved populations. The exploratory and predictive analyses further indicate that digital literacy, mobile device ownership, and educational attainment are the most influential determinants of adoption. These factors exert a stronger impact than age, gender, or household size, highlighting that access to skills and enabling infrastructure remains the primary barrier to meaningful digital inclusion. Importantly, the results confirm that low adoption rates in rural communities are not solely a function of resistance to technology but are largely shaped by structural and educational constraints. Overall, this study contributes to the growing body of literature on digital divide research by integrating machine learning with socio-economic analysis in a low-resource setting. The findings provide empirical evidence that can support data-driven policymaking and targeted interventions aimed at accelerating digital inclusion. By translating predictive insights into practical strategies, this research offers a scalable framework for governments, development agencies, and non-governmental organizations seeking to improve digital participation in marginalized rural communities.

ACKNOWLEDGEMENTS

This research was funded by TETFUND under the Institution Based Research (IBR) Annual Intervention, we therefore acknowledge their immense support.

REFERENCES

Afzal, A., Khan, S., Daud, S., Ahmad, Z., & Butt, A. (2023). Addressing the digital divide: Access and use of technology in education. *Journal of Social Sciences Review, 3*(2), 883–895. https://doi.org/10.54183/jssr.v3i2.326

Ahmad, S., Iliyasu, U., & Jamilu, B. A. (2023). Enhanced predictive model for schistosomiasis. *FUDMA Journal of Sciences, 7*(3), 288–292. https://doi.org/10.33003/fjs-2023-0703-801

Alsheref, F. K., Fattoh, I. E., & Mead, W. (2022). Automated prediction of employee attrition using ensemble model based on machine learning algorithms. *Computational Intelligence and Neuroscience, 2022*. https://doi.org/10.1155/2022/7728668

Baraka, K. (2024). Digital divide and social inequality. *International Journal of Humanity and Social Sciences, 3*(3).

Chukwunonso, R., & Nosike, J. O. (2024). Digitalization in developing countries: Opportunities and challenges. *Nigerian Journal of Arts and Humanities (NJAH, 4).* https://www.researchgate.net/publication/378802704

Cunningham, P., & Delany, S. J. (2021). K-nearest neighbour classifiers—A tutorial. *ACM Computing Surveys, 54*(6). https://doi.org/10.1145/3459665

Farrokh, T. (2024). Data-driven decision-making: Transforming management in the information age. *International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/irjmets49577

Flores, V., & Leiva, C. (2021). A comparative study on supervised machine learning algorithms for copper recovery quality prediction in a leaching process. *Sensors, 21*(6), 1–21. https://doi.org/10.3390/s21062119

García-Avilés, J. A. (2020). Diffusion of innovation. In *The International Encyclopedia of Media Psychology* (pp. 1–8). Wiley. https://doi.org/10.1002/9781119011071.iemp0137

Gbadebo, A. D. (2024). Digital transformation for educational development in Sub-Saharan Africa. *International Journal of Social Science and Religion (IJSSR), 5*(3). https://doi.org/10.53639/ijssr.v5i2.262

Gul Mazloum Yar, F., Yasouri Yasoori, M., & Yasouri, M. (n.d.). Rural development challenges in addition to effective solutions to overcome obstacles. https://doi.org/10.1654/zkdx.2024.29.3-6

Heena, C., & Nidhi, B. (2022). Barriers affecting the effectiveness of digital literacy training programs (DLTPs) for marginalised populations: A systematic literature review. *Journal of Technical Education and Training, 14*(1), 110–127. https://doi.org/10.30880/jtet.2022.14.01.010

Kosasih, A., & Sulaiman, E. (2024). Digital transformation in rural settings: Unlocking opportunities for sustainable economic growth and community empowerment. *Journal of Sustainable Tourism and Entrepreneurship, 5*(2), 129–143. https://doi.org/10.35912/joste.v5i2.2278

Kubus, M. (2019). The problem of redundant variables in random forests. *Acta Universitatis Lodziensis. Folia Oeconomica, 6*(339), 7–16. https://doi.org/10.18778/0208-6018.339.01

Li, W., Wang, W., & Huo, W. (2020). RegBoost: A gradient boosted multivariate regression algorithm. *International Journal of Crowd Science, 4*(1), 60–72. https://doi.org/10.1108/IJCS-10-2019-0029

Liaw, A., & Wiener, M. (2002). Classification and regression by RandomForest. https://www.researchgate.net/publication/228451484

Maria Nurhajati Widjaja, A., Sanjaya, M. H., Fitriati, R., Fitriana, F. W., Keloko, A. B., Sakit Umum Daerah Kabupaten Bekasi, R., Farmasi Bhumi Husada, A., & Widya Dharma Husada Tanggerang, S. T. I. K. (n.d.). Digital health technologies in improving access to care for underserved populations. https://thejoas.com/index.php/

Okocha, D. O., & Dogo, J. S. (n.d.). Digital inclusion in rural areas: Qualitative exploration of challenges faced by people from isolated communities in Southern Kaduna. www.asric.africa

Otchere, D. A., Ganat, T. O. A., Ojero, J. O., Tackie-Otoo, B. N., & Taki, M. Y. (2022). Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *Journal of Petroleum Science and Engineering, 208*. https://doi.org/10.1016/j.petrol.2021.109244

Ponnuru, S. R. (2020). Employee attrition prediction using logistic regression. *International Journal for Research in Applied Science and Engineering Technology, 8*(5), 2871–2875. https://doi.org/10.22214/ijraset.2020.5481

Punnoose, R., & Xavier, X. (2016). Prediction of employee turnover in organizations using machine learning algorithms: A case for extreme gradient boosting. *International Journal of Advanced Research in Artificial Intelligence, 5*(9). www.ijarai.thesai.org

Shipe, M. E., Deppen, S. A., Farjah, F., & Grogan, E. L. (2019). Developing prediction models for clinical use using logistic regression: An overview. *Journal of Thoracic Disease, 11*(Suppl 4), S574–S584. https://doi.org/10.21037/jtd.2019.01.25

Tahmasebi, F. (2023). The digital divide: A qualitative study of technology access in rural communities. *AI and Tech in Behavioral and Social Sciences, 1*(2), 33–39. https://doi.org/10.61838/kman.aitech.1.2.6

Wardhani, F. H., & Lhaksmana, K. M. (2022). Predicting employee attrition using logistic regression with feature selection. *Sinkron, 7*(4), 2214–2222. https://doi.org/10.33395/sinkron.v7i4.11783