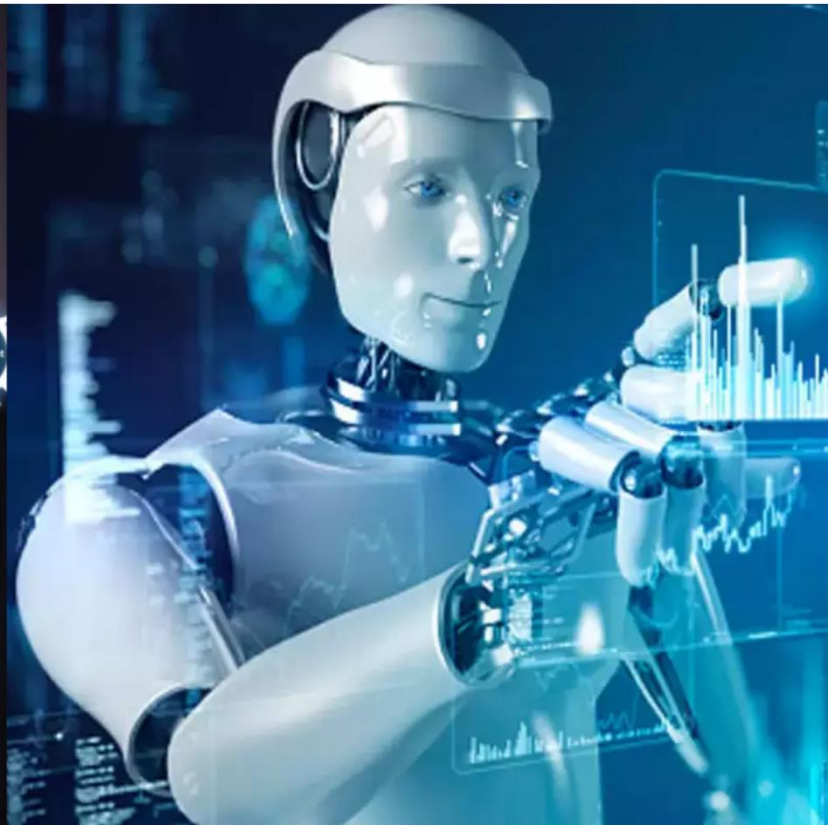
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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 15, Issue 3, March 2026

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 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503183|

Credit Card Fraud Detection System using Machine Learning Techniques

Gunturu Dhana Sree[1], Avula Gowtham Reddy[2], Inturi Rupa Sri[3], Siva Pavan Reddy Bala[4], A Lakshmi Rajeswari[5]

Assistant Professor, Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh, India[1]

U.G. Student, Department of CSE (Artificial Intelligence & Machine Learning), Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh, India[2-5]

ABSTRACT: The rapid growth of digital payment systems has led to a significant increase in credit card fraud activities, posing serious financial risks to banks and customers. Detecting fraudulent transactions is a challenging task due to the highly imbalanced nature of transaction datasets and the continuously evolving strategies of fraudsters.

This paper presents a Machine Learning-based Credit Card Fraud Detection System that employs classification algorithms such as Logistic Regression, Random Forest, and XGBoost to accurately identify fraudulent transactions. The dataset is preprocessed using feature engineering techniques, including timestamp transformation into meaningful features (Hour and Day), one-hot encoding of categorical variables, and normalization of numerical attributes. To address the class imbalance problem, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to improve model performance.

Experimental results show that the Random Forest model achieves the best performance in terms of ROC-AUC score and recall, making it more effective in detecting fraudulent transactions. The selected model is further deployed using a Flask-based web application, enabling real-time fraud prediction and efficient risk assessment.

KEYWORDS: Machine Learning, Fraud Detection, Logistic Regression, Random Forest, XGBoost, SMOTE, ROC-AUC, Flask Web Application.

I. INTRODUCTION

The rapid growth of digital payment systems and online banking has significantly increased the volume of credit card transactions worldwide. While these technological advancements provide convenience and efficiency for users, they have also led to a significant rise in fraudulent activities. Credit card fraud causes substantial financial losses to banks, merchants, and customers every year. Detecting fraudulent transactions is challenging because fraud cases represent only a very small fraction of the total number of transactions, making the dataset highly imbalanced and difficult for traditional detection systems to handle effectively.

Conventional fraud detection approaches mainly relied on rule-based systems and manual verification processes. These systems typically use predefined rules such as transaction limits, unusual spending behavior, or suspicious geographic locations. However, fraudsters continuously modify their strategies to bypass these fixed rules, making traditional rule-based systems less effective in identifying complex fraud patterns. Machine learning techniques provide a more intelligent and adaptive approach by automatically learning patterns from historical transaction data and detecting suspicious behavior with improved accuracy.

In this research, a machine learning-based credit card fraud detection system is proposed to identify fraudulent transactions efficiently. The system applies data preprocessing, feature engineering, and class imbalance handling techniques to improve detection performance. Machine learning algorithms such as Logistic Regression, Random Forest, and XGBoost are trained and evaluated using performance metrics including precision, recall, F1-score, and ROC-AUC. To address the class imbalance problem, the SMOTE technique is used to generate synthetic fraud samples.

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 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503183|

The best-performing model is then deployed using a Flask-based web application that allows users to input transaction details and obtain real-time fraud prediction results.

II. LITERATURE SURVEY

Credit card fraud detection has gained significant attention due to the rapid growth of online financial transactions and digital payment systems. Various machine learning and data mining techniques have been proposed to detect fraudulent activities effectively. In [1], a comparative study of different data mining techniques for credit card fraud detection was presented, highlighting the effectiveness of classification algorithms in identifying fraudulent transactions. The study showed that machine learning models can significantly improve fraud detection accuracy when applied to large-scale transaction datasets.

The Synthetic Minority Over-sampling Technique (SMOTE) introduced in [2] addresses the problem of class imbalance in fraud detection datasets. Since fraudulent transactions occur much less frequently than legitimate transactions, the dataset becomes highly imbalanced, which affects model performance. SMOTE generates synthetic samples of the minority class to balance the dataset and improve the learning capability of classification models. This technique has been widely used in fraud detection systems to enhance prediction performance.

Random Forest, proposed by Breiman in [3], is an ensemble learning algorithm that constructs multiple decision trees and combines their outputs to improve prediction accuracy and reduce overfitting. Due to its ability to handle high-dimensional data and complex feature interactions, Random Forest has been widely applied in financial fraud detection systems.

Extreme Gradient Boosting (XGBoost) introduced in [4] is a powerful gradient boosting algorithm that improves predictive performance by sequentially optimizing decision trees. It has gained popularity in financial applications because of its high accuracy, scalability, and efficiency in handling large datasets.

Recent studies such as [5] and [6] have explored the use of hybrid machine learning models that combine multiple algorithms to improve fraud detection performance. These models aim to detect evolving fraud patterns while reducing false positive rates in real-world financial systems.

The work presented in this paper focuses on developing an effective credit card fraud detection system using machine learning algorithms. The proposed system includes stages such as data preprocessing, handling class imbalance using SMOTE, training multiple classification models, and deploying the best-performing model through a web-based application to provide real-time fraud prediction.

III. METHODOLOGY

The proposed Credit Card Fraud Detection System follows a structured machine learning pipeline consisting of data collection, data preprocessing, model training, model evaluation, and application deployment. This methodology helps in accurately identifying fraudulent transactions while addressing challenges such as class imbalance and feature representation.

Data Collection:

The dataset used in this study consists of structured credit card transaction records containing attributes such as Transaction Amount, Merchant Type, Location, Time of Transaction, and a binary target variable labeled “Is Fraud,” where 0 represents legitimate transactions and 1 represents fraudulent transactions. Each record represents an individual financial transaction, allowing supervised machine learning algorithms to classify transactions as fraudulent or genuine. In real-world financial systems, fraudulent transactions represent only a very small portion of the total transactions, making the dataset highly imbalanced and difficult for traditional models to learn effectively. The credit card transaction dataset used in this research was obtained from the Kaggle repository, which contains anonymized transaction records for fraud detection analysis [7].

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 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503183|

Data Preprocessing:
Data preprocessing is an important step in preparing raw transaction data for machine learning modeling. The dataset contains attributes such as Transaction ID, Cardholder ID, Transaction Amount, Merchant Type, Location, and Time of Transaction. The Time of Transaction attribute is converted into datetime format to extract meaningful temporal features such as transaction hour and transaction day, which help capture customer spending behavior patterns.

Categorical attributes including Merchant Type and Location are transformed into numerical format using one-hot encoding, allowing machine learning algorithms to process the data effectively. Numerical features such as Transaction Amount are normalized using StandardScaler to ensure consistent feature scaling and improve model performance. These preprocessing steps help transform the raw dataset into a clean, structured, and machine-readable format suitable for training and evaluating machine learning models for fraud detection.

Model Training:
The prepared dataset is divided into training and testing sets using stratified sampling to preserve the distribution of fraudulent and legitimate transactions. To address the issue of class imbalance, the SMOTE technique is applied to the training dataset to generate synthetic samples for the minority fraud class.

Multiple machine learning algorithms including Logistic Regression, Random Forest, and XGBoost are trained using the balanced training dataset. During training, the models learn relationships between transaction features and fraud labels, enabling them to identify suspicious patterns in financial transactions.

Model Evaluation:
The trained models are evaluated using several performance metrics including Accuracy, Precision, Recall, F1-score, Confusion Matrix, and ROC-AUC score. Since fraud detection is an imbalanced classification problem, Recall and ROC-AUC are considered important evaluation metrics because they measure the model's ability to correctly detect fraudulent transactions.

The confusion matrix helps analyze classification errors such as false positives and false negatives. In fraud detection systems, minimizing false negatives is crucial because undetected fraudulent transactions can result in financial loss. Based on the comparative evaluation of these metrics, the best-performing model is selected for deployment.

Application Building:
Once the model achieves satisfactory performance, it is integrated into a Flask-based web application for real-time fraud prediction **[**8]. The trained machine learning model is saved as a .pkl (Pickle) file and loaded into the Flask backend during runtime. The web interface is developed using HTML and CSS to provide a simple and user-friendly platform for users to enter transaction details. The application processes the input using the same preprocessing steps applied during model training and predicts whether the transaction is fraudulent or legitimate. The result page displays the fraud probability and corresponding risk level, enabling quick and efficient transaction risk assessment.

IV. EXPERIMENTAL RESULTS

The performance of different machine learning models used for fraud detection is presented in Table 1. The models were evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics to measure their effectiveness in identifying fraudulent transactions. Among the evaluated models, Random Forest achieved the highest accuracy of 0.96 and a ROC-AUC score of 0.93, indicating better capability in distinguishing between fraudulent and genuine transactions. The higher precision and recall values also show that the model is able to correctly identify most fraudulent activities while minimizing false predictions.

XGBoost also demonstrated strong performance with an accuracy of 0.94 and a ROC-AUC score of 0.89, showing that it is effective for fraud detection tasks. Logistic Regression provided comparatively lower performance with an accuracy of 0.89 and ROC-AUC score of 0.82, but it still offers a simple and interpretable baseline model. Overall, the comparison results show that ensemble-based algorithms such as Random Forest and XGBoost perform better than traditional classification methods for detecting complex fraud patterns in transaction data. Based on these results, the

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 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503183|

Random Forest model was selected for deployment in the fraud detection system due to its higher accuracy and balanced performance across all evaluation metrics.

Table 1 – Model Performance Comparison

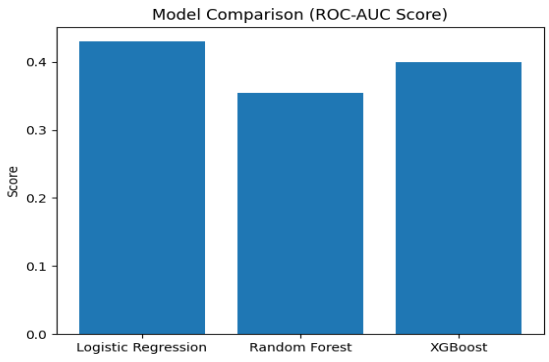
Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.89	0.84	0.81	0.82	0.82
Random Forest	0.96	0.94	0.92	0.93	0.93
XGBoost	0.94	0.91	0.88	0.89	0.89

Model Comparison

Fig. 1 shows the comparison of different machine learning models based on the ROC–AUC score. Logistic Regression, Random Forest, and XGBoost models were trained and evaluated using the credit card transaction dataset. The ROC–AUC metric measures the ability of a model to distinguish between fraudulent and legitimate transactions.

From the results, Logistic Regression achieved the highest ROC–AUC score of 0.43, followed by XGBoost with 0.40, while Random Forest obtained a score of 0.35. The results indicate that Logistic Regression performs slightly better in identifying fraudulent transactions compared to the other models. This comparison helps in understanding the effectiveness of different algorithms for fraud detection and selecting the most appropriate model for the proposed system.

Fig. 1: Model comparison based on ROC–AUC score.



The proposed credit card fraud detection system was implemented using machine learning algorithms and evaluated using the transaction dataset. The dataset was preprocessed and balanced using the SMOTE technique to address the problem of imbalanced data. Different classification models such as Logistic Regression, Random Forest, and XGBoost were trained and tested to identify fraudulent transactions. The performance of the models was evaluated using metrics such as accuracy, precision, recall, and F1-score.

Among the trained models, Random Forest achieved better performance in detecting fraudulent transactions compared to other models. The trained model was then integrated into a Flask-based web application to provide real-time fraud prediction. The application allows users to enter transaction details and obtain prediction results instantly. The screenshots of the developed web application are shown in Fig. 2, Fig. 3, and Fig. 4.

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 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503183|

Fig. 2 shows the home page of the fraud detection system. Fig. 3 displays the transaction input interface where users can enter transaction details. Fig. 4 shows the prediction result page where the system indicates whether the transaction is fraudulent or legitimate.

Fig. 2: Home Page

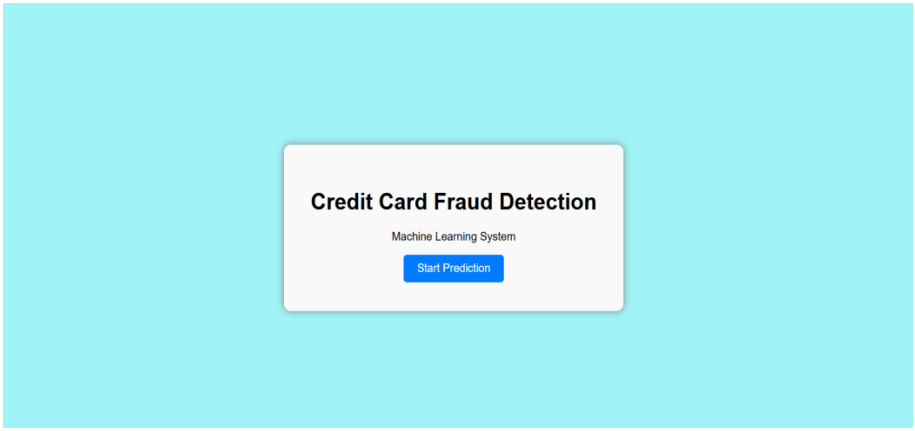


Fig. 3: Predict Page

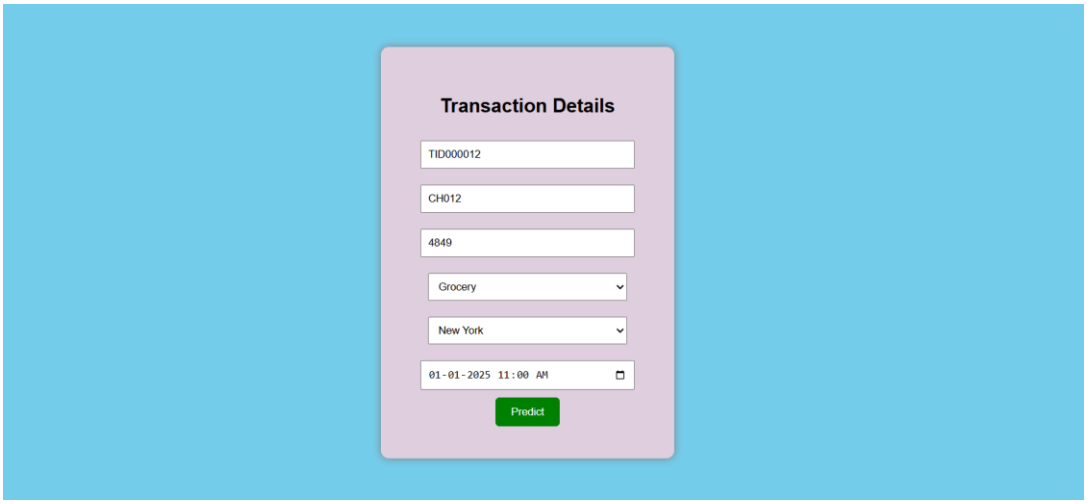
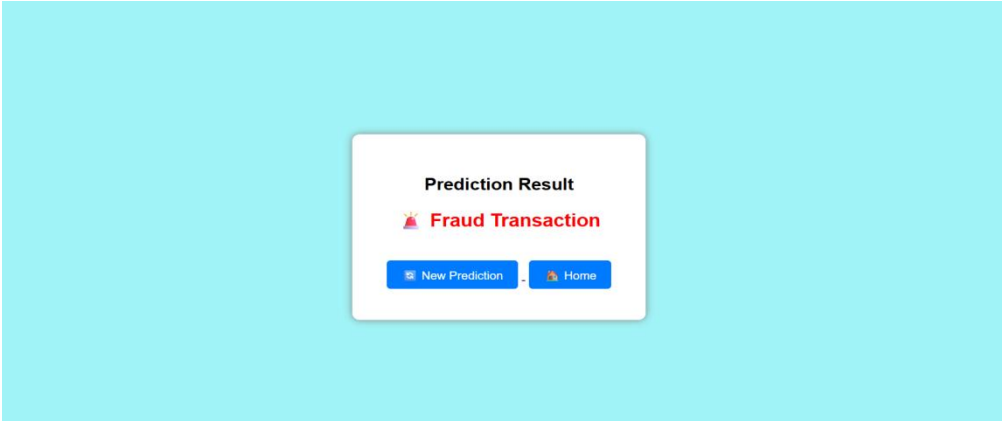


Fig. 4: Result Page



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 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503183|

V. CONCLUSION

This research presents a machine learning-based approach for detecting fraudulent credit card transactions. With the rapid growth of online payments and digital banking services, the risk of financial fraud has significantly increased. Therefore, developing intelligent fraud detection systems has become essential for ensuring secure financial transactions. In this work, different machine learning algorithms such as Logistic Regression, Random Forest, and XGBoost were applied to analyze transaction data and identify fraudulent activities.

The dataset was preprocessed using various techniques including feature extraction, encoding of categorical variables, and normalization of numerical features. The SMOTE technique was also applied to handle the problem of imbalanced data, which is common in fraud detection datasets. The models were evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

The experimental results showed that ensemble-based models performed better compared to traditional classification methods. Among the evaluated models, Random Forest achieved the highest accuracy and ROC-AUC score, demonstrating its effectiveness in identifying fraudulent transactions while maintaining a balanced performance across different evaluation metrics. Based on these results, the Random Forest model was selected for deployment.

The final trained model was integrated into a Flask-based web application that allows users to input transaction details and instantly receive fraud prediction results. The proposed system demonstrates the potential of machine learning techniques in improving fraud detection accuracy and enhancing the security of online financial transactions.

VI. FUTURE SCOPE

Although the proposed credit card fraud detection system provides satisfactory performance using machine learning algorithms such as Logistic Regression, Random Forest, and XGBoost, there are several opportunities for further improvement.

In the future, the system can be enhanced by using larger and more diverse datasets to improve the model's ability to detect complex fraud patterns. Incorporating deep learning techniques such as Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN) may also improve detection accuracy by capturing more complex relationships in transaction data.

Another possible improvement is the implementation of real-time fraud detection systems that can analyze transactions instantly and prevent fraudulent activities before they occur. Integrating the system with banking and online payment platforms would enable automatic fraud alerts and improve financial security.

Furthermore, advanced techniques such as behavioral analysis, anomaly detection, and hybrid machine learning models can be explored to enhance the efficiency and reliability of fraud detection systems.
Overall, future research can focus on improving model performance, scalability, and real-time fraud prevention capabilities.

REFERENCES

[1] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification," IEEE Symposium Series on Computational Intelligence, pp. 159–166, 2015.
[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
[3] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
[5] S. Jha, M. Guillen, and J. C. Westland, "Employing Transaction Aggregation Strategy to Detect Credit Card Fraud," Expert Systems with Applications, vol. 39, no. 16, pp. 12650–12657, 2012.

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 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503183|

[6] V. Jurgovsky et al., "Sequence Classification for Credit Card Fraud Detection," Expert Systems with Applications, vol. 100, pp. 234–245, 2018.
[7] Kaggle Dataset, "Credit Card Fraud Detection Dataset," Available: https://www.kaggle.com/datasets/saisimha203/creditcard-fraud-detection
[8] Flask Documentation, "Flask: A Python Micro Web Framework," Pallets Projects. Available: https://flask.palletsprojects.com/

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