IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

AI-Driven Financial Fraud Detection

Sakshi Dipak Bakare[1], Lalita Balu Chinchole[2], Ratna Sadashiv Chaudhari[3]
Department of Computer Science and Applications[1,2,3]
K.R.T. Arts, B.H. Commerce, and A.M. Science (KTHM College), Nashik, Maharastra,India
sakshibakare0109@gmail.com[1],lalitachinchole29@gmail.com[2],ratnachaudhari@kthmcollege.ac.in[3]

Abstract: *The rapid growth of digital payment systems has led to a significant increase in financial fraud, making accurate and real-time detection essential. Machine learning techniques have proven highly effective in identifying payment-related fraud by detecting hidden and previously unseen patterns in large-scale transactional data. In this study, multiple machine learning models, including Logistic Regression, Decision Tree, Random Forest, XGBoost, and Naïve Bayes, are applied to a labeled dataset of online payment transactions. Due to the highly imbalanced nature of fraud data, appropriate preprocessing techniques are used to enhance model performance. The models are evaluated using key metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that ensemble models, particularly XGBoost and Random Forest, outperform other approaches by achieving higher accuracy and reducing false positives. The proposed approach highlights the effectiveness of AI-based systems in improving fraud detection, minimizing financial risks, and enhancing the security and reliability of digital payment platforms.*

Keywords: Financial Fraud Detection, Machine Learning, AI, Imbalanced Data, XGBoost

I. INTRODUCTION

The use of digital payment systems and online transactions has increased rapidly in recent years. While this has made transactions faster and more convenient, it has also increased the risk of financial fraud.Credit card and payments companies are experiencing a very rapid growth in their transaction volume. In third quarter of 2018, PayPal Inc (a San Jose based payments company) processed 143 billion USD in total payment volume [4]. Fraud in online payments is becoming more advanced, which makes it difficult to detect using traditional methods. Rule-based systems are not very effective because they cannot easily adapt to new types of fraud and often give many false alarms.To solve this problem, researchers are using artificial intelligence (AI) and machine learning (ML) techniques for fraud detection. These techniques can analyze large amounts of transaction data and find hidden patterns that indicate fraudulent activity. Many machine learning models, such as Logistic Regression, Decision Tree, Random Forest, and XGBoost, have been successfully used for this purpose. Among these, ensemble models like Random Forest and XGBoost often give better results.One major challenge in fraud detection is that the dataset is imbalanced. This means that fraudulent transactions are very few compared to normal transactions. Because of this, models may fail to correctly identify fraud cases. Some research studies have improved performance by using feature selection techniques such as Recursive Feature Elimination (RFE) along with models like XGBoost. Initial results demonstrate that traditional classifiers like XGBoost and Random Forest offer superior performance in both datasets, achieving high accuracy, precision, and recall[7]. Even though AI-based methods improve fraud detection, problems like false positives and difficulty in adapting to new fraud patterns still exist. Therefore, there is a need to develop better and more efficient models for accurate fraud detection. In this research, different machine learning models such as Logistic Regression, Decision Tree, Random Forest, XGBoost, and Naïve Bayes are used to detect fraudulent transactions. The models are evaluated using accuracy, precision, recall, and F1-score to find the best-performing method. We compare the effectiveness of these approaches in detecting fraud transactions.[4] This study helps in improving the security of digital payment systems by providing a reliable fraud detection approach. By combining the strengths of XGBoost this research aims to make a significant contribution to improving the security of online transactions and protecting users and businesses

Copyright to IJARSCT
www.ijarsct.co.in

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, May 2026 Impact Factor: 8.2

from fraud threats[1].As digital transactions continue to expand globally, the role of AI in fraud detection will become increasingly essential. Therefore, continuous research and technological innovation are necessary to build secure, transparent, and trustworthy financial ecosystems for the future.

II. LITERRATURE REVIEW

Financial fraud detection has become an important research area due to the rapid increase in online transactions and digital payment systems. Traditional rule-based systems are less effective in detecting modern fraud patterns, so researchers are using machine learning techniques for better accuracy and real-time detection.

J.S Wibowo, B. Hartono, V. Lusiana, "Online Payment Fraud Detection Optimization with XG Boost and Recursive Feature Elimination," (2024) proposed a fraud detection system using XGBoost and Recursive Feature Elimination (RFE). Their study showed that feature selection improves fraud detection performance and helps in handling imbalanced datasets.

M. Venkatesh, B.K. Bai, B. Bhargavi, C. Manasa, and D. Mokshitha, "Online Payment Fraud Detection sing Machine Learning," (2024) applied machine learning models such as Logistic Regression, Decision Tree, and Random Forest for online payment fraud detection. Their results showed that ensemble models provide better performance and higher accuracy compared to traditional methods.

abdulwahab ali almazroi and nasir ayub "Online Payment Fraud Detection Model Using Machine Learning Techiniques." (2023) developed a machine learning-based fraud detection model and found that advanced algorithms reduce false positives and improve fraud detection efficiency.

Jin, J., & Zhang, Y. "The analysis of fraud detection in financial market under machine learning. Scientific Reports" (2025). analyzed the effectiveness of machine learning algorithms in financial fraud detection. Their research compared Logistic Regression, Decision Tree, and Random Forest models and found that ensemble learning methods achieved better fraud classification accuracy and reliability in financial datasets.

L. Theodorakopoulos, A. Theodoropoulou, A. Tsimakis, and C. Halkiopoulos, "Big data-driven distributed machine learning for scalable credit card fraud detection using PySpark, XGBoost, and CatBoost," (2025) developed a distributed machine learning framework using PySpark, XGBoost, and CatBoost for scalable credit card fraud detection. The study showed that ensemble boosting techniques achieved high accuracy and improved scalability for large financial datasets.

R. Singh, "Enhancing financial fraud detection through machine learning: A comparative study of anomaly detection and classification models on imbalanced datasets," (2025) compared anomaly detection and classification models for fraud detection on imbalanced financial datasets. The study found that ensemble models such as Random Forest and XGBoost outperformed traditional machine learning algorithms in fraud classification accuracy.

F. A. Almarshad, M. Zakariah, G. A. Gashgari, and T. Vaiyapuri, "RABEM: Risk-adaptive Bayesian ensemble model for fraud detection," (2025)proposed a Bayesian ensemble framework called RABEM for financial fraud detection. The study demonstrated that Bayesian ensemble techniques improve fraud detection accuracy and adaptability in imbalanced financial datasets.

S. A. Ajagbe, S. Majola, and P. Mudali, "Comparative analysis of machine learning algorithms for money laundering detection," (2025) investigated the effectiveness of machine learning algorithms for detecting money laundering activities in financial transactions. The study evaluated several algorithms, including XGBoost, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN),and Isolation Forest, using anomaly detection techniques. The authors reported that XGBoost achieved the highest performance in terms of accuracy, precision, recall, and F1-score, demonstrating its effectiveness in identifying suspicious financial activities. The study highlighted that machine learning models can significantly enhance anti-money laundering (AML) systems by improving detection efficiency and reducing false positives in large-scale financial environments.

Based on previous studies, ensemble models like Random Forest and XGBoost are considered more effective for detecting fraudulent transactions in online payment systems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

III. METHODOLOGY

This research follows a machine learning-based approach to detect fraudulent transactions in online payment systems. The methodology consists of several steps, including data collection, preprocessing, model building, and evaluation.

3.1 Data Collection

In this project, we have used a Kaggle provided dataset of simulated mobile based payment transactions. Similarly, Kaggle's Synthetic Financial Datasets for Fraud Detection also serve as a valuable resource for training and evaluating fraud detection algorithms[8] .We analyze this data by categorizing it with respect to different types of transactions it contains. The dataset contains five categories of transactions labeled as 'CASH IN', 'CASH OUT', 'DEBIT', 'TRANSFER' and 'PAYMENT'- details are provided in table.

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0	0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1	0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0	0

Fig. 1 Online payment dataset statistics

Dataset Analysis

Dataset Name	Source	Description	Size	Fraud Instances
Online Payment Fraud Detection Dataset	Kaggle	A dataset simulating online payment transaction, including various types like CASH_OUT and TRANSFER, to detect fraudulent activities.	1,048,575 transactions	8,213

TABLE I . Summary Table of Dataset

3.3 Data Preprocessing

After collecting the original data, strict data cleaning measures are taken to ensure the accuracy and integrity of the data. Including removing duplicate records to eliminate data redundancy, processing missing values and selecting filling, interpolation or deletion methods according to data characteristics and business logic, and using statistical methods and business domain knowledge to detect and process abnormal values, thus reducing the negative impact of noise data on subsequent modeling.[8]. Handling missing values and duplicate records. Encoding categorical variables using techniques like one-hot encoding. Normalizing numerical features to improve model performance. Splitting datasets into training, validation, and testing sets(e.g.,70%-15%-15%). Addressing class imbalance.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

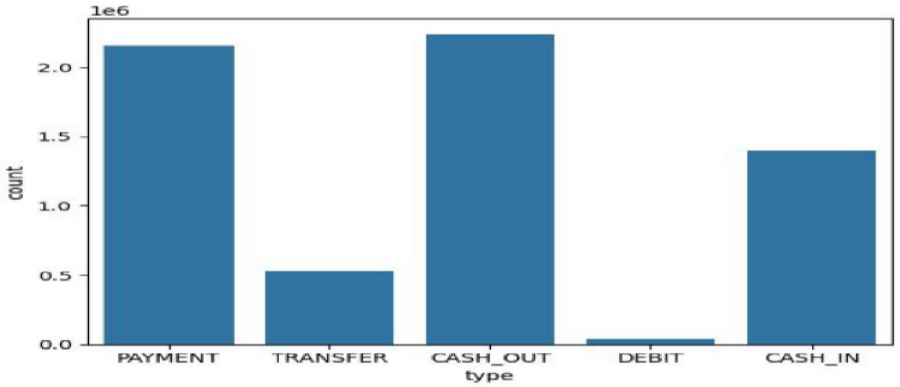

Fig. 2 Distribution of Transaction Types

Given bar chart shows how many times a transaction type appears in the dataset. categories we can see:
PAYMENT and CASH_OUT (appear the most)- These transaction types are very common in our dataset.
DEBIT (has very few values)- This type is rare.
CASH_IN and TRANSFER (appear some time)- Have moderate counts.
This graph is used to understand the type of transactions and find which ones need more focus in fraud detection .it helps to understand distribution of transaction types. important for fraud detection because some transaction types (like TRANSFER and CASH_OUT) are often involved in fraud cases.

3.4 Model Implementation

Different machine learning models are applied to detect fraud:
Logistic Regression
Decision Tree
Random Forest
XGBoost
Naïve Bayes
These models are trained using the prepared dataset. Each model learns patterns to classify transactions as fraud or non-fraud.

3.5 Data Splitting

The dataset is divided into two parts:
Training data (80%) → used to train the model
Testing data (20%) → used to evaluate performance

Dataset Type	Percentage
Training Set	80%
Testing Set	20%

TABLE II. Training & Testing Split

3.6 Model Evaluation

After training the models, the next step is to test and evaluate their performance using a separate validation or test dataset that was not used during training. this ensures that the results reflect how well the model will generalize to new, unseen fraud cases. the evaluation focuses on metrics that are critical for fraud detection, including accuracy, precision, recall, F1-score, ROC-AUC.The terms for the algorithm evaluation were based on the state-of-the-arts measure such as accuracy, F1 scores, recall, precision, confusion matrix and ROC. [9]. The confusion matrix is also examined to

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, May 2026 Impact Factor: 8.2

understand the distribution of true positives, false positives, true negatives, false negatives--where minimizing false negatives is particularly important because undetected fraud can lead to significant financial loss. Descriptions of these performance metrics are provided below.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

This metric measures the number of correct predictions made by model in relation to the total number of predictions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision calculates the proportion of all positive predictions that were actually correct, whether positive or negative.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall measures the proportion of actual positive instances that a model correctly identified.

$$\text{F1- Score} = 2 \text{ X } \frac{\text{Precision X Recall}}{\text{Precision + Recall}}$$

F1-score is a performance metric used in classification problems, especially when the dataset is imbalanced. it balances both precision and recall.

ROC = The (Receiver Operating Characteristics) curve is an analytical method, represents as a graph, that is used to evaluate the performance of a binary diagnostic classification method. it is a graph that plots a classifier's performance at all possible threshold values by showing the trade-off between the true positive rate and the false positive rate.

AUC = The (Area Under Curve) curve, is a performance measurement for classification models that graphically represents how well a model can distinguish between classes across all possible thresholds.

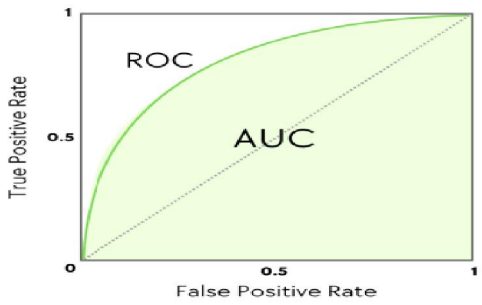

Fig. 2 AUC- ROC Analysis

IV. EXPERIMENTS

In this research, four machine learning algorithms-Logistic Regression, Decision Tree, Random Forest, and XGBoost were applied to an online dataset which is "Online Payment Fraud Detection". The dataset contains multiple transaction-related attributes along with a target variable indicating whether a transaction is fraudulent. All experiments were conducted using python in as online Jupyter Notebook environment on a windows operating system. Libraries such as Numpy, panadas, scikit-learn, matplotlib. NumPy and pandas were used for data preparation, while Scikit-learn and XGBoost were used to build and train the models. each algorithm was evaluated using accuracy, precision, recall, F1-score and confusion matrix. based on these evaluation metrics, the performance of all four algorithms was compared to identify the most effective model for detecting fradulent transactions in the given dataset.

Graph illustrates the performance of four different machine learning models based on their accuracy scores in percentage

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

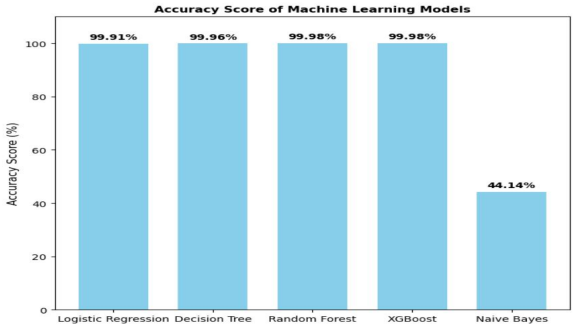

Fig. 3 Accuracy score of ML models

On the X-axis, the graph lists the five machine learning models evaluated: Logistic Regression, Decision Tree, Random Forest, XGBoost, Naïve Bayes. The Y-axis represents the accuracy score as a percentage, ranging from 0% to 100%, showing how accurately each model predicts the correct class in dataset. From the graph, Logistic Regression achieves an accuracy of 99.91%, showing that the dataset fits very well with a linear model. Decision Tree model performs slightly better with 99.96%, indicating its ability to capture non-linear patterns in data.Random Forest and XGBoost both achieve the highest accuracy of 99.98%, which demonstrates the strength of ensemble learning and boosting techniques within dataset.In contrast, Naïve Bayes model shows significantly lower accuracy of 44.14%, suggesting that its assumption of feature independence does not hold true for this dataset. Overall graph clearly shows that Random Forest and XGBoost provide superior performance.

Graph illustrates the error rate of five different machine learning models based on their accuracy scores in percentage.

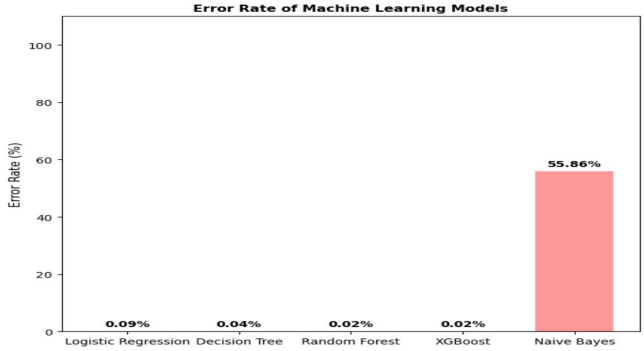

Fig. 4 Error rate of ML models

On the X-axis, the graph lists the five machine learning models evaluated: Logistic Regression, Decision Tree, Random Forest, XGBoost, Naïve Bayes. The Y-axis represents the error rate as a percentage, ranging from 0% to 100%, showing how many predictions each model classified incorrectly. From the graph Logistic Regression shows a very low error rate of 0.09%, indicating that the model made very few incorrect predictions and performed reliably on dataset. Decision Tree model performs even slightly better with an error rate of 0.04%, suggesting its ability to capture underlying patterns more accurately. Both Random Forest and XGBoost achieve the lowest error rates of 0.02%, highlighting the effectiveness of ensemble and boosting techniques in minimizing prediction errors. These models ssuccessfully reduce overfitting and handle complex relationship between features, resulting in highly accurate outcomes. In contrast Naïve Bayes model exhibits a significantly higher error rate of 55.86%, showing that more than half of its predictions are incorrect. Overall graph clearly demonstrates ensemble models such as Random Forest and XGBoost outperform other models by achieving minimal error, while Naïve Bayes performs poorly for this classification.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

Graph illustrates the ROC (Receiver Operating Characteristic) curves for five machine learning models used in fraud detection.

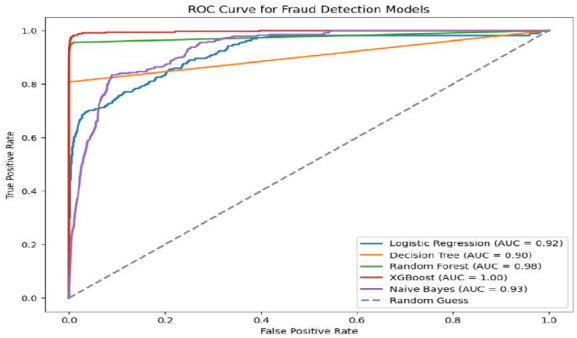

Fig. 5 ROC Curve for fraud detection models

The X- axis shows the False Positive Rate, while the Y-axis shows the True Positive Rate, indicating how well each model distinguishes between fraudulent and legitimate transactions. A curve closer to the top-left corner demonstrates higher predictive performance. among the models, XGBoost performs the best with an AUC of 1.00, showing almost perfect classification with very high true positive rates and minimal false positives.Random Forest also performs strongly with an AUC of 0.98, indicating it is highly reliable for fraud detection. Logistic Regression (AUC=0.92), Naïve Bayes (AUC=0.93), Decision Tree (AUC=0.90) show comparatively lowe but still effective performance. The diagonal dashed line represents random guessing with an AUC of 0.5, which acts as a baseline. Overall the graph clearly indicates that ensemble-based models like XGBoost and Random Forest outperform simpler models in identifying fraud accurately

Graph illustrates the comparison of confusion matrix components for five machine learning models.

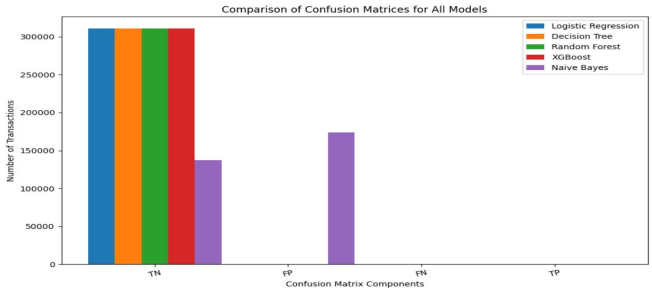

Fig. 6 Comparison of confusion matrices for all models

The Y-axis represents the confusion matrix components (TN,FP,FN,TP) and the X-axis represents the number of transactions corresponding to each component. From the graph it is evident that all models achieve a very high number of True Negatives (TN), meaning they correctly identify most genuine transactions . However, significant variation is observed in the False Positive (FP) and False Negative (FN) counts across models. Naïve Bayes shows comparatively higher FP and FN values, indicating a weaker ability to distinguish between fraudulent and legitimate transactions. In contrast, Logistic Regression, RandomForest, and XGBoost show very low FP and FN values, reflecting superior accuracy and fewer misclassification.Overall the graph clearly highlights the stronger performance of ensemble based models, especially Random Forest and XGBoost, while Naïve Bayes performs the weakest among the compared algorithms.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

Model Comparison

Models	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.9990	0.9827	0.1691	0.2886
Decision Tree	0.9995	0.8168	0.8071	0.8119
Random Forest	0.9997	0.9855	0.8100	0.8892
XGBoost	**0.9998**	**0.9828**	**0.8486**	**0.9108**
Naïve Bayes	0.4414	0.0019	1.0000	0.0038

TABLE III. Model comparison table

The model comparison table evaluates five machine learning algorithms with using performance metrics: accuracy, recall, precision, F1 score. The results show that Logistic Regression, Decision Tree, Random Forest, and XGBoost all achieve very high accuracy above 99%, while Naïve Bayes performs poorly with only 44%. The F1 Score which balances precision and recall, further confirms XGBoost as the best performing model with highest F1 score, followed closely by Random Forest and Decision Tree, while Logistic Regression performs modestly and Naïve Bayes shows very weak performance. Overall, XGBoost emerges as the most effective model for online payment fraud detection based on F1 score and overall metric balance.

Best Model Based on F1 Score

Model	XGBoost
Accuracy	0.9998
Precision	0.9828
Recall	0.8486
F1 Score	0.9108

TABLE IV. Best model table based on F1 score

V. RESULTS & DISCUSSION

In this study, multiple machine learning models were applied to detect fraudulent transactions in an online payment dataset. The models were evaluated using performance metrics such as accuracy, precision, recall, and F1-score to measure their effectiveness. The dataset used in this research is highly imbalanced, where the number of legitimate transactions is significantly higher than fraudulent ones. Due to this imbalance, accuracy alone is not a reliable metric. Therefore, more important metrics such as precision, recall, and F1-score were used to evaluate the performance of the models. Among all the models, XGBoost achieved the best performance in detecting fraudulent transactions. It provided higher accuracy and a better balance between precision and recall, while also minimizing false positives. Random Forest also performed well and showed strong capability in handling imbalanced data due to its ensemble nature. Decision Tree showed moderate performance but was less stable compared to ensemble models. Logistic Regression and Naïve Bayes were faster and simpler models, but they were less effective in capturing complex patterns in the data, which affected their fraud detection capability. The results indicate that ensemble models such as XGBoost and Random Forest are more suitable for fraud detection tasks, especially when dealing with imbalanced datasets. These models are capable of identifying complex patterns and improving overall detection accuracy. Overall, the analysis demonstrates that machine learning techniques are effective in detecting fraudulent transactions, with XGBoost being the most efficient model in this study. However, challenges such as false positives and the need to adapt to evolving fraud patterns still remain, which can be addressed in future research.

VI. CONCLUSION

In this research, a machine learning-based approach was used to detect fraudulent transactions in online payment systems. Different models, including Logistic Regression, Decision Tree, Random Forest, XGBoost, and Naïve Bayes,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, May 2026 Impact Factor: 8.2

were applied and compared to identify the most effective method for fraud detection. The results showed that ensemble models, especially XGBoost and Random Forest, performed better than other algorithms. These models were able to detect fraudulent transactions more accurately and handle the imbalanced dataset effectively. Simpler models such as Logistic Regression and Naïve Bayes showed moderate performance and were less effective in identifying complex fraud patterns. This study highlights the importance of using advanced machine learning techniques for improving fraud detection systems. By using appropriate models and evaluation metrics, it is possible to reduce financial losses and improve the security of digital payment platforms. However, some challenges still remain, such as false positives and adapting to new and evolving fraud patterns. Future work can focus on improving model performance by using deep learning techniques, better data balancing methods, and real-time fraud detection systems. Overall, this research demonstrates that AI-driven approaches can significantly enhance fraud detection and contribute to more secure and reliable financial systems.

VII. FUTURE WORK

Finally, next-generation systems must incorporate privacy-preserving machine learning solutions such as secure multi-party computation and homomorphic encryption to keep up with changing privacy demands. These solutions allow for secure, scalable fraud detection without sacrificing confidentiality or adherence to worldwide data protection regulations.[6]. Future work can focus on improving fraud detection by using deep learning techniques such as neural networks and real-time detection systems. Advanced methods for handling imbalanced data and reducing false positives can also be explored. Additionally, integrating adaptive models can help detect new and evolving fraud patterns more effectively in financial systems.

VIII. REFERENCES

[1] J.S Wibowo, B. Hartono, V. Lusiana, “Online Payment Fraud Detection Optimization with XG Boost and Recursive Feature Elimination,” J. Softw. Eng. Simul., vol.10, no.8, pp.35-42, 2024

[2] M. Venkatesh, B.K. Bai, B. Bhargavi, C. Manasa, and D. Mokshitha, “Online Payment Fraud Detection sing Machine Learning,” LJARIIE, vol.10, no.2, 2024

[3] abdulwahab ali almazroi and nasir ayub “Online Payment Fraud Detection Model Using Machine Learning Techiniques.” IEEE Access, vol.11,2023, pp.141940-141953, doi: 10.1109/ACCESS.2023.3339226.

[4] Oza, A. (2018). Fraud detection using machine learning https://cs229.stanford.edu/proj2018/repor/216.pdf

[5] Jin, J., & Zhang, Y. (2025). *The analysis of fraud detection in financial market under machine learning*. *Scientific Reports, 15*, 29959. https://doi.org/10.1038/s41598-025-15783-2

[6] L. Theodorakopoulos, A. Theodoropoulou, A. Tsimakis, and C. Halkiopoulos, “Big data-driven distributed machine learning for scalable credit card fraud detection using PySpark, XGBoost, and CatBoost,” *Electronics*, vol. 14, no. 9, p. 1754, 2025, doi: 10.3390/electronics14091754.

[7] R. Singh, “Enhancing financial fraud detection through machine learning: A comparative study of anomaly detection and classification models on imbalanced datasets,” in *Proceedings of the International Conference on Business and Intelligent Research*, Atlantis Press, 2025, doi: 10.2991/978-94-6463-906-3_18.

[8] F. A. Almarshad, M. Zakariah, G. A. Gashgari, and T. Vaiyapuri, “RABEM: Risk-adaptive Bayesian ensemble model for fraud detection,” *Scientific Reports*, vol. 15, Art. no. 36796, 2025, doi: 10.1038/s41598-025-20651-0.

[9] S. A. Ajagbe, S. Majola, and P. Mudali, “Comparative analysis of machine learning algorithms for money laundering detection,” *Discover Artificial Intelligence*, vol. 5, no. 1, p. 144, 2025, doi: 10.1007/s44163-025-00397-4.

[10] Ali, A., Razak, S. A., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). *Financial fraud detection based on machine learning: A systematic literature review*. Applied Sciences, 12(19), 9637. https://doi.org/10.3390/app12199637

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

[11] Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). *Financial fraud detection through the application of machine learning techniques: A literature review*. Humanities and Social Sciences Communications, 11, 1130. https://doi.org/10.1057/s41599-024-03606-0

[12] Akre, Z. R. (2024). *Financial fraud detection based on machine and deep learning: A review*. The Indonesian Journal of Computer Science, 13(3). https://doi.org/10.33022/ijcs.v13i3.4059

[13] Husnaningtyas, N., & Dewayanto, T. (2023). *Financial fraud detection and machine learning algorithm (unsupervised learning): Systematic literature review*. Jurnal Riset Akuntansi dan Bisnis Airlangga, 8(2). https://doi.org/10.20473/jraba.v8i2.49927

[14] Naqvi, S. H. (2026). *Fraud detection using machine learning in financial transactions: A systematic review*. Veredas do Direito, 23(2). https://doi.org/10.18623/rvd.v23.n2.4187

[15] Chen, Y., Zhao, C., Xu, Y., & Nie, C. (2025). *Year-over-year developments in financial fraud detection via deep learning: A systematic literature review*. arXiv. https://arxiv.org/abs/2502.00201

[16] Psychoula, I., Gutmann, A., Mainali, P., Lee, S. H., Dunphy, P., & Petitcolas, F. A. P. (2021). *Explainable machine learning for fraud detection*. arXiv. https://arxiv.org/abs/2105.06314