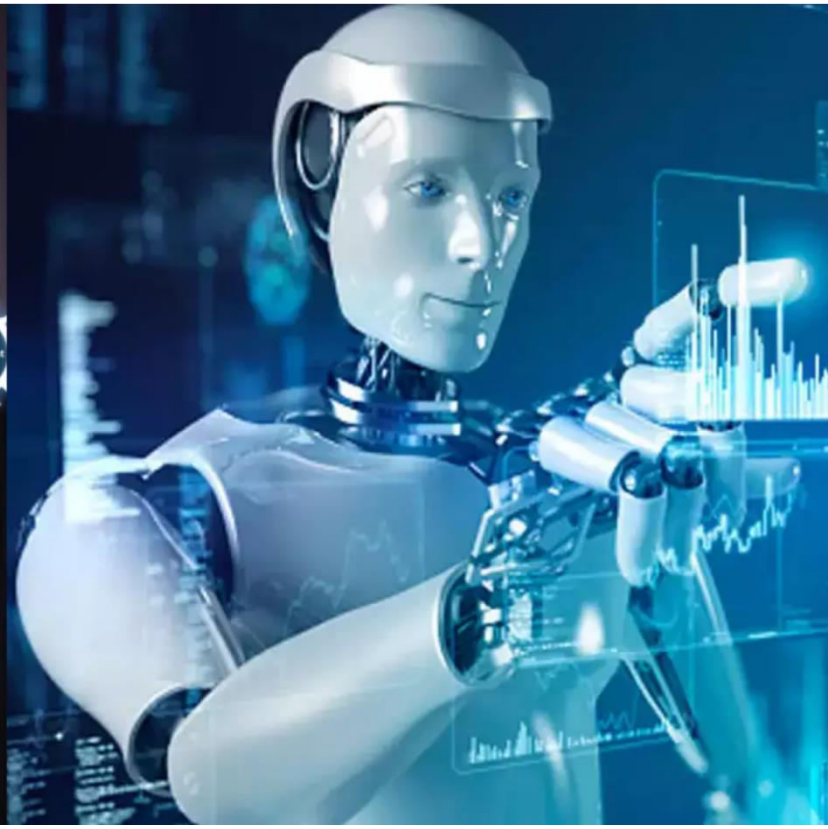
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404337|

CREDITSAFE AI: Detecting Fraudulent Transactions with Random Forest

T. Raja Mohan Reddy, A. Manasa, M. Rishitha, Y. Manaswini, D.Vinay Babu, K.Mahesh

Assistant Professor, Department of AI&DS, Sree Venkateswara College of Engineering, Rajupalem, Nellore, India

U.G. Student, Department of AI&DS, Sree Venkateswara College of Engineering, Rajupalem, Nellore, India

U.G. Student, Department of AI&DS, Sree Venkateswara College of Engineering, Rajupalem, Nellore, India

U.G. Student, Department of AI&DS, Sree Venkateswara College of Engineering, Rajupalem, Nellore, India

U.G. Student, Department of AI&DS, Sree Venkateswara College of Engineering, Rajupalem, Nellore, India

U.G. Student, Department of AI&DS, Sree Venkateswara College of Engineering, Rajupalem, Nellore, India

ABSTRACT: With the growing use of credit cards for both online and offline purchases, the need for effective fraud detection systems has become more important than ever. This paper presents an AI-driven approach using the Random Forest (RF) algorithm to detect fraudulent credit card transactions. The system analyzes transaction features to classify activities as either fraudulent or legitimate, addressing the challenge of imbalanced datasets where fraudulent cases are relatively rare. To overcome this imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied, and hyperparameter tuning is used to improve model performance. Developed using Python and key libraries such as Scikit-learn, Pandas, NumPy, and Streamlit, the system achieves an accuracy of 98% and an F1-score of 98%. This model can assist financial institutions in minimizing fraud risk, enhancing security, and improving transaction monitoring. Future enhancements may include the integration of deep learning techniques and real-time anomaly detection to further strengthen fraud prevention systems.

KEYWORDS: Credit Card Fraud Detection, Random Forest Algorithm, SMOTE, Imbalanced Dataset, Fraudulent Transactions, Machine Learning, Hyperparameter Tuning, Financial Security, Classification Models, Real-time Anomaly Detection**.**

I. INTRODUCTION

In today's digital world, credit cards have become one of the most common modes of payment. However, with the increase in online and cashless transactions, credit card fraud has also risen sharply. Fraud occurs when unauthorized individuals use stolen or fake credit card information to make purchases or withdraw money. These fraudulent transactions cause huge financial losses to banks, businesses, and customers.

Traditional fraud detection methods, which often rely on fixed rules or manual reviews, are no longer sufficient to detect modern fraud techniques. Fraudsters are constantly evolving their methods to bypass existing systems. To address this challenge, machine learning models offer a smarter and more adaptive solution.

This paper uses the Random Forest algorithm, a powerful ensemble learning technique, to build a fraud detection system. Random Forest combines multiple decision trees to improve the accuracy and stability of predictions. By analyzing key transaction details such as amount, time, location, and type, the model can identify patterns that indicate fraudulent behavior. The goal of this project is to create a reliable and accurate system that can detect fraud in credit card transactions and help reduce financial risk.

II. LITERATURE SURVEY

Credit card fraud detection has attracted significant research attention, particularly with the increase in online transactions. Many machine learning techniques have been explored, and Random Forest has consistently shown

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404337|

promising results due to its accuracy, robustness, and ability to handle imbalanced datasets where fraudulent transactions are rare.

Dal Pozzolo et al. (2015) [1] highlighted the challenge of class imbalance in fraud detection datasets. They applied under sampling techniques in conjunction with Random Forest and demonstrated that this combination significantly enhanced the detection of rare fraud cases without increasing false positives. Their work emphasized the importance of balancing datasets to improve model effectiveness in fraud detection.

Bahnsen et al. (2016) [2] focused on cost-sensitive learning for fraud detection. Rather than solely improving classification accuracy, their approach aimed at minimizing financial losses due to misclassified transactions. By leveraging Random Forest, they created a model that prioritized high-risk fraudulent transactions, ultimately helping financial institutions mitigate potential losses.

Carcillo et al. (2017) [3] conducted an analysis on streaming credit card data and tested several machine learning algorithms. They found that Random Forest outperformed other models, particularly when integrated with techniques like SMOTE and under sampling to manage class imbalance. Their study also explored the role of active learning in improving fraud detection for real-time systems.

Ayesha et al. (2018) [4] compared multiple machine learning models, including Random Forest, using publicly available banking datasets. They concluded that Random Forest achieved the highest accuracy, precision, and recall, demonstrating its suitability for handling noisy, high-dimensional fraud detection data, which is typical of credit card transaction datasets.

Jha et al. (2019) [5] introduced a hybrid model that combined Random Forest with clustering algorithms. Their approach focused on analyzing customer behavior patterns to detect hidden fraud cases. The hybrid model outperformed standalone models, showcasing the benefits of integrating different machine learning techniques for enhanced fraud detection.

Based on these studies, it is clear that Random Forest is one of the most reliable and widely used algorithms for credit card fraud detection. It offers high accuracy, effectively handles imbalanced datasets, and can be further improved by incorporating other techniques for more robust performance

III. METHODOLOGY

This study presents the design and implementation of a credit card fraud detection system using the Random Forest algorithm. The process consists of the following key stages:

A. Data Collection and Preprocessing

The dataset contains detailed information on each transaction, including features such as transaction amount, time, merchant details, and customer location. The initial step involves cleaning the data by removing missing or redundant values. Subsequently, feature scaling and normalization techniques are applied to ensure all attributes are on a uniform scale, thereby enhancing the learning efficiency of the model.

B. Handling Imbalanced Data

Credit card fraud detection inherently involves an imbalanced dataset, as fraudulent transactions are significantly less frequent than legitimate ones. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) is employed. SMOTE generates synthetic instances of the minority class (fraudulent transactions), which helps to balance the dataset and improve the model's ability to detect fraud effectively.

C. Model Training

The Random Forest algorithm is chosen due to its robustness and ability to handle complex, high-dimensional datasets. The model is trained on the preprocessed and balanced dataset. Hyperparameters such as the number of trees and the maximum depth of trees are optimized using grid search combined with cross-validation to enhance overall performance.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404337|

D. Model Evaluation

To assess the model's effectiveness, performance metrics such as accuracy, precision, recall, and F1-score are computed. These metrics provide a comprehensive understanding of the model's ability to correctly identify fraudulent transactions while minimizing false positives and false negatives.

E. Deployment

The fraud detection system is developed using Python and libraries including Scikit-learn, Pandas, NumPy, and Streamlit. Post-training, the model is deployed as a web-based application, enabling real-time fraud detection for financial institutions.

IV. IMPLEMENTATION

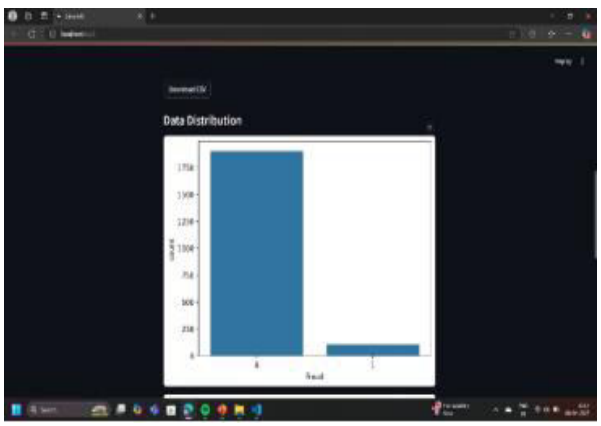
The provided code implements a real-time credit card fraud detection system using a trained Random Forest model. The system utilizes Streamlit to create an easy-to-use web interface for fraud prediction. Initially, the trained Random Forest model (fraud_model.pkl) and the scaler (scaler.pkl) used to normalize the transaction amount are loaded. The application presents a form where users can input transaction details, including anonymized features (V1 to V28) and the transaction amount. Once the user submits the data, the transaction amount is scaled using the previously saved scaler, and the model predicts whether the transaction is fraudulent or legitimate. The result is then displayed on the interface, alongside the model's confidence score, with a clear label indicating the status of the transaction (fraudulent or legitimate).

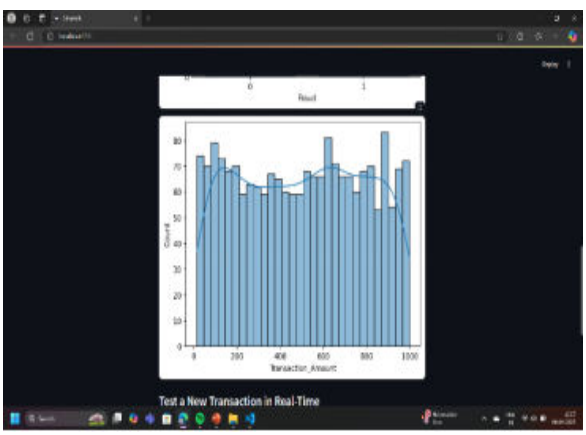
The code also contains the training process for the Random Forest model. It begins by loading and preparing the credit card transaction dataset, separating the features and class labels. The transaction amount is standardized using the Standard Scaler, and the data is split into training and testing sets. To address the class imbalance between fraudulent and legitimate transactions, the model is trained with the class weight='balanced' parameter. The Random Forest model is constructed using scikit-learn's Random Forest Classifier with 100 estimators. After training, both the model and scaler are saved as fraud_model.pkl and scaler.pkl. These saved files are later loaded in the Streamlit app to make real-time fraud predictions. This setup enables the system to efficiently detect fraud in credit card transactions based on the provided input data.

V. CONCLUSION

In this paper, we developed a fraud detection system using the Random Forest algorithm to detect fraudulent credit card transactions. By cleaning the data, addressing the imbalance in the dataset with SMOTE, and fine-tuning the model settings, we created a system that performs with high accuracy. This system offers a practical solution for real-time fraud detection, helping banks and financial institutions reduce risks and prevent fraud effectively.

In the future, we plan to enhance the system by integrating deep learning models and real-time anomaly detection to improve its ability to spot fraud and adapt to new fraud patterns.





International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404337|

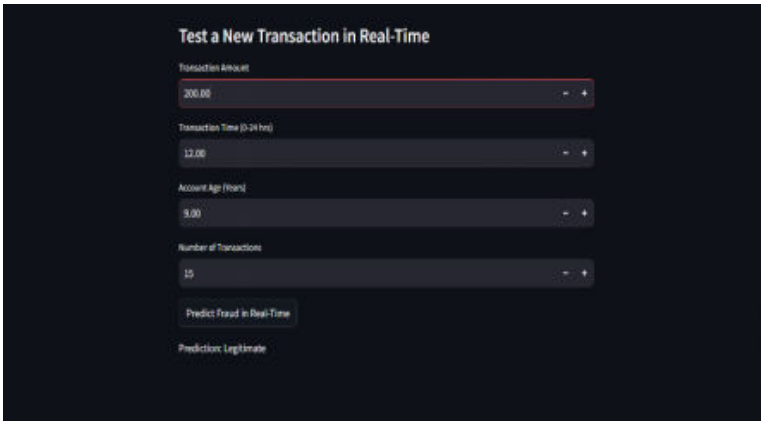


FIGURE: Results of our credit card fraud Detection

REFERENCES

1. Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences, 479, 448–455.
2. Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. Decision Support Systems, 95, 91–101.
3. Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, 51, 134–142.
4. Zareapoor, M., & Shamsolmoali, P. (2015). Application of credit card fraud detection: based on bagging and ensemble classifier. Procedia Computer Science, 48, 679–685.
5. Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., & Nandi, A. K. (2017). Credit card fraud detection using AdaBoost and majority voting. IEEE Access, 6, 14277–14284.
6. Save, P., Tiwarekar, P., Jain, K. N., & Mahyavanshi, N. (2017). A novel idea for credit card fraud detection using decision tree. International Journal of Computer Applications, 161(13), 0975–8887.
7. Sorournejad, S., Zojaji, Z., Atani, R. E., & Monadjemi, A. H. (2016). A survey of credit card fraud detection techniques: data and technique oriented perspective. ArXiv preprint arXiv:1611.06439.
8. Singh, A., & Jain, A. (2019). Adaptive credit card fraud detection techniques based on feature selection method. In Advances in Computer Communication and Computational Sciences (pp. 167–178). Springer.
9. Li, Z., Liu, G., Wang, S., Xuan, S., & Jiang, C. (2018). Credit card fraud detection via kernel-based supervised hashing. In 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (pp. 1272–1279). IEEE.
10. Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. (2018). Random forest for credit card fraud detection. In 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC) (pp. 1–6). IEEE.
11. Cruz, R., & Hammad, M. (2019). Credit card fraud detection using deep learning models. International Journal of Advanced Computer Science and Applications (IJACSA), 10(6), 219-225.
12. Gu, F., & Jiao, Z. (2021). Credit card fraud detection using deep learning and feature engineering. Journal of Intelligent & Fuzzy Systems, 40(3), 5119-5129.
13. Kaur, S., & Kaur, H. (2020). A review on machine learning techniques for credit card fraud detection. Journal of King Saud University - Computer and Information Sciences, 32(3), 254-264.
14. Ravisankar, P., & Nandi, A. K. (2020). Credit card fraud detection using machine learning techniques: A survey. Computers, Materials & Continua, 64(3), 1355-1377.
15. Alsewari, A. M., & Elhadi, A. M. (2021). Credit card fraud detection using hybrid machine learning algorithms. Journal of Computational and Theoretical Nanoscience, 18(7), 3476-3484.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details