International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 9, Issue 12, pp 72-77, December-2025
https://doi.org/10.47001/IRJIET/2025.912010

Identifying Fraudulent Credit Card Transactions Using Ensemble Machine Learning

[1]Riza Peerzade, [2]Riya Jadhav, [3]Sanika Pathare, [4]Manjushri Jadhav, [5]Prof. Nita Pawar, [6]Prof. Nita Pawar

[1,2,3,4]Student, Computer Engineering Diploma, Ajeenkya D. Y. Patil School of Engineering, Charholi, Pune, India
[5]Guide, Professor, Computer Engineering Diploma, Ajeenkya D. Y. Patil School of Engineering, Charholi, Pune, India
[6]HoD, Professor, Computer Engineering Diploma, Ajeenkya D. Y. Patil School of Engineering, Charholi, Pune, India

***Abstract* - Credit card fraud is a big problem for banks and their consumers, and it costs a lot of money around the world. The data is quite unbalanced, making it difficult to identify fake transactions, which make up a very small fraction of all transactions. This study examines the utilization of machine learning methodologies for fraud detection, employing a publicly accessible dataset of credit card transactions conducted by European cardholders in September 2013. The dataset contains 284,807 transactions from two days, yet just 492 of these (0.172%) were found to be fraudulent. Principal Component Analysis (PCA) has changed all of the input variables except for the transaction amount and time to keep them private. In this work, we explore the difficulties of finding credit card fraud and try to look at the newest improvements in fraud detection methods, datasets, and evaluation standards. We list and assess the pros and cons of different ways to identify fraud. This study introduces a viable and reproducible methodology for detecting credit card fraud through supervised machine learning applied to a commonly utilized credit card dataset characterized by transactions over two days with significant imbalance. We talk about preprocessing, how to address very uneven class distributions, how to choose models, how to use evaluation metrics that work for rare-event detection, how to set up experiments, and what analysis we propose. We compare standard classifiers, including logistic regression, random forest, NB, and neural networks. We also talk about deployment issues and suggest future work.**

Keywords: Principal Component Analysis, Machine learning, Credit card fraud detection, ensemble learning.

I. INTRODUCTION

Digital payment systems have grown quickly, and they have improved the global financial system by making it easier and faster for millions of consumers to use. This move has also made it easier for people to conduct fraud, especially when they use credit cards. Credit card fraud costs banks and customers money, and it also makes people less likely to trust digital financial services. According to recent statistics from the industry, fraud costs billions of dollars every year. This phenomenon is why both businesses and academia are so interested in researching how to find fraud. Finding fraud is not always simple. First, false transactions are much less common than actual ones. This means that the datasets are quite unbalanced because the fake class only makes up less than 1% of all transactions. Because of this imbalance, common ways of classifying things don't work because models tend to favor the majority class. Second, fraud patterns evolve; thus, models need to be able to adapt to new behaviors. Third, false positives can make consumers angry for no reason, and false negatives can cost a lot of money. So, it's very crucial to build fraud detection models that are robust, accurate, and can grow. Machine learning presents us great tools for dealing with these difficulties by detecting hidden patterns in transaction data. This study utilizes a publicly available dataset of European credit card transactions, integrating anonymized data obtained from Principal Component Analysis (PCA), alongside transaction amount and timing. We focus on creating predictive models that not only have high accuracy but also improve precision and recall. This makes it easier to identify rare cases of fraud while cutting down on false positives. Using resampling approaches and ensemble models, we intend to develop a complete framework that can be used as a starting point for real-world systems that look for fraud.

II. LITERATURE SURVEY

Abdul Salam, Mustafa, *et al.* (2024) [1] a federated learning–based fraud detection framework that enables multiple financial organizations to collaboratively train models without exposing their raw datasets. The authors employ data balancing methods such as SMOTE and ADASYN to mitigate severe class imbalance. Their framework achieves notable improvements in detecting fraudulent transactions while maintaining strong privacy guarantees. Experimental testing reports high recall and stable performance across distributed environments. The study underscores the promise of privacy-conscious machine learning for practical fraud detection deployments.

ISSN (online): 2581-3048
Volume 9, Issue 12, pp 72-77, December-2025
https://doi.org/10.47001/IRJIET/2025.912010

Mosa, Diana T., *et al.* (2024) [2] the present CCFD, a hybrid fraud detection approach that merges meta-heuristic optimization with machine learning classification. Algorithms like GA and PSO are utilized to identify the most informative feature subsets, thereby enhancing predictive accuracy. Their findings reveal that these optimized models consistently outperform conventional ML classifiers when dealing with rare fraud instances. The system demonstrates improved precision and F1-scores across diverse credit card datasets. This study highlights the importance of optimization-driven feature selection in fraud analytics.

Chung, Jiwon & Kyungho Lee (2023) [3] a detection strategy centered on maximizing recall, a key metric for minimizing missed fraud cases. The method combines LDA, KNN, and linear regression to strengthen separation between classes in highly imbalanced data. Results show that incorporating LDA before KNN significantly boosts detection capabilities. The authors report consistently higher recall rates compared to standard machine learning models. Their work stresses that reducing false negatives is crucial for operational banking systems.

Nuthalapati, Aravind (2023) [4] an intelligent credit card fraud detection model employing a mix of machine learning algorithms to enhance security. The approach includes comprehensive preprocessing, feature ranking, and classifier adjustments to improve predictive accuracy. Experimental evaluations indicate competitive performance across multiple benchmark datasets. The findings suggest that ML-driven solutions can effectively support and improve traditional rule-based fraud detection systems. The study advocates scalable ML integration to strengthen financial protection mechanisms.

Afriyie, Jonathan Kwaku, *et al.* (2023) [5] the design a supervised machine learning framework capable of both detecting and predicting fraudulent credit card activity. Their methodology incorporates thorough data preprocessing, feature evaluation, and multi-metric performance assessment. Tree-based models emerge as the top-performing classifiers, with superior accuracy and recall. The study also emphasizes the necessity of real-time detection for financial institutions. Their findings reinforce the value of ML-based systems for proactive fraud prevention.

Alfaiz, Noor Saleh & Suliman Mohamed Fati (2022) [6] an enhanced machine learning model focused on improving classification outcomes in imbalanced fraud detection datasets. The authors integrate sampling approaches with optimized ML algorithms to boost performance. Their results show substantial gains in precision, recall, and F1-score. The work demonstrates that selecting an effective combination of balancing techniques and classifiers can greatly reduce financial loss due to fraud. The improved framework is well-suited for large-scale transactional data environments.

Moradi, Tarif & Homaei (2025) – Systematic Review [7] the comprehensive review explores the evolution of machine learning techniques applied to credit card fraud detection, ranging from conventional algorithms to advanced deep learning solutions. The authors discuss key challenges, including severe class imbalance, anonymized feature spaces, and real-time processing needs. They highlight new trends such as federated training, graph-based anomaly detection, and ensemble architectures. The review also notes limitations like insufficient benchmark datasets and scalability issues. It offers a detailed roadmap to guide future research in the field.

Theodorakopoulos, Leonidas, *et al.* (2025) [8] a distributed, big data–oriented fraud detection platform using PySpark in combination with XGBoost and CatBoost. Designed to manage millions of records efficiently, the system leverages cluster computing for accelerated processing. Experiments indicate that gradient boosting–based models achieve high accuracy and rapid training times in large-scale settings. The authors demonstrate that computational parallelism significantly enhances detection efficiency. The framework is optimized for real-time deployment in banking and fintech environments.

Al-Maari, Al-Anood, *et al.* (2025) [9] a fine-tuned ensemble learning model that merges multiple machine learning techniques for more dependable fraud detection. Hyperparameter optimization is applied to maximize the ensemble's predictive strength. Their experiments show notable improvements in AUC, F1-score, and recall when compared to individual models. The study highlights the resilience of ensemble methods against noisy and imbalanced data. Overall, the work validates hybrid ML approaches as powerful tools for safeguarding credit card systems.

Khalid, Abdul Rehman, *et al.* (2024) [10] an ensemble-based strategy to enhance the accuracy and robustness of fraud detection systems. The authors combine several classifiers through techniques such as stacking and majority voting. Their evaluation shows superior performance across key metrics, particularly in precision and F1-score. They also discuss how model diversity helps capture intricate fraud patterns more effectively. The enhanced ensemble proves well-suited for deployment in real-world financial environments.

III. METHODOLOGY

Data Collection

We gather data from various sources, including the UCI ML repository, Kaggle, and several real-time data sources.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 12, pp 72-77, December-2025

https://doi.org/10.47001/IRJIET/2025.912010

Prior to executing the classification activity, the data must be pre-processed in order to enhance the outcome by addressing the missing values and removing the redundant features contained in the chosen dataset. The target variable Class indicates whether a transaction is fraud (1) or genuine (0).

Preprocessing and normalization

There may be a lot of useless information and gaps in the data. Data preparation is carried out to handle this portion. Numerous data preprocessing techniques, such as data cleansing, data transformation, and data reduction, have been utilized at this step.

- Data cleaning: It addresses noisy data, missing information, etc. Different strategies have been adopted when some data in the information is incomplete, such as filling in the gaps or disregarding the tuples. Data may contain null values that are incomprehensible to machines. This noisy data may result from poor data collecting, incorrect data input, etc. Regression, clustering, and the binning approach are used to address it.
- Data transformation: This technique is used to change the data into the form that is suited for the mining procedure. Normalization, attribute choice, and discretization are involved in this technique.

Feature Extraction and Selection

From the data input, this procedure retrieves a variety of features. The extracted features are then standardized using a feature selection threshold, which eliminates redundant and unnecessary features for training. The normalized data with relational characteristics is used to extract a variety of hybrid attributes, and training is carried out by selecting an optimization strategy. The hybrid method has been used for feature selection from fully extracted features—selecting the best quality increases classification accuracy. Many irrelevant features appear during the feature extraction, which need to be eliminated when we choose the parts. The benefit of this method is that it provides a respective feature selection for the individual feature set.

Classification

After the module has been successfully executed, the selected features are given as input to the training module, which produces comprehensive background knowledge for the overall system. After we get the training model, we can feed the testing data into it and get a classification prediction. The testing stage includes preprocessing of testing text, vectorization, and classification of the testing text. The module testing evaluates the system's predictive performance using hybrid machine learning methods. Various machine learning models are applied. Class balance is handled using cost-sensitive learning and threshold adjustment. Performance is evaluated with precision, recall, and F1-score to ensure reliable fraud detection in real-world banking scenarios.

IV. SYSTEM IMPLEMENTATION

The proposed credit card fraud detection framework is structured with a modular architecture that enables efficient processing of high-volume transactional data while supporting accurate, real-time decision-making. The system is composed of several coordinated components, including data acquisition, preprocessing, feature engineering, model training and classification, and alert management. Initially, transaction records are gathered from various financial sources and go through preprocessing to handle missing entries, eliminate noise, and normalize the data. Subsequently, key attributes are extracted and refined through hybrid feature selection techniques to enhance predictive performance. The optimized data is then provided to machine learning models—such as logistic regression, random forest, naïve Bayes, and neural networks—to classify and detect potentially fraudulent activities. Suspicious transactions are flagged, and automatic notifications or comprehensive analytical reports are delivered to security teams and system administrators. The proposed architecture prioritizes scalability, robustness, and maintainability, making it suitable for deployment in real-world banking environments while ensuring strong data security and operational efficiency.

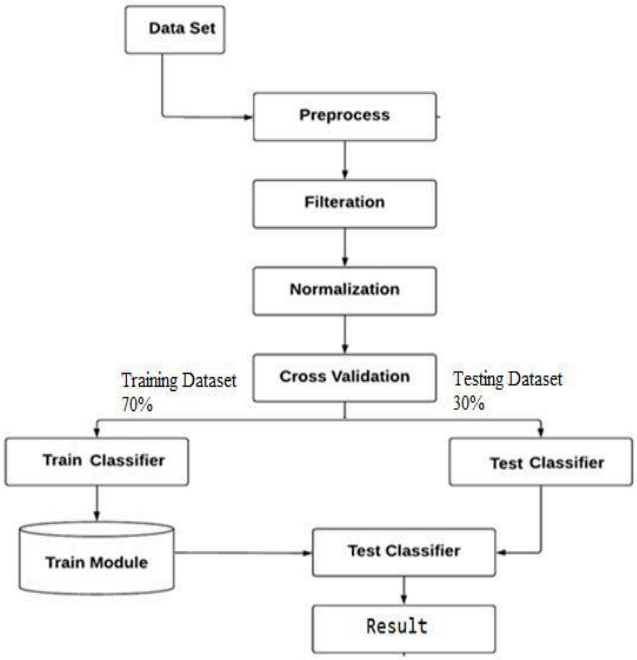


Figure 1: System Architecture of Proposed System

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

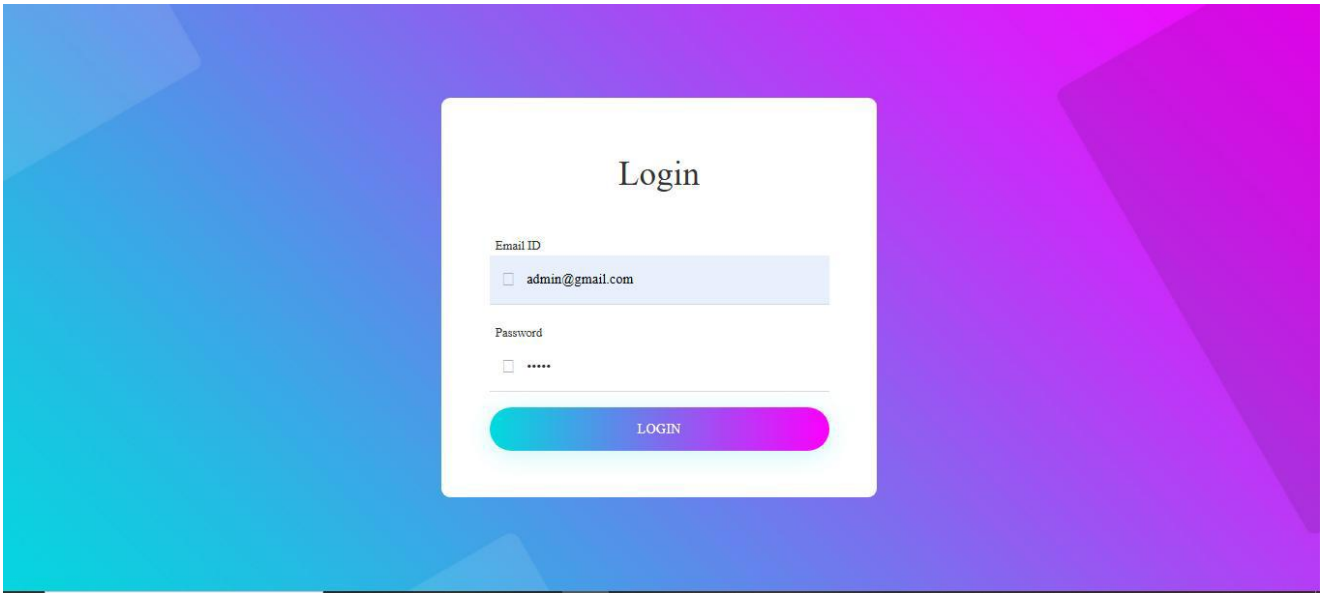
Volume 9, Issue 12, pp 72-77, December-2025

https://doi.org/10.47001/IRJIET/2025.912010

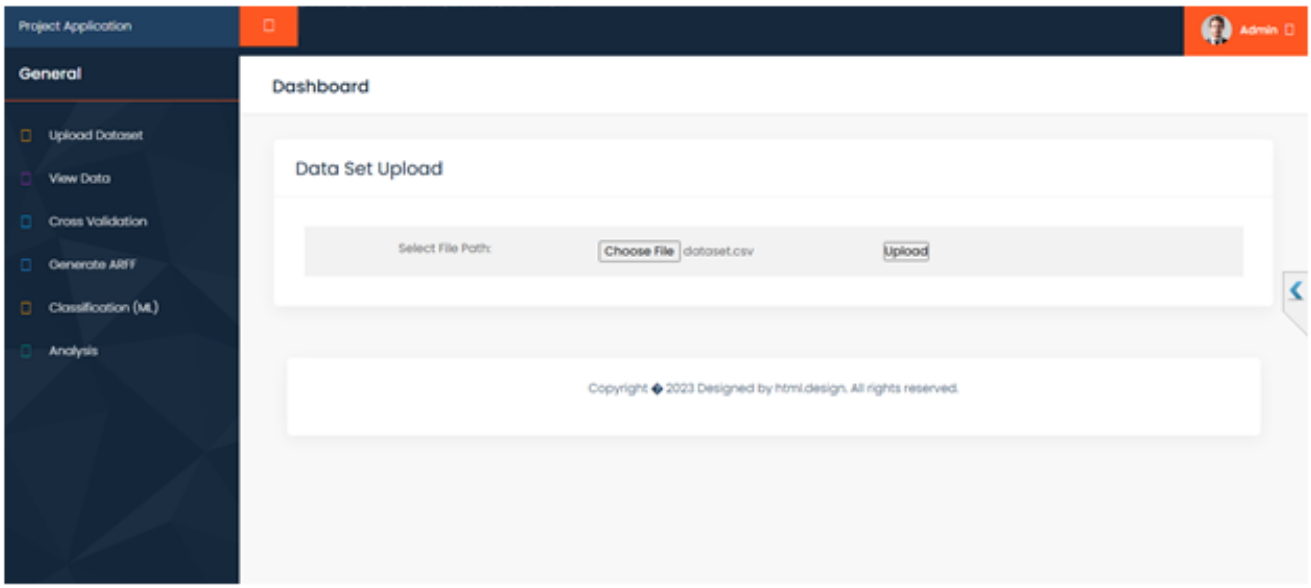
V. RESUTLS

The system workflow commences with a secure login interface designed to authenticate authorized users, including system administrators and data analysts. After successfully authenticating with a registered email address and password, the page directs the user to the primary application dashboard. The initial functional step on the dashboard is uploading a CSV dataset, allowing users to browse, choose a file from their computer, and submit it to the system for storage and subsequent data preparation. The dashboard also provides many features, including viewing datasets, running cross-validation, creating ARFF files, implementing classification models, and analyzing results. Upon uploading the initial dataset, the user advances to the Training and Testing File Upload area, where distinct ARFF files for model training and testing are chosen via dropdown menus. The system initiates data segmentation, trains machine learning models, and evaluates performance upon pressing the Process button. This phase facilitates categorization with algorithms like Random Forest, Naïve Bayes, and Support Vector Machines, allowing for precise model evaluation. These components collectively establish an efficient and systematic framework for data management and fraud detection experimentation.

Login Page



File Upload



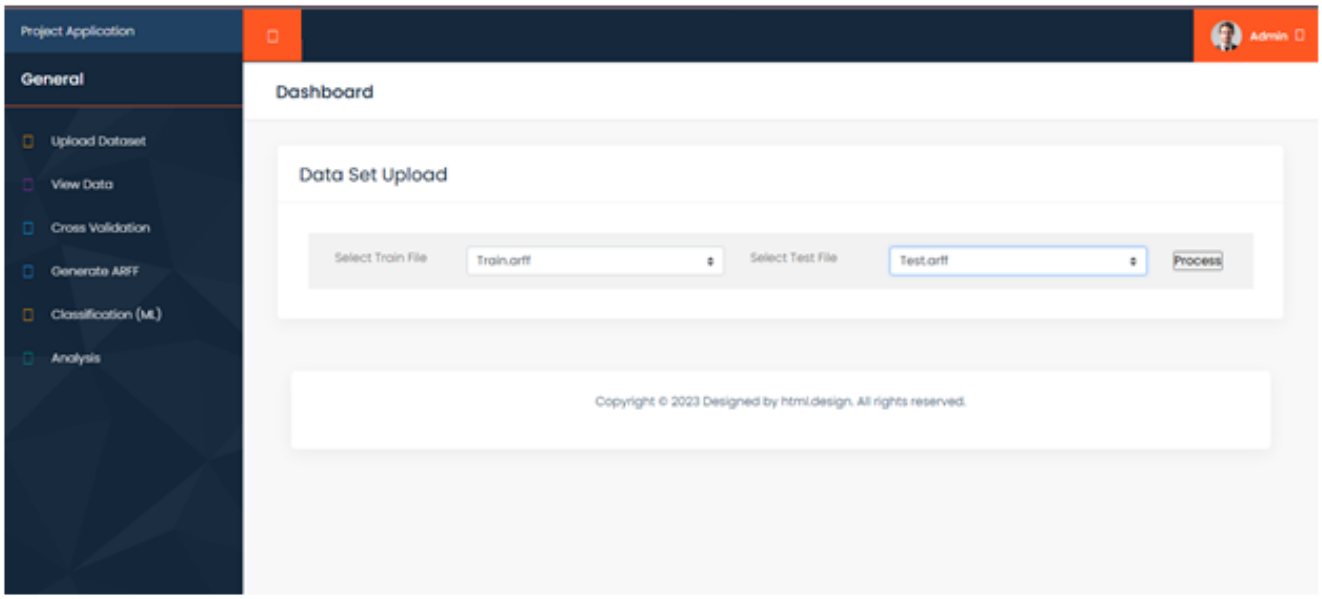
International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 12, pp 72-77, December-2025

https://doi.org/10.47001/IRJIET/2025.912010

Training and Testing File Upload



VI. CONCLUSION

This method proved accurate in identifying fraudulent transactions and minimizing the number of false alerts. Machine learning is appropriate in this kind of application area. The use of this algorithm in a credit card fraud detection system results in detecting or predicting the fraud probably in a very short span of time after the transaction has been made. This will eventually prevent the banks and customers from great losses and also will reduce risks. Machine learning-based credit card fraud detection models provide an effective way to identify and prevent fraudulent transactions in real time. By analysing transaction patterns, these models can accurately distinguish between legitimate and suspicious activities, reduce financial losses, and enhance customer trust. With proper data preprocessing and model optimization, machine learning approaches outperform traditional rule-based systems and can adapt to evolving fraud patterns. Implementing such models in digital payment systems strengthens security and ensures a safer online transaction environment.

REFERENCES

[1] Abdul Salam, Mustafa, *et al.* "Federated learning model for credit card fraud detection with data balancing techniques." *Neural Computing and Applications* 36.11 (2024): 6231-6256.

[2] Mosa, Diana T., *et al.* "CCFD: Efficient credit card fraud detection using meta-heuristic techniques and machine learning algorithms." *Mathematics* 12.14 (2024): 2250.

[3] Chung, Jiwon, and Kyungho Lee. "Credit card fraud detection: an improved strategy for high recall using KNN, LDA, and linear regression." *Sensors* 23.18 (2023): 7788.

[4] Nuthalapati, Aravind. "Smart fraud detection leveraging machine learning for credit card security." *Educational Administration: Theory and Practice* 29.2 (2023): 433-443.

[5] Afriyie, Jonathan Kwaku, *et al.* "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions." *Decision Analytics Journal* 6 (2023): 100163.

[6] Alfaiz, Noor Saleh, and Suliman Mohamed Fati. "Enhanced credit card fraud detection model using machine learning." *Electronics* 11.4 (2022): 662.

[7] Moradi, Fatemeh, M. Tarif, and M. Homaei. "A systematic review of machine learning in credit card fraud detection." *Preprint, MDPI AG* (2025).

[8] Theodorakopoulos, Leonidas, et al. "Big data-driven distributed machine learning for scalable credit card fraud detection using PySpark, XGBoost, and CatBoost." *Electronics* 14.9 (2025): 1754.

[9] Al-Maari, Al-Anood, *et al.* "Optimized Credit Card Fraud Detection Leveraging Ensemble Machine Learning Methods." *Engineering, Technology & Applied Science Research* 15.3 (2025): 22287-22294.

[10] Khalid, Abdul Rehman, *et al.* "Enhancing credit card fraud detection: an ensemble machine learning approach." *Big Data and Cognitive Computing* 8.1 (2024): 6.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 12, pp 72-77, December-2025

https://doi.org/10.47001/IRJIET/2025.912010

Citation of this Article:

Riza Peerzade, Riya Jadhav, Sanika Pathare, Manjushri Jadhav, Prof. Nita Pawar, & Prof. Nita Pawar. (2025). Identifying Fraudulent Credit Card Transactions Using Ensemble Machine Learning. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(12), 72-77. Article DOI https://doi.org/10.47001/IRJIET/2025.912010

****

