# Credit Card Fraud Detection Using Machine Learning

Mr. Chetan Bhankhede [1], Mr. Sumedh Khonde [2], Mr. Kalpak Ramidwar[3], Miss. Shravani Suddalwar[4,]
Miss. Latika Dagwar[5], Miss. Samrudhi Tatewar[6]
[1]*Department Of Computer Engineering*
[2]*Sant Gadge Baba Amravati University, Amravati*
[3]*Jagdambha College of Engineering and Technology, Yavatmal*

***Abstract*: Credit card fraud has emerged as a major problem because of the fast development of online and digital payment systems. It is difficult to identify fraudulent transactions efficiently because of the highly imbalanced nature of the transaction data and the ever-changing patterns of credit card fraud. This paper proposes a machine learning-based system for credit card fraud detection using classification algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM). The proposed system examines the characteristics of transactions to detect anomalies in real-time. The experimental results demonstrate that ensemble learning methods, particularly Random Forest, perform better than traditional approaches in terms of accuracy and fraud detection rates. The proposed system can help banks prevent financial losses and ensure transaction security.**

**Keywords: Credit Card Fraud Detection, Machine Learning, Random Forest, Classification, Imbalanced Dataset, Financial Security, Transaction Data Analysis, Fraud Prevention**

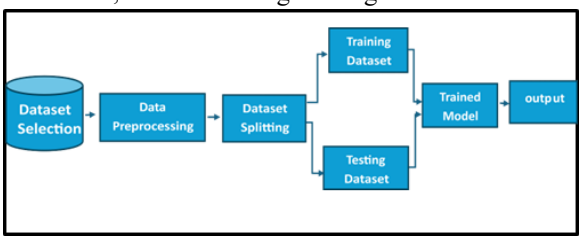
## I.INTRODUCTION

Credit card fraud involves unauthorized, deceptive use of another person's credit card information to make purchases or steal funds, resulting in billions in annual losses. As digital payments rise rapidly via e-commerce and contactless methods, fraud techniques become more sophisticated, rendering traditional, manual, rule-based systems ineffective. Machine Learning (ML) solves this by identifying complex, non-linear patterns in high-volume, real-time transaction data that human analysts cannot detect. This research aims to develop a robust, efficient ML model to accurately detect fraudulent transactions while minimizing false positives. What is Credit Card Fraud? It is a form of identity theft where fraudsters illegally obtain and use credit card information to charge transactions without the owner's consent, costing financial institutions and users billions annually. Growth of Digital Payments: Digital transactions are increasing exponentially, driven by e-commerce, contactless payments, and a shift towards a cashless economy. This rapid adoption creates more opportunities for fraud. Importance of Fraud Detection: Effective detection is crucial to prevent substantial financial losses for banks and consumers, protect customer trust, and ensure compliance with financial regulations. Challenges in Traditional Detection: Traditional systems rely on static, rule-based techniques that cannot keep pace with new fraud methods, leading to high false-positive rates (valid transactions flagged as fraud) and missed fraud cases (false negatives). How Machine Learning Helps: ML algorithms (e.g., Random Forest, Neural Networks) learn from historical data to identify hidden, complex, and evolving fraudulent patterns in real-time, adapting faster to new threats than traditional methods. Aim of the Research Paper: This paper aims to analyze, evaluate, and compare various machine learning algorithms to develop a high-accuracy model that can distinguish between legitimate and fraudulent transactions efficiently.

## II. MTHODOLOGY

This paper uses a structured approach to design a credit card fraud detection system using machine learning algorithms Fraud detection is a binary classification problem in which every transaction is classified as either fraud or legitimate. The dataset is available in open-source form from a Kaggle, which is a platform for shared a dataset for data analysis research. It contains 31 features, of which 28 variables

have been anonymized to maintain confidentiality. Amount are the amount of money transacted. Credit card fraud detection is a major financial problem that causes huge financial losses each year. To overcome this problem, a structured machine learning approach is used to identify fraud transactions from legitimate transactions using Machine Learning. The first method is the use of a different machine learning algorithm on the dataset. The dataset is pre-processed by dealing with the missing value, dealing with the class imbalance, and feature engineering to enhance the




performance of the model. The Kaggle is website for data analysis that provides the datasets, where we got the dataset, it's free to Open-Source Python library, which provides a variety of a simple and efficient tools for data analysis and machine learning. Using the Jupiter Notebook platform, we developed a python application to the method recommended in this paper. Machine learning (ML) techniques have been increasingly used in fraud detection due to their ability to automatically identify complex patterns in large datasets. Traditional rule based systems rely on human-defined rules, which are inflexible and not very effective in dealing with new and complex fraud patterns. On the other hand, ML-based systems learn patterns from past transaction data and continue to adapt to new fraudulent patterns. The trained system is then used to identify fraudulent patterns in new, unlabelled data. The hybrid approach enhances the effectiveness of the fraud detection system by using labelled data for training and also for identifying anomalies in unlabelled data. This helps in improving the detection of known fraudulent patterns as well as unknown fraudulent patterns, thus making the system effective.

1. Problem Definition and Goal Setting: The problem statement, objectives, and key performance indicators (KPIs) need to be defined. What is considered a fraudulent transaction, and what is the desired level of accuracy or detection rate.

2. Data Collection: Data collection is a significant aspect in the development of an efficient credit card fraud detection system. The quality, variability, and volume of data have a direct effect on the efficiency of the system in identifying fraudulent transactions. The following section discusses the key aspects and techniques involved in data collection for credit card fraud detection. The continuation provides a general insight into the key aspects and techniques of data collection in credit card fraud detection.

3. Model Development: Machine learning algorithms are trained on the prepared dataset to identify the patterns of fraudulent and genuine transactions. Based on the nature and availability of the labelled dataset, supervised, unsupervised, or deep learning models can be employed. The comparison of the results will be discussed in the latter part of this paper to identify the best approach for fraudulent credit card transactions4.Data Preprocessing; Duplicate observations are removed, errors are corrected, and inconsistent or inaccurate data is handled. Data preprocessing is a significant step before the application of machine learning algorithms since different machine learning algorithms require different handling of input variables, and the accuracy of the training data directly influences the accuracy of the predictive model. The primary objectives of data preprocessing include data cleaning, bias removal, handling missing data, and improving data variability. The dataset contains both numerical and categorical variables, and the categorical variables are encoded before the training of the model. Outlier deletion is also done to remove noise in the data. Feature scaling is done to ensure that the independent variables are measured on a common scale, and a Box-Cox transformation is used to eliminate feature skewness.

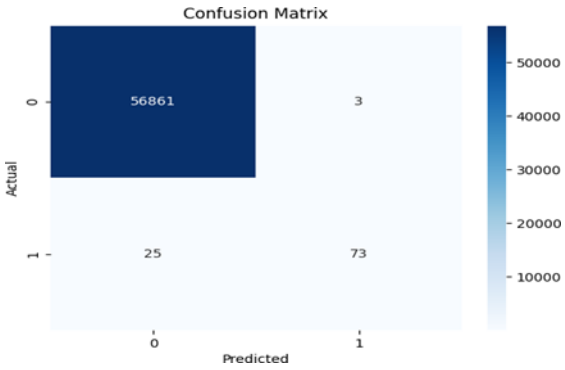
## III. MACHINE LEARNING MODEL

II.A. Model Selection the Random Forest Classifier is used for credit card fraud detection because it's robust and works well with complex data. It's an ensemble learning algorithm that combines multiple decision trees to make predictions. Each tree is trained on a random sample of data, and they vote on the final outcome. This approach helps reduce overfitting and improves performance. Random Forest is particularly useful for credit card fraud detection because it can

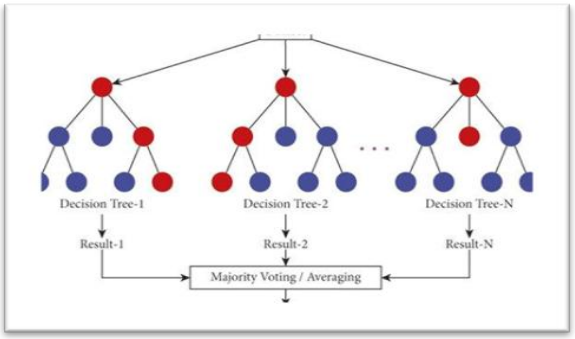
handle imbalanced datasets, where fraudulent transactions are rare. It also captures complex patterns and provides feature importance scores, making it a great choice for this research. This helps in understanding which factors contribute most to detecting fraud.

II.B. Working of Random Forest Algorithm Random Forest is a machine learning algorithm that uses ensemble learning to produce a single final classification output by creating multiple decision trees and aggregating their predictions. The first step in building a Random Forest model is to create multiple bootstrap sample sets from the original dataset by randomly selecting data points from the dataset using replacement. The original dataset is repeated multiple times to create multiple training sets, which will be used to train each individual decision tree When creating a decision tree using a bootstrap sample set, a random subset of features will be selected for consideration at each node while building the decision tree. Randomly selecting a subset of available features at each node will reduce the correlation between the individual decision trees and improve the overall performance of the model with respect to its ability to generalize to previously unseen data. Each individual decision tree is created to full depth without pruning in order to facilitate the identification of complex relationships in the training data. In the end, after all of the decision trees have been created, the predicted values produced by each decision tree are combined via a votingmechanism. In the case of classification tasks, the class with the most votes have the highest likelihood of being the correct class label. The ensemble nature.

following: Reduced variance allows for the Random Forest model to be much less sensitive to noise and



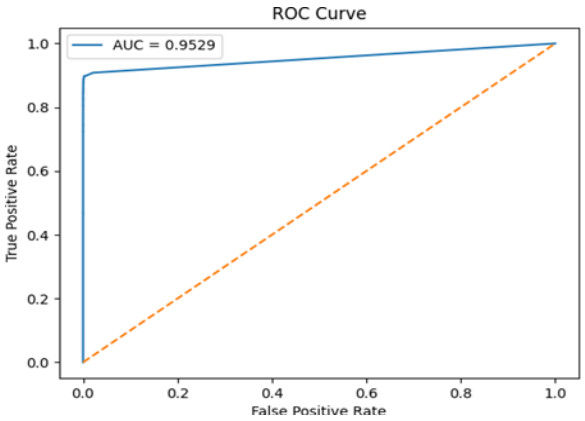

outliers. The likelihood of overfitting, a problem that can occur within a single decision tree, is reduced due to the use of multiple decision trees in the Random Forest. By utilizing multiple decision trees, the Random Forest is able to produce stable and accurate classification results, making it an attractive choice for many complex




distributions are skewed. C.5 Confusion Matrix The confusion matrix summarizes prediction outcomes in terms of true positives, true negatives, false positives, and false negatives, offering detailed insight into classification behavior.

## IV. RESULT AND DISCUSSION

Result: The performance of the proposed credit card fraud detection model was evaluated using multiple metrics to account for the highly imbalanced nature of the dataset. The Random Forest classifier demonstrated effective discrimination between fraudulent and legitimate transactions, achieving reliable performance across all selected evaluation measures. The model attained high precision, indicating a low false positive rate and reduced misclassification of legitimate transactions as fraud. Additionally, a strong recall value was observed, reflecting the model's capability to successfully identify a significant proportion of actual fraudulent transactions, which is critical for minimizing financial losses. The F1-score further confirmed a balanced trade-off between precision and recall, making the model suitable for imbalanced classification scenarios. The confusion matrix analysis revealed a substantial.




number of true positives and true negatives, with relatively fewer false negatives. Furthermore, the ROC-AUC score indicated strong class separability, demonstrating consistent performance across varying. satisfaction and avoiding unnecessary transaction blocks.

classification thresholds. Overall, the experimental results validate the effectiveness and robustness of the Random Forest model for credit card fraud detection

1. Discussion: The experimental results demonstrate that the Random Forest classifier is well-suited for credit card fraud detection, particularly in the presence of highly imbalanced data. The model achieves a favorable balance between precision and recall, indicating effective fraud identification while minimizing false alarms for legitimate transactions. This balance is crucial in real-world financial systems, where both missed frauds and unnecessary transaction blocks can have significant consequences. The high recall obtained by the model highlights its ability to detect a substantial portion of fraudulent transactions, thereby reducing potential financial losses. At the same time, strong precision values suggest that the model does not excessively misclassify genuine transactions as fraud. The F1-score further confirms the stability of the model by capturing the trade-off between these two metrics. Analysis of the confusion matrix reveals that false negatives are relatively limited compared to true positives, which is desirable in fraud detection scenarios. Moreover, the ROC-AUC score indicates strong class separability, demonstrating consistent performance across varying decision thresholds. Overall, the discussion confirms the robustness, reliability, and practical applicability of the Random Forest approach for credit card fraud detection.

## V ADVANTAGES

1. High Detection Accuracy: The application of machine learning algorithms facilitates high detection accuracy of fraudulent transactions by learning complex patterns from past data, thereby improving the performance of fraud detection.

2.Handling Imbalanced Data Effectively: The proposed system applies preprocessing and resampling methods to handle class imbalance effectively, thereby ensuring that fraudulent transactions are not missed despite their low prevalence.

3. Adaptive Learning Capability: Unlike rule-based systems, the machine learning algorithms applied in the proposed system adapt to new and evolving patterns of fraud, thereby making the system resilient to new fraud patterns.

4. Reduction in False Positives: By considering

5. Real-Time Fraud Detection: The application of trained machine learning algorithms to the transaction processing system enables real-time classification of transactions, thereby facilitating immediate action against fraudulent transactions.

6. Scalability and Flexibility: The system design facilitates scalability, thereby enabling the system to process large transaction data efficiently without compromising performance.

7. Cost Efficiency for Financial Institutions: Early detection of fraudulent transactions helps in preventing financial losses and costs associated with manual fraud analysis processes.

8. Privacy Preservation: The application of anonymized transaction features helps in preserving customer privacy while facilitating effective fraud detection.

9. Improved Decision Support: The system ensures effective decision support for banks and financial

institutions by helping in automated approval or rejection of transactions through predictive analysis.

10. Applicability to Real-World Systems: The proposed method can be easily incorporated with the current banking system and hence applicable to real-world systems.

## VI. FUTURE SCOPE

1. Integration of Deep Learning Models: Future developments could involve the integration of deep learning models like Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) to identify complex patterns in transaction data.

2. Real-Time Stream Processing: The system can be extended to process real-time transactions using tools like Apache Kafka or Spark Streaming to enable real-time fraud detection for high-frequency transactions.

3. Hybrid Fraud Detection Model: A hybrid model that combines supervised and unsupervised learning approaches can be developed to identify known and unknown fraud patterns more accurately.

4. Adaptive Model Retraining: Automatic model transaction data can be developed to enable continuous learning and adaptation to new fraud patterns.

5. Explainable AI Integration: Future developments could involve the integration of explainable machine learning approaches to enable transparency in fraud predictions, which can help financial institutions understand the rationale behind transaction decisions.

6. Cross-Channel Fraud Detection: The system can be extended to support the analysis of various transaction channels like online payments, mobile wallets, and point-of-sale transactions for complete fraud detection.

7. Cloud-Based Deployment of the System: The deployment of the fraud detection system on cloud platforms can help enhance scalability, availability, and processing speed, along with lowering infrastructure costs.

8. Integration of Behavioral Analysis: Behavioral analysis of users, including transaction rate, location, and spending behavior, can be integrated into the system to improve the accuracy of fraud detection.

9. Mobile and Web Application Integration: The system can be integrated with mobile and web applications to send real-time alerts and notifications to users and administrators.

10. Enhanced Compliance with Financial Regulations: Future developments of the system can be made to ensure complete compliance with financial regulations and data protection laws, ensuring safe and legal operations of fraud detection.

## VII. CONCLUSION

This paper has proposed a machine learning-based solution for credit card fraud detection using past transaction data. Different classification techniques have been tested, and the Random Forest method has been found to be most effective for identifying fraudulent transactions. The proposed system is more accurate, has fewer false positives, and can be easily incorporated into a real-time financial transaction system.

## REFFERENCE

[1] S. Dalal and S. K. Bansal, “Credit Card Fraud Detection Using Machine Learning Algorithms”, International Journal of Computer Applications, vol. 97, no. 2, pp. 6–9, July 2014.

[2] N. Malini and M. Pushpa, “Analysis on Credit Card Fraud Identification Techniques Based on KNN and Outlier Detection”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 6, no. 5, pp. 541–544, 2017.

[3] S. Carillo, A. Bontempi, and D. Dal Pezzulo, “Scarff: A Framework for Credit Card Fraud Detection Using Streaming Data,” Information Fusion, vol. 41, pp. 182–194, 2018.

[4] S. Bhattacharyya, S. Jha, K. Thara Kunnel, and J. C. Westland, “Data Mining for Credit Card Fraud: A Comparative Study”, Decision Support Systems, vol. 50, no. 3, pp. 602–613, Feb. 2011.

[5] A. Mishra and S. Tiwari, “Credit Card Fraud Detection Using Random Forest Algorithm”,

International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 6, pp. 1201–1205, 2020.

[6] R. Saxena and S. Sharma, "Credit Card Fraud Detection Using Neural Network", International Journal of Computer Science and Information Technologies (IJCSIT), vol. 7, no. 2, pp. 680–683, 2016.

[7] A. Dalal, P. Gupta, and R. Singh, "Credit Card Fraud Detection Using Machine Learning and Data Analytics", International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, pp. 119–124, 2018.

[8] P. S. Patil, V. A. Tambe, and S. R. Shinde, "Credit Card Fraud Detection Using Decision Tree and Support Vector Machine", International Journal of Advanced Research in Computer Engineering & Technology, vol. 4, no. 5, pp. 1818–1822, 2015.

[9] A. K. Jain, A. K. Gupta, and S. Tyagi, "A Novel Approach for Credit Card Fraud Detection Using Machine Learning", International Journal

[10] A. Sharma and A. Panigrahi, "A Review of Credit Card Fraud Detection Using Machine Learning", International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, no. 7, pp. 287–294, 2018.

[11] R. B. Patel and J. Mehta, "Credit Card Fraud Detection Using Hybrid Machine Learning Models", International Journal of Engineering and Advanced Technology (IJEAT), vol. 9, no. 3, pp. 2142–2147, 2020.

[12] S. R. Borkar and S. P. Patil, "Credit Card Fraud Detection Using Artificial Neural Networks", International Journal of Computer Trends and Technology (IJCTT), vol. 41, no. 2, pp. 94–98, 2016.

[13] P. K. Dhamija and S. K. Bhardwaj, "Credit Card Fraud Detection Using Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 9, pp. 2278–3075, 2019.

[14] Kaggle & ULB Machine Learning Group, "Credit Card Fraud Detection Dataset", European Cardholders Dataset, 2013.