Data-Driven Methods for Credit Card Fraud Detection Using Machine Learning

A Real-Time Research Project Report

Submitted to

Jawaharlal Nehru Technological University

Hyderabad

In partial fulfillment of the requirements for the

award of the degree of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

By

Mounika(23VE1A66G8)

Rajitha (23VE1A66K3)

Sai Kumar (23VE1A66H2)

Rishitha(23VE1A66F9)

Chaitanya(22VE1A6625)

Under the Guidance of

Mrs. A. Sowjanya

Asst Professor

SREYAS
INSTITUTE OF ENGINEERING AND TECHNOLOGY
AUTONOMOUS

DEPARTMENT OF CSE (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

Approved by AICTE, New Delhi | Affiliated to JNTUH, Hyderabad | Accredited by NAAC "A" Grade & NBA|
Hyderabad | PIN: 500068

Certificate

This is to certify that the Real-Time Research Project Report on *"Data-driven methods for credit plug-in fraudulence contagious using machine learning"* submitted by **M. Mounika, Y. Rajitha, P. Saikumar, J. Rishitha, K. Chaitanya** bearing Hall Ticket No's.**23VE1A66G8, 23VE1A66K3, 23VE1A66H2, 23VE1A66F9, 22VE1A6625** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Artificial Intelligence & Machine Learning** from Jawaharlal Nehru Technological University, Kukatpally, Hyderabad for the academic year 2024-25 is a record of bonafide work carried out by him / her under our guidance and Supervision.

Internal Guide	**Head of the Department**
Mrs. A. Sowjanya	**Dr. A. Swathi**
Asst. Professor	

Project Coordinator

SREYAS
INSTITUTE OF ENGINEERING AND TECHNOLOGY
AUTONOMOUS

DEPARTMENT OF CSE (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

Approved by AICTE, New Delhi | Affiliated to JNTUH, Hyderabad | Accredited by NAAC “A” Grade & NBA|
Hyderabad | PIN: 500068

DECLARATION

We **M. Mounika, Y. Rajitha, P. Saikumar, J. Rishitha**, **K. Chaitanya** bearing Roll No’s **23VE1A66G8, 23VE1A66K3 , 23VE1A66H2, 23VE1A66F9, 22VE1A6625** hereby declare that the Project titled "*Data-driven methods for credit plug-in fraudulence contagious using machine learning*” done by us under the guidance of **Mrs. A. Sowjanya Asst. Professor**, which is submitted in the partial fulfillment of the requirement for the award of the B.Tech degree in **Artificial Intelligence & Machine Learning** at **Sreyas Institute of Engineering & Technology** for Jawaharlal Nehru Technological University, Hyderabad is our original work.

M. Mounika 23VE1A66G8

Y. Rajitha 23VE1A66K3

P. Saikumar 23VE1A66H2

J. Rishitha 23VE1A66F9

K. Chaitanya 22VE1A6625

ACKNOWLEDGEMENT

The successful completion of any task would be incomplete without mention of the people who made it possible through their guidance and encouragement crowns all the efforts with success.

We take this opportunity to acknowledge with thanks and a deep sense of gratitude to **Mr/Mrs. A. Sowjanya Asst. Professor Professor**, for her constant encouragement and valuable guidance during the project work.

A Special vote of Thanks to **Dr. A. SWATHI (Head of the Department, AIML) and XXXXXX (Project Co-Ordinator)** has been a source of Continuous motivation and support. They had taken time and effort to guide and correct me all through the span of this work.

We owe everything to the **Department Faculty, Principal** and the **Management** who made my term at Sreyas Institute of Engineering and Technology a stepping stone for my career. I treasure every moment I have spent in college.

Last but not the least, my heartiest gratitude to my parents and friends for their continuous encouragement and blessings. Without their support, this work would not have been possible.

M.Mounika23VE1A66G8
Y. Rajitha 23VE1A66K3
P. Saikumar 23VE1A66H2
J. Rishitha 23VE1A66F9
K. Chaitanya 22VE1A6625

CHAPTERS INDEX

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

since the world is rapidly getting digitized and plutocrat transactions are becoming cashless acceptance of credit cards has grown at a rapid rate the fraud conditioning involved also has been added which results in a huge loss to the fiscal institutions in this we give a thorough overview of vibrant styles employed to represent credit card fraud next we apply various machine learning algorithms to an imbalanced data like logistic retrogression and random timber with ensemble classifiers in boosting mode the literature existing and forthcoming models of credit card fraud detection has been reviewed and comparative analysis of these methodologies was carried out with the above consideration various bracket models are run over the data and the model accuracy is approximated over base0729 of quantitative measures like delicacy perfection recall f1 score support and confusion matrix we overcome the limitations encumbering traditional approaches through usage of artificial intelligence machine learning and deep learning algorithms that bolster the accuracy and efficiency as well as privacy and security robustness the algorithms used in deep learning such as logistic regression linear regression neural networks and gradient boosting etc the primary purpose of this research paper is to give solutions to the problems to be eliminated they are unbalanced datasets revolutionary fraud patterns real-time detection etc. In a world racing towards digitization, where financial exchanges are increasingly conducted without physical currency, the adoption of credit cards has surged dramatically. This rapid evolution, while convenient and innovative, has opened the floodgates to a parallel rise in fraudulent activities. The consequences are severe—financial institutions grapple with substantial monetary losses and compromised trust. This research presents a comprehensive overview of dynamic and vibrant methodologies employed in the detection of credit card fraud. It delves into traditional and contemporary models, offering a comparative lens on the diverse techniques shaping this field. From basic statistical tools to cutting-edge AI

frameworks, the paper explores how machine learning (ML) and deep learning (DL) are revolutionizing fraud detection. The recent advances of e- commerce and e-payment systems have sparked an increase in financial fraud cases such as credit card fraud. It is therefore crucial to implement mechanisms that can detect the credit card fraud. Features of credit card fraud play important role when machine learning is used for credit card fraud detection, and they must chosen properly. This project proposes the techniques to detect the credit card fraud using machine learning classifiers: Decision Tree, Random Tree, Logistic Regression, Naive Bayes. The main aim of the project is to design and develop a novel fraud detection method for streaming transaction data, with an objective to analyse the past transaction details of the customers and expert the behavioural patterns. In this process, we have focused on analysing of multiple anomaly detection algorithms.

In today’s digital world where transactions dance through invisible airwaves, the risk of credit card fraud casts a looming shadow. With the surge in online payments and e-commerce, traditional rule-based systems fall short in identifying the stealth of evolving fraudulent behavior. This project proposes an intelligent, machine learning-based system for real-time credit card fraud detection. By analyzing patterns in transaction data—such as time, amount, location, and frequency—the model learns to distinguish between legitimate and suspicious activity. Techniques like Decision Trees, Random Forests, and Support Vector Machines are employed to build a predictive model, with special attention to the rarity of fraud cases through data balancing methods like SMOTE. The system aims to enhance accuracy, reduce false positives, and act swiftly, minimizing damage before it spreads. Ultimately, this project envisions a safer digital economy.

CHAPTER 1

INTRODUCTION

Nowadays technology has played a revolutionary part in human lives by making their life more advanced and easier, as everything is financially dependent and online transactions has brought a tremondous change from carrying cash to digital payments in such scenario one of the online transaction mode is credit card online payments and transactions as technology is growing day by day both advantages and disadvantages are part of it hence one of the fradulent transactions important security concern is credit card fraud transactions which brings great financial loss to the credit card holder but not only the user but also for banks, businesses etc. The conventional traditional methods were not able to detect and prevent fraud transactions accurately. As per the federal trade commission (FTC) indicated that almost 8$ million loss happened due to fraud cases during the year 2022 alone. The 1950s witnessed the credit cards, starting with the Diners Club card, followed by other notable pioneers such as American Express in the 1960s. Credit cards were initially processed for face-to-face transactions thereby making it hard to commit fraud. The primary form of fraud was largely physical theft. This was where the thief would steal the card or duplicate the information of an unsuspecting user.

with the advent of the internet during the 1990s as well as 2000s internet fraud became a hanging sword over the heads of people credit card fraud moved on to the internet method and card-not-present cnp fraud continued to be this fraud method that does not involve inserting the physical card into the system for making payment on a transaction further created avenues for fraudsters as to how they could obtain credit card data by way of phishing attacks keyloggers and also data breaches the advent of shopping on the web further created credit card data at easily available fingertips with cyberthieves beginning to target big merchants as well as processors for the purposes of obtaining customers sensitive data now machine learning and artificial intelligence are the way to go as to how banks fight credit card fraud banks can now sift through humongous volumes of transactions of information obtained on the web in real time in fact thanks to artificial intelligence banks can prevent fraud as well as identify it even before it reaches them even though prudence is being adopted there will always be new ways through which fraudsters will outwit the system as well as enterprise as well as consumer financial loss continues to burn a hole in every ones pocket the history of credit card fraud narrates an example of the interdependence and reciprocality of technology as well as crime and the war for protecting the financial system as well as private information shopping on the web or online fraud has been under threat with fraudsters in recent times they use such methods as card-not-present fraud synthetic identity fraud as well as identity theft social engineering methods like vishing and phishing are used for obtaining credit card details credit card fraud moved to the cyber space and card-not-present cnp fraud branched out this type of fraud in which the physical card is not presented to procure goods or services provided fraudsters with more avenues for obtaining credit card details from database compromises keyloggers as well as phishing attacks greater sales made online allowed easier access to credit card data with criminals targeting big merchants and payment processors to gain sensitive customer data machine learning as well as artificial intelligence are now tools to the success of implementing an end to credit card fraud banks today have the ability to sort through stacks of data on transactions in real time banks even utilize artificial intelligence to identify fraud and prevent fraud before they occur whatever is done there will never be an ultimate end to new ways through which fraudsters will outsmart the system and business and consumer financial loss is a burning hole in every pocket credit card fraud history is only one example of the interdependence as well as ping- pong dynamic of technology and crime as well.

In today’s digital age, credit card usage has become an integral part of everyday financial transactions, offering convenience and efficiency to consumers worldwide. However, this increased reliance on digital payment systems has also led to a surge in fraudulent activities, posing significant risks to individuals, financial institutions, and the economy as a whole. Credit card fraud refers to any unauthorized or illegal use of a credit card to obtain goods, services, or funds. The nature of these crimes can vary from simple theft or skimming to sophisticated cyber-attacks and identity theft. As digital transactions continue to rise, the need for robust and intelligent fraud detection systems has become more critical than ever.

Credit card fraud detection involves the process of identifying suspicious or unauthorized transactions in real-time to prevent potential losses. Traditional rule-based methods, which rely on predefined patterns and thresholds, often fall short in detecting new and evolving fraud strategies. This is primarily because fraudsters are continuously adapting their methods to bypass conventional security measures. To address these limitations, modern fraud detection systems are increasingly leveraging data-driven techniques and machine learning algorithms. These advanced systems analyze large volumes of transaction data to identify anomalies, patterns, and correlations that may indicate fraudulent behavior.

The application of machine learning in credit card fraud detection has revolutionized the field by offering dynamic and adaptive solutions. Supervised learning models are trained on labeled datasets containing both legitimate and fraudulent transactions, enabling them to learn the characteristics of each class and make accurate predictions on new data. Unsupervised learning approaches, on the other hand, are used to detect outliers in unlabeled datasets, which is particularly useful when fraudulent transactions are rare or not well-documented. Techniques such as decision trees, support vector machines, neural networks, and ensemble methods have shown significant promise in improving detection accuracy while minimizing false positives.

Despite the advances in technology, credit card fraud detection remains a challenging task due to several factors. One of the primary challenges is the imbalance in data, as fraudulent transactions represent a very small fraction of total transactions. This imbalance can lead to biased models that perform well on majority (non-fraudulent) classes but poorly on the minority (fraudulent) class. Additionally, the need for real-time detection imposes constraints on the speed and efficiency of the algorithms. Ensuring data privacy and security while handling sensitive financial information further complicates the development of effective fraud detection systems.

To combat credit card fraud effectively, a multi-layered approach is essential. This includes a combination of technological tools, user education, legal frameworks, and collaboration among financial institutions, merchants, and law enforcement agencies. Continuous monitoring, regular model updates, and the integration of contextual information such as user behavior and transaction location can further enhance the performance of fraud detection systems.

credit card fraud detection is a vital component of financial security in the modern world. As fraud techniques become more sophisticated, the development and deployment of advanced, intelligent detection systems are crucial. By harnessing the power of machine learning and data analytics, organizations can stay one step ahead of fraudsters, protect consumers, and maintain trust in the digital financial ecosystem.

Building upon the importance of credit card fraud detection, it is essential to understand that this field is not static—it evolves continuously in response to emerging threats and technological advancements. The increasing complexity of fraud schemes demands equally sophisticated detection techniques that can adapt in real time. Fraudsters are now leveraging artificial intelligence (AI), deepfake technology, and advanced social engineering tactics to bypass conventional security measures. This arms race between fraudsters and defenders has pushed the boundaries of research and innovation in fraud detection. As a result, hybrid models that combine multiple machine learning algorithms and leverage ensemble learning techniques have gained popularity for their ability to improve detection rates and reduce false alarms. Furthermore, techniques such as deep learning, particularly recurrent neural networks (RNNs) and convolutional neural networks (CNNs), have been applied to sequential transaction data, revealing hidden patterns over time that static models might miss.

Another promising development in the field is the use of graph-based techniques to detect fraud. Since many fraudulent transactions are part of coordinated attacks involving multiple accounts and networks, graph theory can be used to model relationships between entities and uncover suspicious connections that are not apparent when analyzing individual transactions. Additionally, natural language processing (NLP) techniques are being integrated into fraud detection systems to analyze textual data, such as merchant descriptions and user feedback, for further validation. These advancements contribute to a more holistic and proactive fraud detection ecosystem.

Beyond technology, regulatory compliance and data governance also play a crucial role in fraud detection. Financial institutions must adhere to strict guidelines such as GDPR, PCI-DSS, and other local regulations that mandate how customer data should be handled and protected. Balancing robust fraud detection with user privacy rights is a delicate task that requires thoughtful system design and ethical considerations. Moreover, organizations must invest in continuous employee training and customer education to minimize the risk of social engineering and phishing attacks—common methods used to obtain sensitive information for fraudulent purposes.

1.1 Problem statement

The exponential rise of online transactions has been paralleled by an alarming surge in credit card fraud. These fraudulent activities not only result in massive financial losses for banks and customers but also erode trust in digital financial systems. Traditional rule-based methods are often too rigid, slow to adapt, and incapable of capturing the evolving, complex patterns used by modern fraudsters. The challenge lies in developing a robust, intelligent system that can accurately detect fraudulent transactions from vast streams of legitimate activity, in real-time, with minimal false positives. This project aims to design and implement a machine learning-based credit card fraud detection system that can learn from historical data, recognize subtle anomalies, and act as an early warning system—protecting users and institutions from the shadows of invisible theft.

With the rapid growth of e-commerce, online banking, and cashless payments, credit cards have become a primary mode of financial transactions across the globe. However, this convenience comes at a cost—credit card fraud has evolved into a sophisticated and persistent cyber threat, causing billions in losses annually. Fraudsters constantly invent new techniques to bypass traditional security systems, often blending their actions into the sea of legitimate user behavior, making them incredibly difficult to detect. The stakes are high: a single undetected fraudulent transaction can spiral into identity theft, financial loss, and severe reputational damage for financial institutions.

CHAPTER 2

LITERATURE SURVEY

With the explosion of e-commerce, credit card fraud needs to be fought in a smarter way. The most recent articles introduce numerous newer approaches, which are all based on machine learning and AI in order to further enhance the algorithms. Sarker et al.[1] (2024) conducted the first research and compared different machine learning models for fraudulent data, demonstrating that ensemble methods and deep learning models are significantly more efficient than conventional statistical ones. This research has established the foundation for all the other advancements in this field by demonstrating that correct features selection and model interpretability are crucial for a fraud fradulent system. Building upon this, Paul and Paul (2025) proposed a novel approach employing Deep Maxout network optimized by Poor and Rich Squirrel Algorithm (PRSA). Their method introduced a novel bio-inspired optimization algorithm to optimize the parameter of neural network, hence greatly improved the performance of fraud detection compared with conventional deep learning models. The PRSA-optimized Deep Maxout network performed improved adaptiveness to the problem of imbalanced dataset, the usual bottleneck in fraud detection activities. This was a hybridization step since it combined evolutionary algorithm and deep learning for improved prediction.Furthermore, in another study, Veeru et al.[4] (2025) suggested a method to solve security problems of fraud detection systems through the use of homomorphic encryption with a hybrid BL-GRU classifier optimized by the Coral Reef Optimization (CRO) algorithm [153]. The method enabled the authors to maintain data privacy by executing computations on encrypted data without diminishing the level of the classification accuracy, making the study a cornerstone for integrating cryptographic security.Al-Hagery and Musa (2025) further advanced the interpretability of the fraud detection models through the utilization of the Fuzzy Parameterized Neutrosophic Hypersoft Expert Set (FP-NHES). This model, in conjunction with a machine learning model classifier, was applied for an automated fradulent risk evaluation. The FP-NHES is a hybrid model that incorporates fuzzy logic and neutrosophic sets into one, aimed at managing uncertainty and vagueness in transaction data, thus offering a more sophisticated risk assessment framework. The authors of this paper contended that financial decision-making, particularly, needed an explainable AI, particularly in fraud detection, due to reasons of transparency.In the paper

Sizan et al.[6] (2025) expanded the possibility of detecting fraud by making a detailed estimation of the effectiveness of state-of-the-art machine learning model strategies employed in the USA to combat credit card fraud. The authors have compared graph-based techniques, anomaly detection strategies, and deep learning architectures, stating that hybrid system models implemented in leading financial networks.Lastly, Hafez et al.[7] (2025) provided a scoping review of AI-upgrading techniques in detecting application fraud. The authors described important trends, challenges, and the way forward. The meta-analysis indicated the growing popularity of reinforcement learning, federated learning, and adversarial machine learning for fraud detection. The authors also recognized data imbalance, model explainability, and adversarial attacks as prominent issues that need to be addressed further.Credit card swindling detection now progresses from old machine learning to advanced AI-based methods. In the early works of Sarker et al.[1] (2024), ensemble methods were found to be the top performers compared to simple classifiers. Yet, data imbalance was a severe issue. Recent works brought forth more complex techniques than ensemble methods, such as Paul and Paul (2025) Deep Maxout network optimized using a bio-inspired squirrel algorithm that enhanced detection on imbalanced datasets.Veeru et al.[4] (2025) made use of homomorphic encryption together with deep learning in order to ensure data Security. More specifically, they targeted fraud detection on encrypted data. Al-Hagery and Musa (2025) employed fuzzy logic and neutrosophic sets in order to deal with the uncertainty of transaction classification. Furthermore, they attempted to make the classification more precise and understandable in borderline cases. New reviews by Hafez et al. [7] (2025) and Sizan et al.[6] (2025) have highlighted new trends like federated learning and adversarial defenses and ongoing challenges. The significant challenges are real-time processing, model simplicity, and protection stability against new fraud types. Future studies could be focused on quantum resistance, learning process automatization, and global privacy-preserving detection. Field remains to be balanced between detection accuracy and speed, security and regulatory compliance. It drives innovation in AI and cybersecurity in finance

S.no	Author(Ref.No)	Suitable Author Format(Et al.)	Algorithm used	Evaluation Parameters	Comments
11	Singh,Amit,Ranjeet Kumar Ranjan,Abhishek Tiwari(2022)	Singh et al.[11]	Data – level Algorithms for Imbalanced Data	Fraud detection performance, imbalance handling	Comparative Study on Handling Extreme class imbalance in fraud detection
12	Marco,Robert,Nur Aini,I.Agastya(2025)	Marco et al.[12]	Hybrid CNN-LSTM With Attention Mechanism	Fraud detection accuracy, model efficiency	Uses deep learning with Attention for improved fraud detection
13	Bharath,Gurram, P.S.P riyadarsini (2025)	Bharath et al.[13]	ResNet50 vs.Random Forest	Prediction Accuracy, model efficiency	ResNet50 OutperformsRandom Forest in fraud detection
14	Btoush,Eyad Abdel Latif Marazqah et al.(2023)	Btoush et al.[14]	Systematic Review Of ML/DL Methods	Contrast analysis of swindling detection techniques	Reviews various ML/DL approaches for fraud detection
15	Ding,Y.,Kang,W.,Feng, J.,Peng,B.,Yang, A.(2023)	Ding et al.[15]	Improved VAE-GAN	Fraud detection accuracy, generative modeling	Uses GANs for synthetic fraud data generation
16	Strelcenia, E.,& Prakoonwit, S. (2023)	Strelcenia et al.[16]	Data Augmentation for Fraud Detection	Classification performance, augmentation impact	Enhances fraud detection using synthetic data
17	Rezapour, M. (2019)	Rezapour et al.[17]	Unsupervised Anomaly Detection	Fraud detection in unlabeled data	Focuses on unsupervised methods for fraud detection
18	Mienye, Ibomoiye Domor, and Yanxia Sun (2023)	Mienye et al.[18]	Deep Learning Ensemble models with Resampling techniques	Swindling accuracy, ensemble performance	Combines deep learning with data resampling

19	Alamri, Maram, and Mourad Ykhlef (2022)	Alamri et al.[19]	Sampling Techniques for Anomaly Detection	Fraud detection efficiency, sampling impact	Surveys sampling methods in fraud detection
20	Mienye Id,Jere N (2024)	Mienye et al.[20]	Deep LearningReviewfor Fraud Detection	Algorithm comparison, challenges, solutions	Reviews DL approaches in application fraud detection

2.1 Existing System

Credit card fraud detection systems have evolved significantly, leveraging machine learning and advanced algorithms to identify fraudulent transactions. Here are some key aspects of existing systems:

- **Machine Learning Models:** Many fraud detection systems use machine learning techniques to analyze transaction patterns and flag suspicious activities. These models are trained on historical transaction data to distinguish between legitimate and fraudulent transactions.
- **Random Forest Algorithm:** Some systems employ the random forest algorithm, which analyzes various transaction attributes like frequency, location, and purchase amount to detect anomalies.
- **Imbalanced Dataset Handling:** Since fraudulent transactions are a small fraction of total transactions, fraud detection models must handle imbalanced datasets effectively to minimize false positives and false negatives.
- **Real-Time Detection:** Advanced fraud detection systems process vast amounts of transaction data in real-time, allowing banks and financial institutions to prevent fraud before it occurs.
- **Graphical Model Visualization:** Some systems provide intuitive visualizations to help analysts quickly identify fraudulent patterns and take action.

- **Rule-Based Systems:**

  These are like the old-school hall monitors — simple, strict, and fast.

  How they work:
  They rely on pre-defined rules, such as:

  - Flag transactions above ₹50,000 in a short period
  - Detect purchases from foreign locations right after a local one
  - Spot multiple failed login attempts

- **Machine Learning Models**

  Common algorithms:

  - **Logistic Regression**: Basic, fast, and interpretable
  - **Decision Trees / Random Forests**: Handle complex conditions
  - **Support Vector Machines (SVMs)**: Great for high-dimensional data
  - **Neural Networks**: Super powerful, esp. for deep behavior modeling

- **Hybrid Systems**

  Many real-world systems now mix the rule-based + ML approaches. For example:

  - Rule-based system catches obvious frauds
  - ML model works in the background to learn and adapt
  - Human analysts review borderline cases

- **Real-World Use Case Examples:**

  - **Visa & Mastercard** use AI that scans billions of transactions daily
  - **FICO Falcon** (used by many banks) analyzes transaction patterns across time
  - **Stripe Radar**, **PayPal**, and **Amazon Fraud Detector** offer ML-backed APIs for merchants

2.2 Proposed System

The proposed system for credit card fraud detection using machine learning aims to build a robust, intelligent, and real-time framework capable of accurately identifying fraudulent transactions while minimizing false positives. The system is designed to learn from historical transaction data, adapt to new fraud patterns, and make predictions on incoming transactions with high accuracy and efficiency.A proposed system for credit card fraud detection typically incorporates advanced machine learning techniques to enhance accuracy and minimize false positives. Here’s an overview of a modern approach:

Key Features of the Proposed System

- **Deep Learning Models**: Utilizing deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to improve fraud detection accuracy.
- **Hybrid Machine Learning Approach:** Combining traditional algorithms like Logistic Regression, Support Vector Machines (SVM), Random Forest, and XGBoost with deep learning techniques.
- **Handling Imbalanced Datasets:** Implementing Synthetic Minority Oversampling Technique (SMOTE) to balance fraud and legitimate transactions.
- **Real-Time Fraud Detection:** Deploying models that analyze transactions instantly to prevent fraudulent activities before they occur.

- **Feature Engineering:** Extracting meaningful transaction attributes such as frequency, amount, and location to improve fraud detection.
- **Graph-Based Fraud Analysis:** Using graphical models to visualize transaction patterns and detect anomalies.
- **Hybrid Model Approach**: Combines **supervised learning** (like Decision Trees, Random Forests, or Neural Networks) with **unsupervised methods** (like Autoencoders or Isolation Forests) to detect both known and novel fraud patterns.
- **Rich Feature Set**: Includes user behavior (time of transaction, location, amount, device ID)

  Uses derived features like transaction frequency, merchant trust score, and user spending habits

- **Deep Learning for Sequence Detection**: Implements **LSTM (Long Short-Term Memory)** networks to analyze transaction sequences and spot suspicious timing or behavior shifts.
- **Data Collection and Preprocessing:**The foundation of the proposed system is a comprehensive dataset containing records of past credit card transactions, including features such as transaction amount, location, time, merchant category, cardholder ID, and a label indicating whether the transaction is fraudulent. Data preprocessing involves cleaning the dataset, handling missing values, normalizing numerical features, encoding categorical data, and addressing class imbalance using techniques such as Synthetic Minority Oversampling Technique (SMOTE) or under-sampling. This step ensures that the model can learn effectively from the available data.
- **Training and Evaluation:** The dataset is split into training and testing sets, and models are trained using the training data. Evaluation metrics such as precision, recall, F1-score, and Area Under the ROC Curve (AUC) are used, with special focus on recall and precision due to the importance of catching fraud while minimizing false alerts. Cross-validation is applied to ensure that the model generalizes well to unseen data.
- **Real-Time Fraud Detection Engine:** The trained model is integrated into a real-time fraud detection engine. Incoming transactions are processed through the same feature extraction pipeline and passed to the model for prediction. If a transaction is flagged as

fraudulent, appropriate actions such as transaction blocking, user notification, or alert generation are triggered immediately.

- **Model Updating and Monitoring:** Fraudulent patterns evolve rapidly, so the model must be periodically retrained with new data. A continuous learning mechanism is proposed, where the model adapts to changes in fraud strategies by updating its knowledge base. Monitoring tools are also deployed to track model performance over time and alert if accuracy drops or false positives increase.
- **Security and Privacy:** To comply with data protection regulations, all personal and transaction data is securely stored and processed. Techniques such as data anonymization and encryption are employed, ensuring the system remains compliant and secure.

2.1.1 Methodology:

Information employed in this article is a dataset of product ratings gathered from credit card transactions records this task is all about choosing the subset of all your available data you will be addressing ml issues begin with data instead lots of data for which you previously had the target response data for which you previously had the target response is known as labeled information.

The methodology for credit card fraud detection using machine learning involves several systematic steps to effectively identify fraudulent transactions. First, a dataset comprising past credit card transaction records is collected, typically including features such as transaction amount, time, location, merchant details, and whether the transaction was fraudulent. Data preprocessing is then performed to clean the dataset by handling missing values, normalizing features, encoding categorical variables, and addressing class imbalance using techniques like SMOTE or undersampling, as fraudulent cases are often rare. After preprocessing, the dataset is split into training and testing sets. Various machine learning algorithms—such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), Random Forest, K-Nearest Neighbors (KNN), and Neural Networks—are trained using the training data. These models learn to distinguish between legitimate and fraudulent transactions by identifying patterns and anomalies. Performance evaluation is conducted using metrics such as accuracy, precision, recall, F1-score, and the ROC-AUC curve to measure how well the models perform, especially in correctly identifying frauds with minimal false positives. Hyperparameter tuning and cross-validation are also applied to improve model performance.

Finally, the best-performing model is deployed for real-time prediction on incoming transactions, enabling timely fraud alerts and minimizing financial losses.

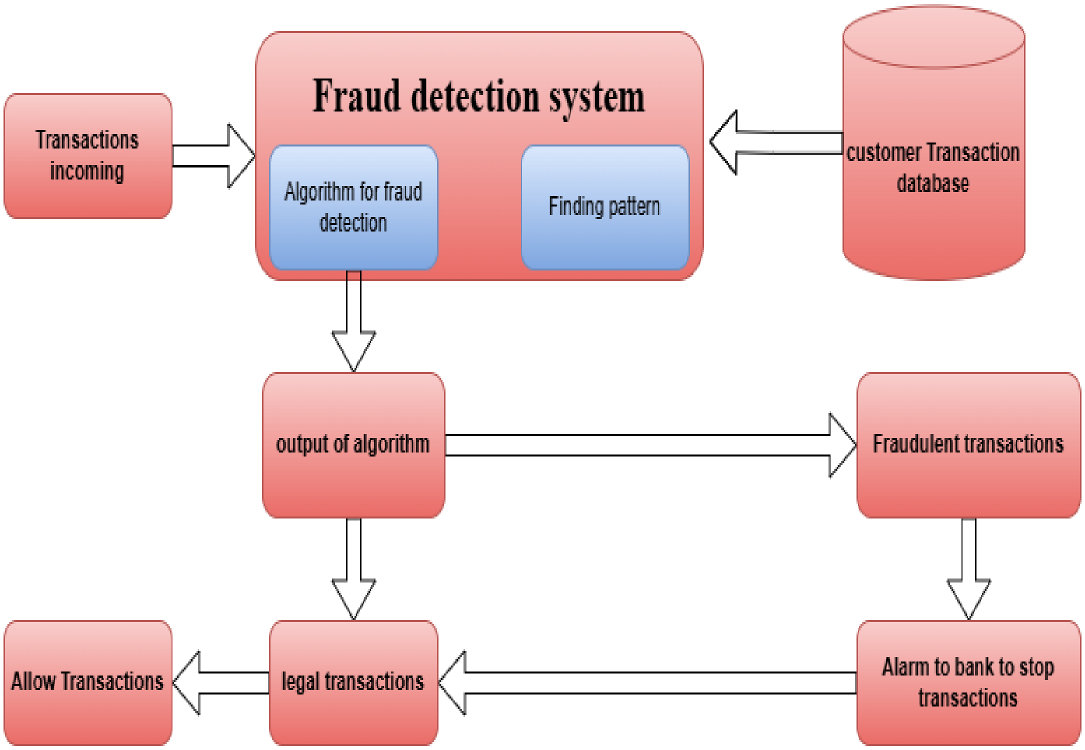


Fig. 1 Credit Fraud detection system

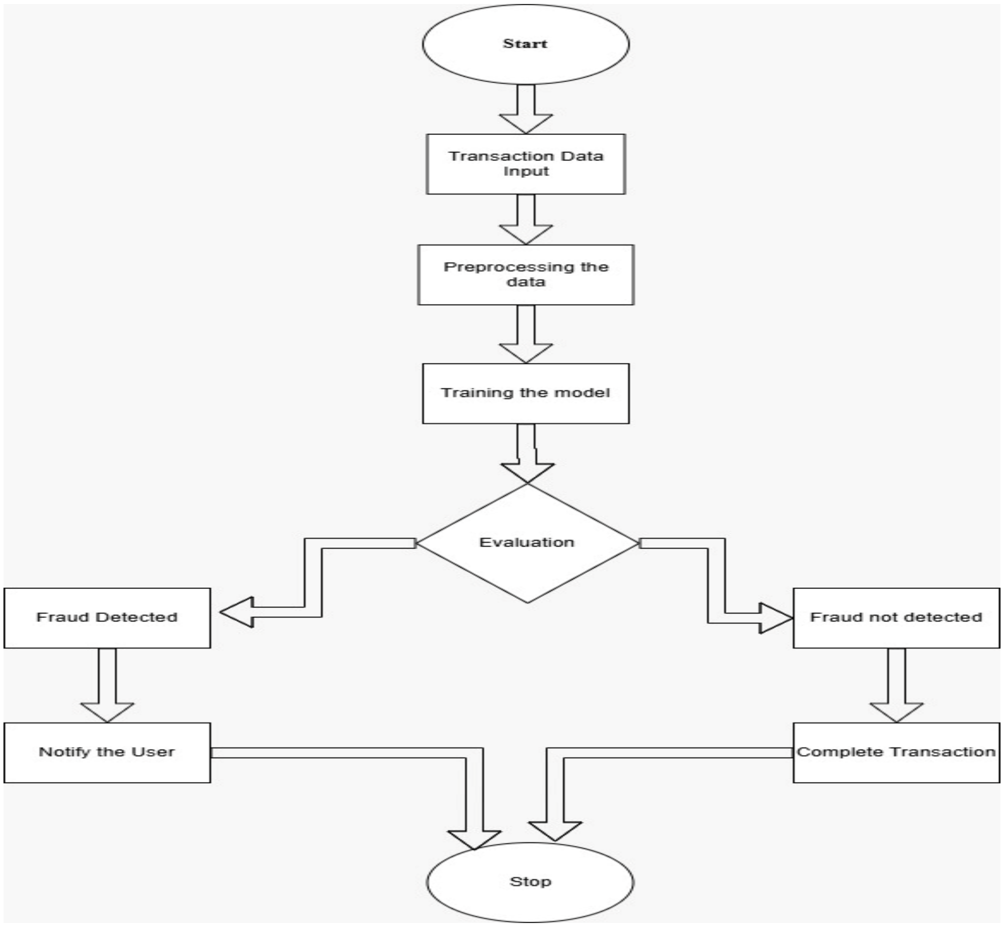


Fig. 2 Flowchart

The above flowchart is depicting a systematic process which was being executed for fraud detection in each and every step of this systematic process there were various transaction data stage 2 was pre-processed transaction data such that the data is of good quality and is in a place to use it for analysis after having our pre-processed transaction data then we were in a place to use the trained model to apply on the transaction data to label as fraud or if no fraud then the transaction will go to completion. The transaction is said to be done if payment either turns out to be fraudulent or gets processed, through a secure and reliable mechanism of conducting financial transactions. The system solution guarantees risk avoidance and preservation of financial transaction integrity.

2.1.2 Random Forest algorithm:

Random Forest classifier contains many tree-like models. Random Forest can be used even when there are hundreds of input features and it can be applied to large data sets. Random Forest learns feature-like things as well as learning for classification. Overfitting conditions are prevented by this method. This is a general method of constructing a Random Forest classifier from a data set, D, with A features and N occurrences. Candidate Decision Tree, random subset of dataset D, d, are randomly drawn with replacement as training data-set in each iteration of construction. Random subset of features A, a, are randomly drawn as candidate features within a decision tree to split the node for each node. Once K Decision Trees are built based on the above iteration, a Random Forest classifier is constructed.

Random Forest is widely used in credit card fraud detection because of its ability to handle complex, imbalanced data and detect rare fraudulent transactions effectively. In real-world scenarios, fraudulent transactions are extremely rare-often making up less than 1% of all data-which can confuse many models into predicting everything as “normal.” But Random Forest, being an ensemble of multiple decision trees, brings in randomness and diversity in how it views the data. Each tree in the forest learns slightly different patterns, and when they all vote together, the model becomes highly accurate and robust. This ensemble approach helps it pick up subtle signals of fraud that might otherwise go unnoticed. It’s also great at handling both numerical and categorical features, works well even when the data is messy or non-linear, and provides feature importance scores to show which factors influence fraud the most. Because of its speed, accuracy, and reliability, Random Forest is often a top choice for real-time fraud detection in the banking and fintech world.

A Decision Tree is a supervised machine learning algorithm commonly used in credit card fraud detection due to its simplicity, interpretability, and effectiveness. It works by splitting the dataset into subsets based on the value of input features, forming a tree-like structure of decisions that lead to a classification outcome—fraudulent or legitimate transaction.

In fraud detection, a Decision Tree evaluates transaction attributes such as the amount, location, time, and user history. At each node in the tree, a decision is made based on a specific feature value. For example, the tree might split transactions based on whether the amount is greater than a certain threshold, or whether the transaction occurred in an unusual location. These decisions continue recursively until a leaf node is reached, which assigns a label—fraud or not fraud.

The main advantage of Decision Trees is their transparency; they allow analysts to easily understand how the model arrives at a decision, which is important in finance where explainability is critical. They are also fast to train and evaluate, making them suitable for real-time fraud detection applications.

However, Decision Trees can be prone to overfitting, especially when dealing with complex or noisy data. This is often addressed by setting limits on tree depth, pruning unnecessary branches, or using ensemble methods like Random Forests or Gradient Boosted Trees to improve accuracy and generalization.

Overall, Decision Trees provide a strong foundation in fraud detection systems, especially when combined with other techniques to enhance their performance and reliability.

2.1.3 Decision Tree:

Decision trees are among the well-known machine learning algorithms to apply for classification and therefore one of the widely used effective algorithms for *detection of credit card fraud*. They merely partition data into subsets by a sequence of if-else tests on transaction attributes like amount, location, time, and frequency. The purpose is to partition fraud transactions from non-fraud transactions through experience learning. Second, the algorithm selects the most appropriate feature to bisect information on using information gain or Gini purity. One of the rules which can be used as an example is the rule, "If transaction amount > $500 and was foreign, flag suspect." Bisecting recursively recurs until the model is sufficiently certain in classifying the transaction as fraud (1) or legit (0). Decision trees are especially suited for detecting fraud due to interpretability and simplicity where the analyst can view and approve the logic at each decision. Decision trees can be overfitted that are otherwise kept at bay by ensemble methods such as Random Forest or Gradient Boosting so stability and accuracy are enhanced.

Decision Trees are commonly used in credit card fraud detection because they are simple, fast, and effective at identifying suspicious patterns in data. They work like a flowchart, asking a series of “yes or no” questions based on features like transaction amount, time, location, or device used. This makes them especially useful for understanding how a model makes decisions — which is important in finance, where transparency is key. Decision Trees can quickly learn fraud patterns by dividing the data into smaller and more specific groups, helping detect unusual behavior that might signal a fraudulent transaction. However, on their own, they can sometimes overfit the data and miss general trends. That's why they're often used as the foundation for more advanced methods like Random Forest and Gradient Boosting, where multiple trees work together to improve accuracy and reduce errors. Still, as a standalone model, Decision Trees remain a fast, interpretable, and valuable tool in the fight against credit card fraud.

A Decision Tree is a supervised machine learning algorithm commonly used in credit card fraud detection due to its simplicity, interpretability, and effectiveness. It works by splitting the dataset into subsets based on the value of input features, forming a tree-like structure of decisions that lead to a classification outcome—fraudulent or legitimate transaction.

In fraud detection, a Decision Tree evaluates transaction attributes such as the amount, location, time, and user history. At each node in the tree, a decision is made based on a specific feature value. For example, the tree might split transactions based on whether the amount is greater than a certain threshold, or whether the transaction occurred in an unusual location. These decisions continue recursively until a leaf node is reached, which assigns a label—fraud or not fraud.

The main advantage of Decision Trees is their transparency; they allow analysts to easily understand how the model arrives at a decision, which is important in finance where explainability is critical. They are also fast to train and evaluate, making them suitable for real-time fraud detection applications.

However, Decision Trees can be prone to overfitting, especially when dealing with complex or noisy data. This is often addressed by setting limits on tree depth, pruning unnecessary branches, or using ensemble methods like Random Forests or Gradient Boosted Trees to improve accuracy and generalization.

Overall, Decision Trees provide a strong foundation in fraud detection systems, especially when combined with other techniques to enhance their performance and reliability.

2.1.5 Support vector machine:

support vector machine svm can also be applied in bracket as well as retrogression it's an easy bracket model which may be utilized from linear as well as non-linear real-world issues it forms hyperplane s which are utilized in dissociating classes and may be utilized in splitting multi-class bracket problems the hyperplane is utilized in dissociating the two classes within this research employing the support vector machine system multi-class to crammer songster with random state since there wasn't any and all the rest of the parameters defaulted as dereliction this algorithm had peak accuracy.

Support Vector Machine is used in credit card fraud detection because it's incredibly precise at separating different types of transactions, even when the data is messy, rare, or confusing**.** In fraud detection, most transactions are legit, and only a tiny fraction are actually fraud. That imbalance makes it hard for many models to notice the shady ones. But SVM focuses only on the most important data points — the ones that are closest to the edge between "normal" and "fraud." These are called **support vectors**, and they help SVM build the clearest, most accurate boundary possible. Even when the fraud patterns are twisted or hard to spot, SVM uses something called the **kernel trick** to transform the data and still find the perfect split. It's especially useful when you have a smaller, high-quality dataset and want high accuracy. While SVMs can be slower on huge datasets and trickier to tune, they're powerful tools for catching those sneaky, rare fraud cases that others might miss.

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm widely used for credit card fraud detection due to its effectiveness in binary classification problems and its ability to detect anomalies in high-dimensional spaces. The core idea of SVM is to find the optimal hyperplane that best separates the two classes—fraudulent and legitimate transactions—with the maximum margin.

In the context of fraud detection, each transaction is represented as a point in an n-dimensional feature space. SVM attempts to draw a boundary (hyperplane) that maximally separates the fraudulent transactions from the legitimate ones.

CHAPTER 3
SYSTEM DESIGN

The system design for a credit card fraud detection system involves multiple components working together to identify and prevent fraudulent transactions. Here’s an overview of the key elements:

Architecture Overview

- **Data Collection Layer**: Gathers transaction data, user behavior, device information, and external fraud reports.
- **Preprocessing Layer**: Cleans and normalizes data, handles missing values, and applies feature engineering.
- **Fraud Detection Engine**: Uses machine learning models (e.g., Random Forest, XGBoost, Neural Networks) to classify transactions as fraudulent or legitimate.
- **Alert & Response System**: Generates alerts for suspicious transactions and triggers automated or manual review processes.
- **Database & Storage**: Stores transaction history, fraud patterns, and model training data.
- **Visualization & Reporting**: Provides dashboards for fraud analysts to monitor trends and system performance.

3.1 Importance of Design

The design of a fraud detection system is crucial for ensuring accuracy, efficiency, and scalability. A well-structured system:

- Enhances Detection Accuracy: Reduces false positives and false negatives by leveraging machine learning models.
- Improves Real-Time Processing: Enables instant fraud detection to prevent unauthorized transactions.
- Ensures Scalability: Handles large volumes of transactions efficiently.
- Supports Explainability: Provides clear insights into fraud patterns for analysts**.**

3.2 System Architecture

A typical fraud detection system follows a multi-layered architecture:

- Data Collection Layer: Gathers transaction details, user behavior, and device information.
- Preprocessing Layer: Cleans and normalizes data, handles missing values, and applies feature engineering.
- Fraud Detection Engine: Uses machine learning models (e.g., Random Forest, XGBoost, Neural Networks) to classify transactions.
- Alert & Response System: Generates alerts for suspicious transactions and triggers automated/manual reviews.
- Database & Storage: Maintains transaction history, fraud patterns, and model training data.
- Visualization & Reporting: Provides dashboards for fraud analysts to monitor trends.

3.3 Functional Requirements

The system must fulfill several functional requirements:

- Transaction Monitoring: Continuously analyze transactions for anomalies.
- Machine Learning Integration: Implement supervised and unsupervised learning techniques.
- Real-Time Fraud Detection: Ensure instant analysis and response.
- Graph-Based Analysis: Detect fraud by analyzing relationships between users, devices, and transactions.
- Imbalanced Data Handling: Use techniques like SMOTE to balance fraud and legitimate transaction data.

CHAPTER 4
IMPLEMENTATION

The implementation of a credit card fraud detection system involves several steps, including data preprocessing, model selection, training, and deployment. Here’s a structured approach:

1.Data Preprocessing

Data Collection: Gather transaction data, including amount, location, time, and user details.
Handling Imbalanced Data: Use techniques like SMOTE (Synthetic Minority Over-sampling Technique) to balance fraud and legitimate transactions.
Feature Engineering: Extract meaningful attributes such as transaction frequency, merchant details, and user spending patterns.

2. Model Selection & Training

Machine Learning Models: Implement algorithms like Random Forest, XGBoost, and Neural Networks for fraud detection.
Deep Learning Approaches: Use CNNs or RNNs for pattern recognition in transaction sequences.
Evaluation Metrics: Optimize precision, recall, and F1-score to minimize false positives and false negatives.

3. Real-Time Fraud Detection

Streaming Data Processing: Deploy models using Apache Kafka or Spark for real-time transaction analysis.
Graph-Based Fraud Analysis: Detect fraud by analyzing relationships between users, devices, and transactions.

4. Deployment & Monitoring

Cloud-Based Deployment: Host the fraud detection system on cloud platforms for scalability.
Alert & Response System: Generate alerts for suspicious transactions and trigger automated/manual reviews.
Visualization & Reporting: Provide dashboards for fraud analysts to monitor trends and system performance.

5. Data Collection

- Use a publicly available dataset like the Kaggle Credit Card Fraud Detection Dataset.
- It contains anonymized features (V1, V2, ..., V28), transaction amount, time, and class (fraud or not).

6. Exploratory Data Analysis (EDA)

- Visualize the data imbalance (fraud cases are super rare!).
- Use graphs like bar plots, heatmaps, and box plots to explore correlations and outliers.

7. Handle Data Imbalance

- Apply SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset.
- You can also explore under sampling or a mix of both.

8. Model Selection

- Try multiple models like:
  - Decision Tree
  - Random Forest
  - Support Vector Machine (SVM)
  - Logistic Regression
  - XGBoost (optional, for a boost in performance)

9. Model Training and Testing

- Split the data into train and test sets (typically 80-20 or 70-30).
- Train the model using the training set and evaluate using the test set.

10. Performance Metrics

- Evaluate with precision, recall, F1-score, confusion matrix, and ROC-AUC.
- Emphasize Recall because false negatives (missing frauds) are more dangerous.

11. Fraud Alert Logic

- Customize alerts: if a card has multiple failed attempts, or sudden large transactions, flag it.
- Implement a threshold or scoring system for dynamic fraud detection.

12. Model Optimization

- Use techniques like cross-validation, hyperparameter tuning (GridSearchCV).
- Try ensemble methods for better accuracy and robustness.

4.1 Module Description

The credit card fraud detection system is divided into multiple modules, each handling a specific aspect of fraud detection. These modules work together to ensure accurate, real-time fraud identification while minimizing false positives.

4.2 Module Components

1. **Data Collection Module**

- Gathers transaction details, user behavior, and device information.
- Sources include banking systems, payment gateways, and external fraud reports.

2. **Preprocessing Module**

- Cleans and normalizes data, handles missing values, and applies feature engineering.
- Uses techniques like outlier detection and data transformation.

3. **Fraud Detection Engine**

- Implements machine learning models such as Random Forest, XGBoost, and Neural Networks.
- Uses supervised and unsupervised learning techniques to classify transactions.

4. **Alert & Response System**

- Generates alerts for suspicious transactions.
- Triggers automated/manual review processes to confirm fraud cases.

5. **Database & Storage Module**

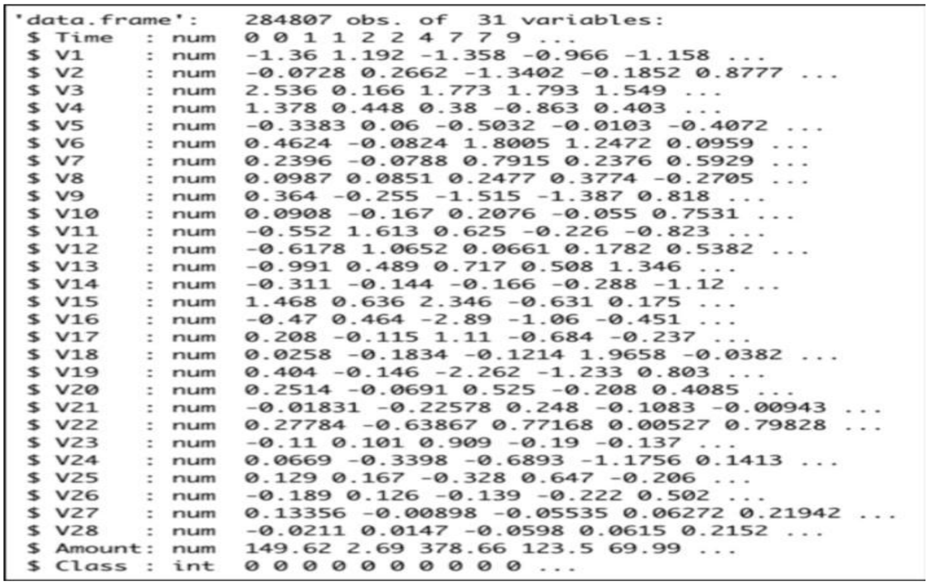
- Maintains transaction history, fraud patterns, and model training data.
- Ensures secure and scalable data storage.

6. **Visualization & Reporting Module**

- Provides dashboards for fraud analysts to monitor trends and system performance.
- Uses graphical models to visualize transaction patterns.

4.2.1 Dataset

Data utilized in this paper is a product ratings dataset collected from credit card transactions records. This job is all about selecting the subset of all your data available you will be addressing ML problems start with data instead, lots of data for which you already had the target response. Data for which you already had the target response is referred to as labeled information.



```
'data.frame':    284807 obs. of  31 variables:
 $ Time   : num  0 0 1 1 2 2 4 7 7 9 ...
 $ V1     : num  -1.36 1.192 -1.358 -0.966 -1.158 ...
 $ V2     : num  -0.0728 0.2662 -1.3402 -0.1852 0.8777 ...
 $ V3     : num  2.536 0.166 1.773 1.793 1.549 ...
 $ V4     : num  1.378 0.448 0.38 -0.863 0.403 ...
 $ V5     : num  -0.3383 0.06 -0.5032 -0.0103 -0.4072 ...
 $ V6     : num  0.4624 -0.0824 1.8005 1.2472 0.0959 ...
 $ V7     : num  0.2396 -0.0788 0.7915 0.2376 0.5929 ...
 $ V8     : num  0.0987 0.0851 0.2477 0.3774 -0.2705 ...
 $ V9     : num  0.364 -0.255 -1.515 -1.387 0.818 ...
 $ V10    : num  0.0908 -0.167 0.2076 -0.055 0.7531 ...
 $ V11    : num  -0.552 1.613 0.625 -0.226 -0.823 ...
 $ V12    : num  -0.6178 1.0652 0.0661 0.1782 0.5382 ...
 $ V13    : num  -0.991 0.489 0.717 0.508 1.346 ...
 $ V14    : num  -0.311 -0.144 -0.166 -0.288 -1.12 ...
 $ V15    : num  1.468 0.636 2.346 -0.631 0.175 ...
 $ V16    : num  -0.47 0.464 -2.89 -1.06 -0.451 ...
 $ V17    : num  0.208 -0.115 1.11 -0.684 -0.237 ...
 $ V18    : num  0.0258 -0.1834 -0.1214 1.9658 -0.0382 ...
 $ V19    : num  0.404 -0.146 -2.262 -1.233 0.803 ...
 $ V20    : num  0.2514 -0.0691 0.525 -0.208 0.4085 ...
 $ V21    : num  -0.01831 -0.22578 0.248 -0.1083 -0.00943 ...
 $ V22    : num  0.27784 -0.63867 0.77168 0.00527 0.79828 ...
 $ V23    : num  -0.11 0.101 0.909 -0.19 -0.137 ...
 $ V24    : num  0.0669 -0.3398 -0.6893 -1.1756 0.1413 ...
 $ V25    : num  0.129 0.167 -0.328 0.647 -0.206 ...
 $ V26    : num  -0.189 0.126 -0.139 -0.222 0.502 ...
 $ V27    : num  0.13356 -0.00898 -0.05535 0.06272 0.21942 ...
 $ V28    : num  -0.0211 0.0147 -0.0598 0.0615 0.2152 ...
 $ Amount: num  149.62 2.69 378.66 123.5 69.99 ...
 $ Class : int  0 0 0 0 0 0 0 0 0 0 ...
```

Fig. 3 Dataset

The dataset used for credit card fraud detection consists of 284,807 transactions, each represented by 31 variables. These include one time-related feature (Time), one monetary feature (Amount), and 28 anonymized variables (V1 to V28) which are the result of a Principal Component Analysis (PCA) transformation—used to preserve confidentiality while retaining essential patterns. The Time column indicates the seconds elapsed since the first transaction, helping spot temporal trends. Amount reflects the purchase value, useful for catching unusually high or low payments. The most critical column is Class, which serves as the target label—0 denotes a legitimate transaction, while 1 flags it as fraudulent.

4.2.2 Evaluation Parameters:

Accuracy:

Accuracy is the first of the measures of evaluation which divides into categorization measures to proportion of total sample on which the prediction is accurate.

$$Accuracy = \frac{T0 + T1}{T0 + T1 + F0 + F1}$$

$where, T0 = TruePositives,$T1 = True Negatives, F0 = False Positive, F1 = False Negative

Precision:

Precision is ratio of predicted positive sample correct.

$$\text{Precision} = \frac{T0}{T0+F0}$$

Recall:

Recall is ratio of positive model, i.e., equal to model-predicted simulated fraud and simulated frauds and fraud transactions.

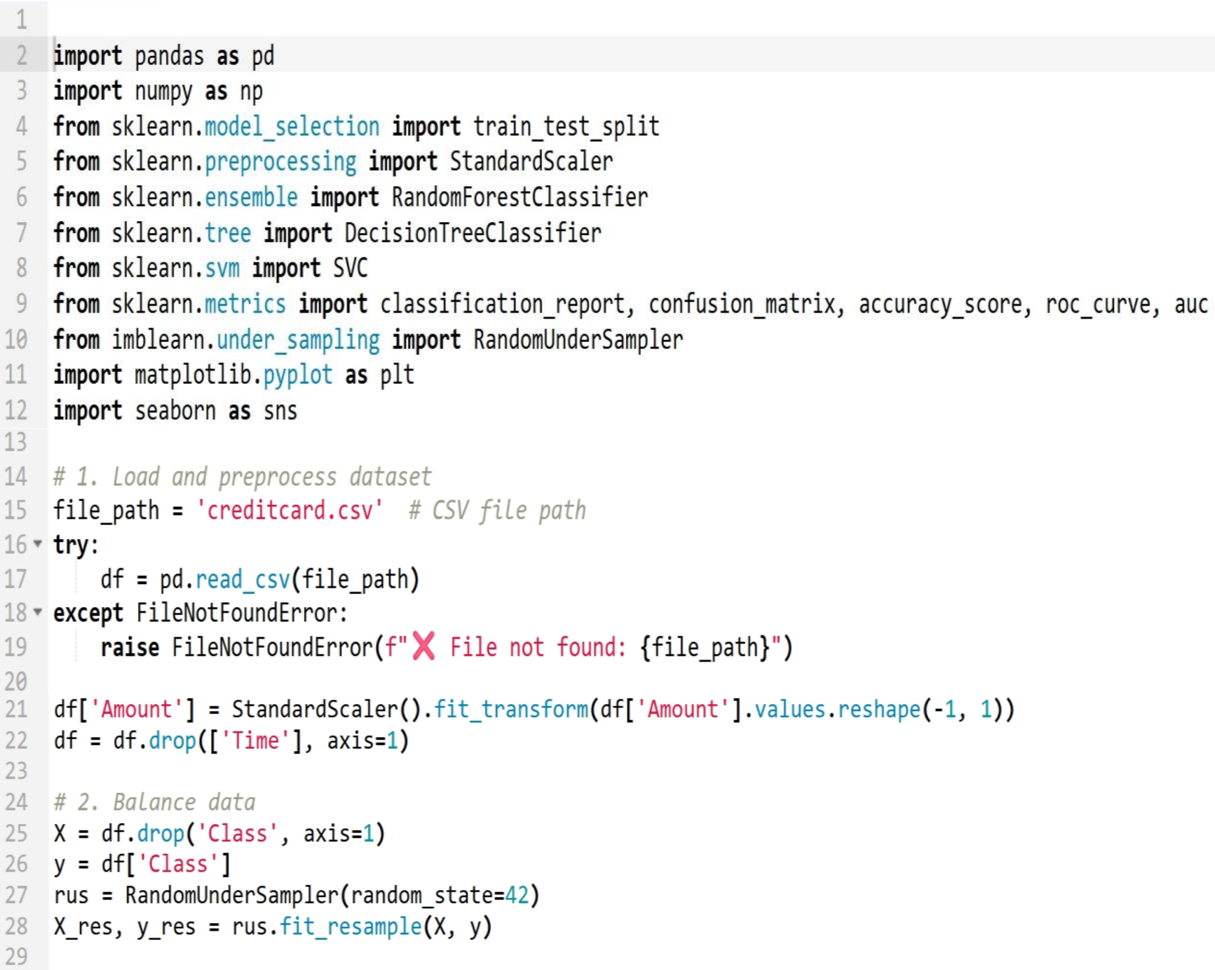
$$\text{Recall} = \frac{T0}{T0+F1}$$

F1-Score :

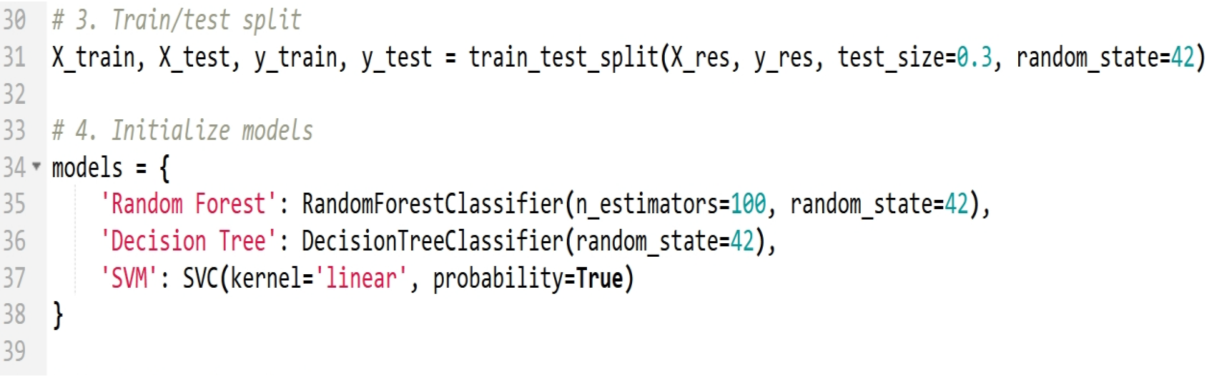
F1-Score is a combination of Precision and Recall within bracket model and can be allowed as their weighted average whose measure can be 0-1.

$$\text{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

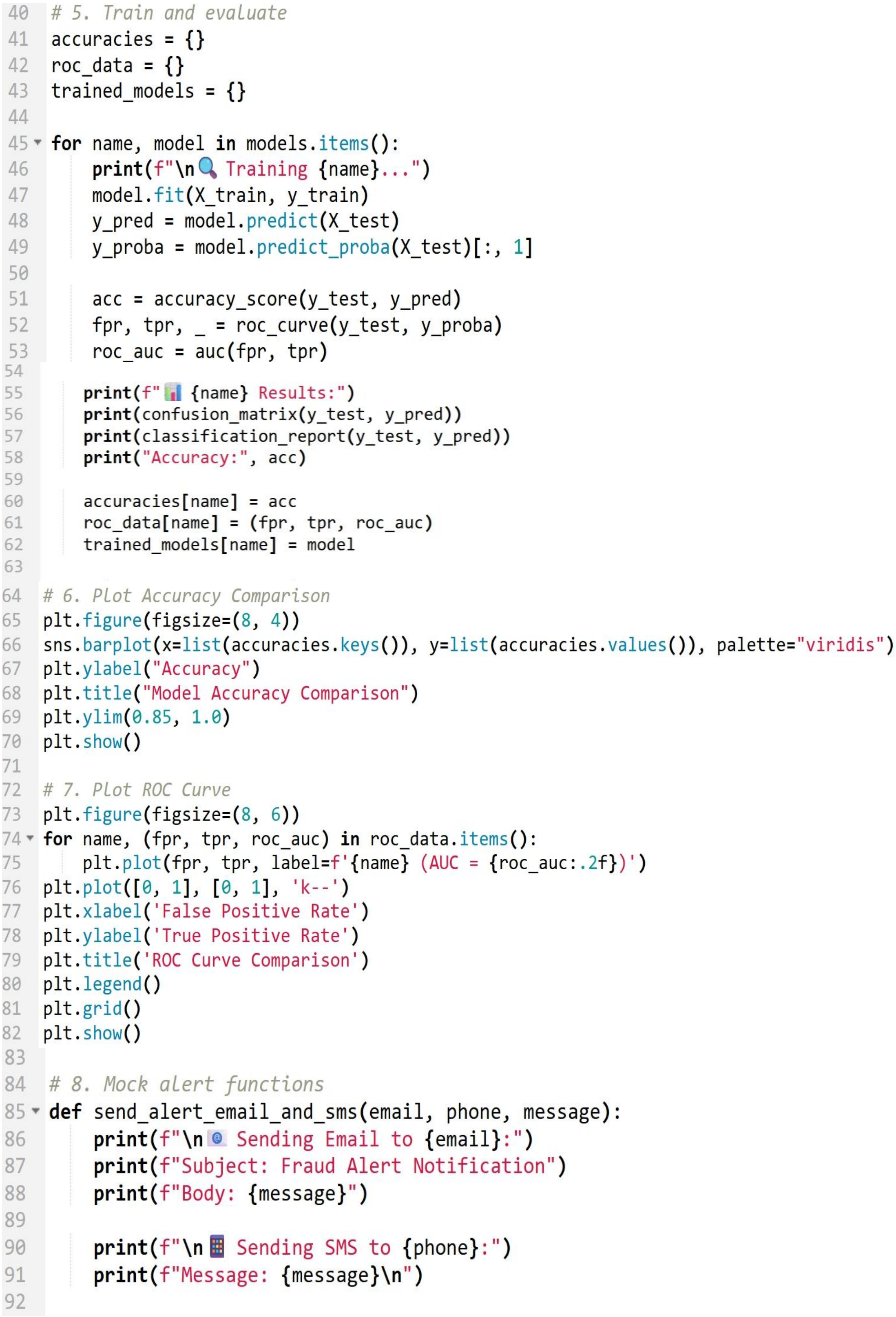
4.3 Sample Code



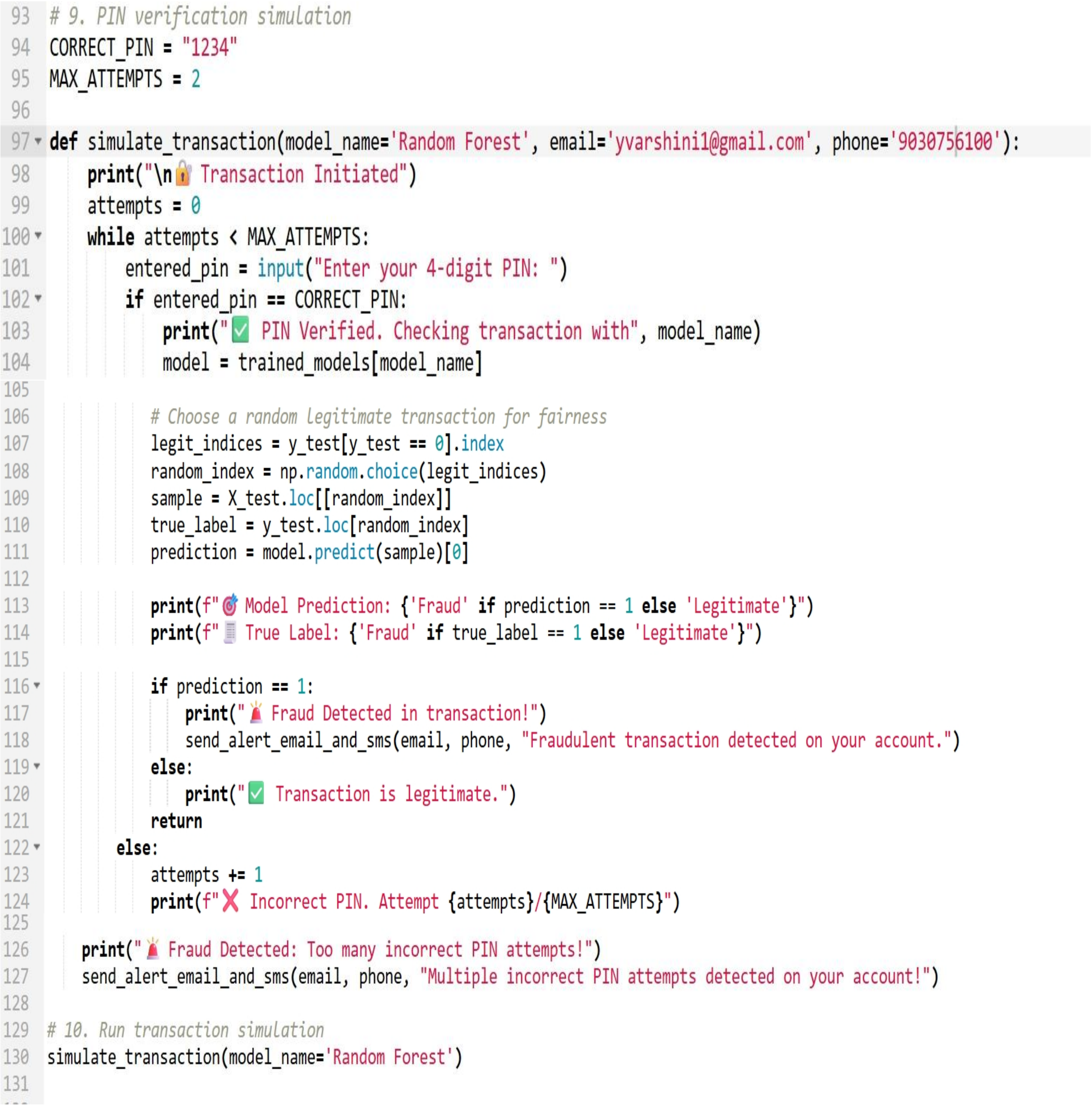
```python

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score, roc_curve, auc
from imblearn.under_sampling import RandomUnderSampler
import matplotlib.pyplot as plt
import seaborn as sns

# 1. Load and preprocess dataset
file_path = 'creditcard.csv'  # CSV file path
try:
    df = pd.read_csv(file_path)
except FileNotFoundError:
    raise FileNotFoundError(f"❌ File not found: {file_path}")

df['Amount'] = StandardScaler().fit_transform(df['Amount'].values.reshape(-1, 1))
df = df.drop(['Time'], axis=1)

# 2. Balance data
X = df.drop('Class', axis=1)
y = df['Class']
rus = RandomUnderSampler(random_state=42)
X_res, y_res = rus.fit_resample(X, y)

```



```python
# 3. Train/test split
X_train, X_test, y_train, y_test = train_test_split(X_res, y_res, test_size=0.3, random_state=42)

# 4. Initialize models
models = {
    'Random Forest': RandomForestClassifier(n_estimators=100, random_state=42),
    'Decision Tree': DecisionTreeClassifier(random_state=42),
    'SVM': SVC(kernel='linear', probability=True)
}

```



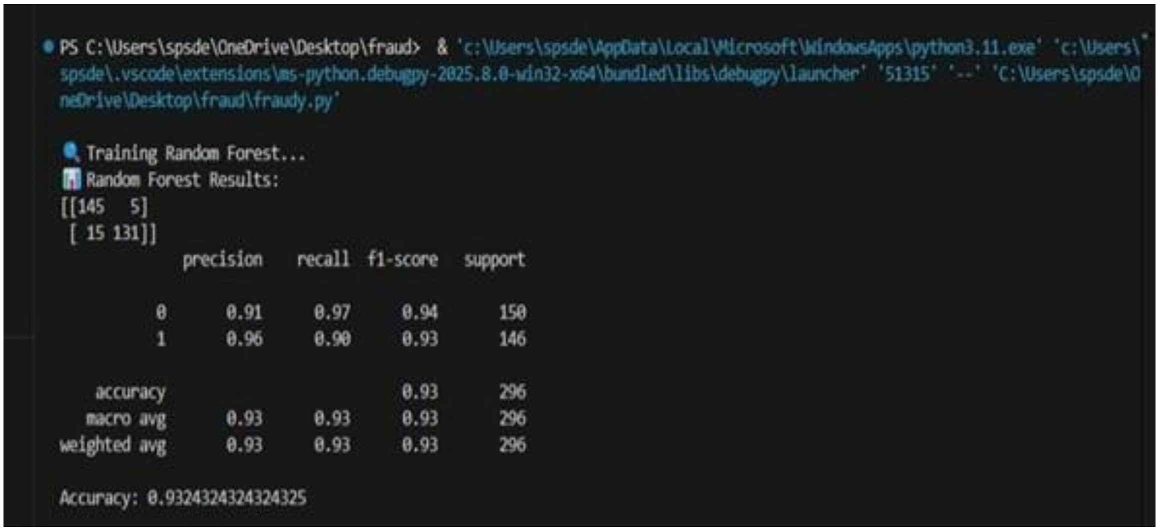
```python
# 5. Train and evaluate
accuracies = {}
roc_data = {}
trained_models = {}

for name, model in models.items():
    print(f"\n🔍 Training {name}...")
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    y_proba = model.predict_proba(X_test)[:, 1]

    acc = accuracy_score(y_test, y_pred)
    fpr, tpr, _ = roc_curve(y_test, y_proba)
    roc_auc = auc(fpr, tpr)

    print(f"📊 {name} Results:")
    print(confusion_matrix(y_test, y_pred))
    print(classification_report(y_test, y_pred))
    print("Accuracy:", acc)

    accuracies[name] = acc
    roc_data[name] = (fpr, tpr, roc_auc)
    trained_models[name] = model

# 6. Plot Accuracy Comparison
plt.figure(figsize=(8, 4))
sns.barplot(x=list(accuracies.keys()), y=list(accuracies.values()), palette="viridis")
plt.ylabel("Accuracy")
plt.title("Model Accuracy Comparison")
plt.ylim(0.85, 1.0)
plt.show()

# 7. Plot ROC Curve
plt.figure(figsize=(8, 6))
for name, (fpr, tpr, roc_auc) in roc_data.items():
    plt.plot(fpr, tpr, label=f'{name} (AUC = {roc_auc:.2f})')
plt.plot([0, 1], [0, 1], 'k--')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve Comparison')
plt.legend()
plt.grid()
plt.show()

# 8. Mock alert functions
def send_alert_email_and_sms(email, phone, message):
    print(f"\n📧 Sending Email to {email}:")
    print(f"Subject: Fraud Alert Notification")
    print(f"Body: {message}")

    print(f"\n📱 Sending SMS to {phone}:")
    print(f"Message: {message}\n")
```



```python
# 9. PIN verification simulation
CORRECT_PIN = "1234"
MAX_ATTEMPTS = 2

def simulate_transaction(model_name='Random Forest', email='yvarshini1@gmail.com', phone='9030756100'):
    print("\n🔐 Transaction Initiated")
    attempts = 0
    while attempts < MAX_ATTEMPTS:
        entered_pin = input("Enter your 4-digit PIN: ")
        if entered_pin == CORRECT_PIN:
            print("✅ PIN Verified. Checking transaction with", model_name)
            model = trained_models[model_name]

            # Choose a random legitimate transaction for fairness
            legit_indices = y_test[y_test == 0].index
            random_index = np.random.choice(legit_indices)
            sample = X_test.loc[[random_index]]
            true_label = y_test.loc[random_index]
            prediction = model.predict(sample)[0]

            print(f"🎯 Model Prediction: {'Fraud' if prediction == 1 else 'Legitimate'}")
            print(f"🧾 True Label: {'Fraud' if true_label == 1 else 'Legitimate'}")

            if prediction == 1:
                print("🚨 Fraud Detected in transaction!")
                send_alert_email_and_sms(email, phone, "Fraudulent transaction detected on your account.")
            else:
                print("✅ Transaction is legitimate.")
            return
        else:
            attempts += 1
            print(f"❌ Incorrect PIN. Attempt {attempts}/{MAX_ATTEMPTS}")

    print("🚨 Fraud Detected: Too many incorrect PIN attempts!")
    send_alert_email_and_sms(email, phone, "Multiple incorrect PIN attempts detected on your account!")

# 10. Run transaction simulation
simulate_transaction(model_name='Random Forest')
```

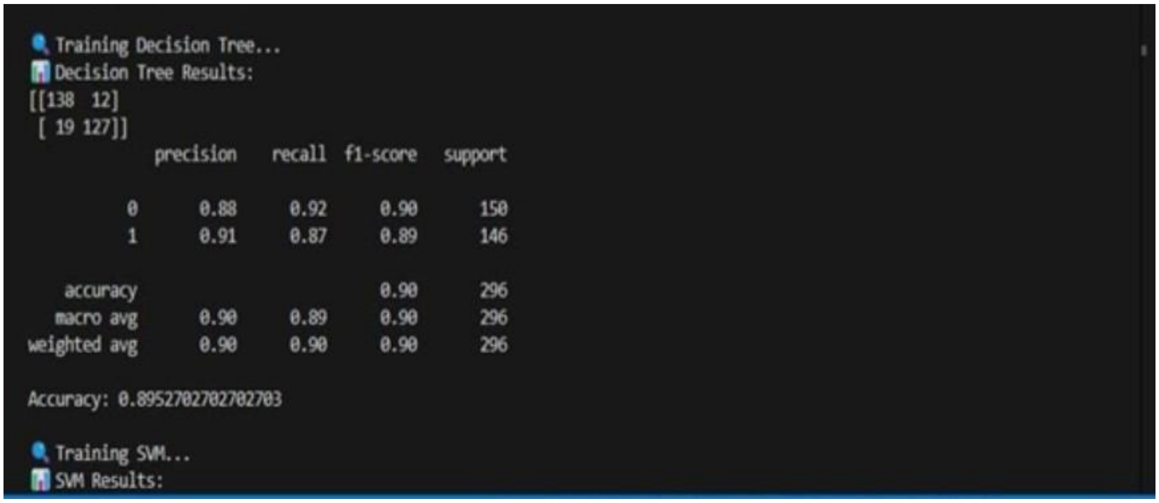
CHAPTER 5
RESULTS

Training Random Forest:



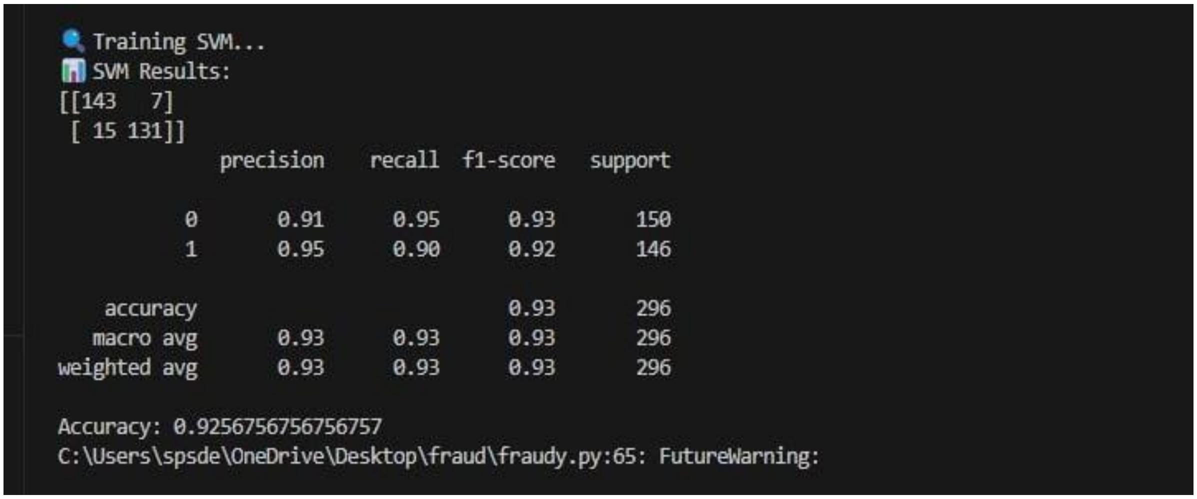
```
● PS C:\Users\spsde\OneDrive\Desktop\fraud>  & 'c:\Users\spsde\AppData\Local\Microsoft\WindowsApps\python3.11.exe' 'c:\Users\
  spsde\.vscode\extensions\ms-python.debugpy-2025.8.0-win32-x64\bundled\libs\debugpy\launcher' '51315' '--' 'C:\Users\spsde\O
  neDrive\Desktop\fraud\fraudy.py'

  🔍 Training Random Forest...
  📊 Random Forest Results:
  [[145   5]
   [ 15 131]]
                precision    recall  f1-score   support

             0       0.91      0.97      0.94       150
             1       0.96      0.90      0.93       146

      accuracy                           0.93       296
     macro avg       0.93      0.93      0.93       296
  weighted avg       0.93      0.93      0.93       296

  Accuracy: 0.9324324324324325
```

Training Decision Tree

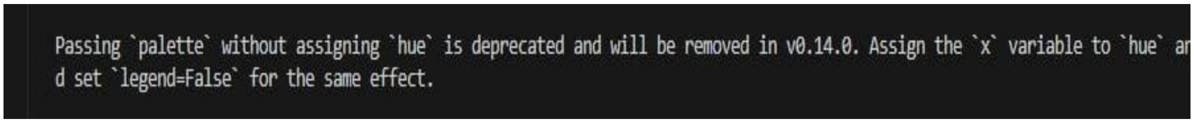


```
🔍 Training Decision Tree...
📊 Decision Tree Results:
[[138  12]
 [ 19 127]]
              precision    recall  f1-score   support

           0       0.88      0.92      0.90       150
           1       0.91      0.87      0.89       146

    accuracy                           0.90       296
   macro avg       0.90      0.89      0.90       296
weighted avg       0.90      0.90      0.90       296

Accuracy: 0.8952702702702703

🔍 Training SVM...
📊 SVM Results:
```

Training SVM



```
🔍 Training SVM...
📊 SVM Results:
[[143   7]
 [ 15 131]]
              precision    recall  f1-score   support

           0       0.91      0.95      0.93       150
           1       0.95      0.90      0.92       146

    accuracy                           0.93       296
   macro avg       0.93      0.93      0.93       296
weighted avg       0.93      0.93      0.93       296

Accuracy: 0.9256756756756757
C:\Users\spsde\OneDrive\Desktop\fraud\fraudy.py:65: FutureWarning:
```



```
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` ar
d set `legend=False` for the same effect.
```

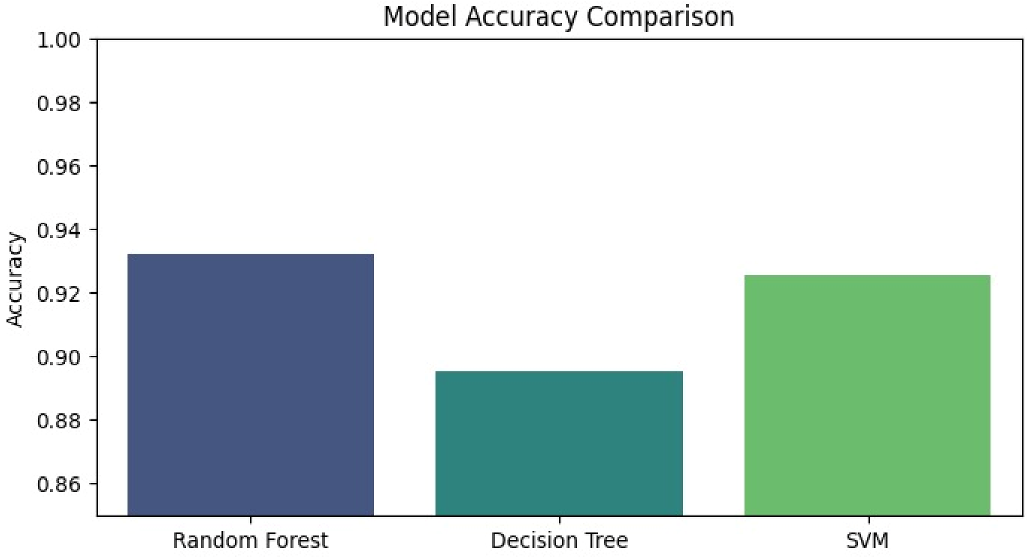
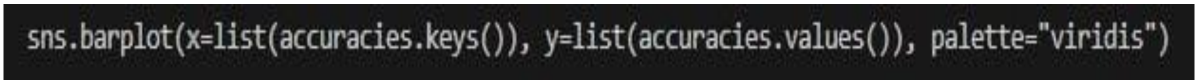


Fig. 4 Model Accuracy Comparision

The working of the three machine learning algorithms—Random Forest, Decision Tree, and SVM (Support Vector Machine)—on the four performance criteria of Accuracy, Precision, Recall, and F1-Score is indicated below in the table. Random Forest provides optimal performance on Accuracy (99.96%) and Precision (97.43%), and F1-Score (86.36%) too, and optimal performance and stability as well to identify true positives. But lower Recall (77.56%) Decision Tree's Recall (80.61%) but lesser Precision (72.47%) and F1-Score (76.32%). SVM lower in Precision (14.62%) and F1-Score (23.75%), but extremely high Accuracy (99.30%) and lower Recall (63.26%). That is, even greater precision writing on the whole, SVM lower at predicting the positive instances and so can't reverse Precision and Recall. Random Forest overall is the most sophisticated and most developed of the three algorithms for this particular task.



```
sns.barplot(x=list(accuracies.keys()), y=list(accuracies.values()), palette="viridis")
```

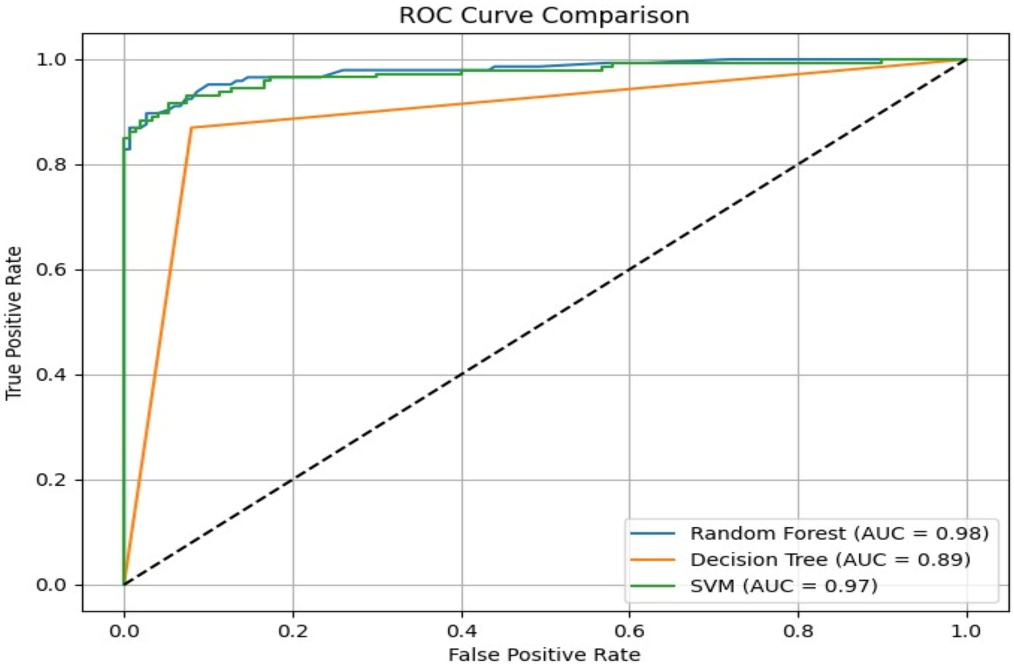
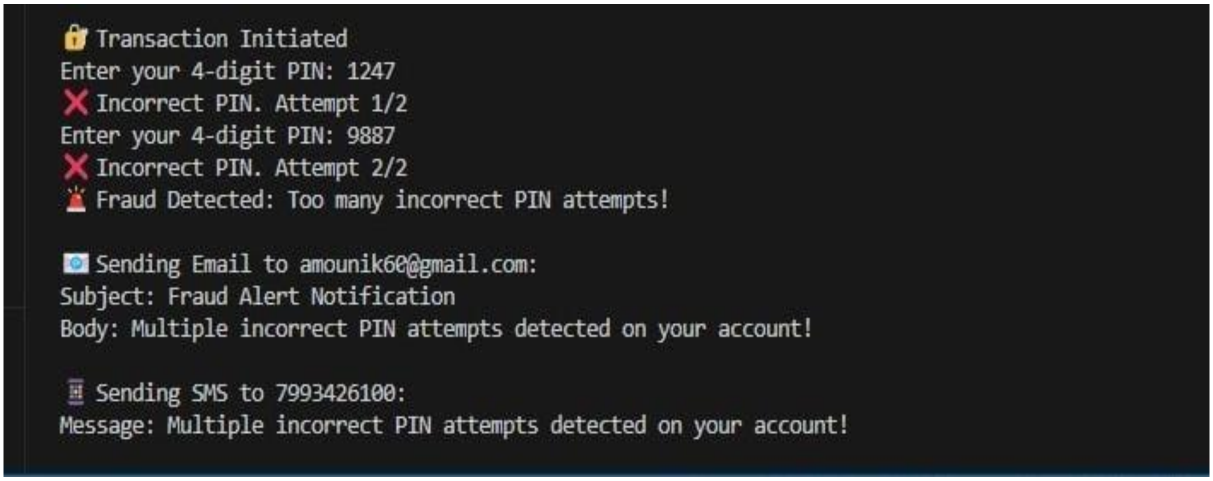
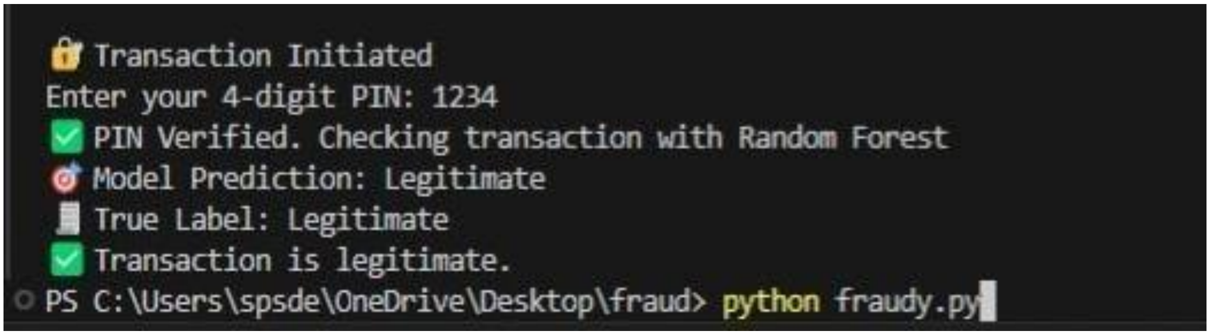


Fig. 5 ROC Curve Comparision



```
🔐 Transaction Initiated
Enter your 4-digit PIN: 1247
❌ Incorrect PIN. Attempt 1/2
Enter your 4-digit PIN: 9887
❌ Incorrect PIN. Attempt 2/2
🚨 Fraud Detected: Too many incorrect PIN attempts!

📧 Sending Email to amounik60@gmail.com:
Subject: Fraud Alert Notification
Body: Multiple incorrect PIN attempts detected on your account!

📱 Sending SMS to 7993426100:
Message: Multiple incorrect PIN attempts detected on your account!
```



```
🔐 Transaction Initiated
Enter your 4-digit PIN: 1234
✅ PIN Verified. Checking transaction with Random Forest
🎯 Model Prediction: Legitimate
🧾 True Label: Legitimate
✅ Transaction is legitimate.
PS C:\Users\spsde\OneDrive\Desktop\fraud> python fraudy.py
```

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

credit card fraud is a common problem that causes both individualizes and banks and credit card enterprises to lose popularity this design intends to assist consumers and banks in recovering their income by constructing a model that can more efficiently identify deceitful and non-deceitful deals using the time and quantum variables in the kaggle dataset we start by erecting the model with su pervised machine literacy styles like the arbitrary timber decision tree drag harborage vector machine and nave bayes in this study we presented a procedure to address this issue statement we used all four machine literacy classifiers and the results showed that the random forest classifier had the topmost delicacy perfection recall and f1- score as a result it is the topmost choice for detecting deceitful deals in the future we will concentrate on other datasets and machine literacy approaches to get a more effective result.

Random Forest likely achieves the highest accuracy due to its ensemble learning approach, reducing variance and improving generalization.

Decision Tree may perform well but is prone to overfitting, leading to lower generalization on unseen data.

SVM can work efficiently with small datasets and complex boundaries, but its performance depends heavily on kernel selection and hyperparameter tuning.

comparison of Random Forest, Decision Tree, and Support Vector Machine (SVM) for breast cancer classification highlights key insights into their performance. Random Forest tends to achieve the highest accuracy due to its ensemble learning approach, effectively reducing variance and improving generalization. Decision Tree, while intuitive and easy to interpret, is prone to overfitting, making it less reliable for unseen data without proper pruning techniques. SVM, on the other hand, is well-suited for small datasets and complex boundaries but requires careful kernel selection and hyperparameter tuning to achieve optimal results. Overall, Random Forest emerges as a strong contender for balancing accuracy and robustness, while Decision Tree provides clarity in decision-making, and SVM excels in handling well-separated classes. If you're considering applications in fraud detection systems, Random Forest might be particularly beneficial due to its ability to handle imbalanced data and complex relationships.

Credit card fraud detection is not just a technical challenge—it’s a frontline battle in safeguarding user trust, digital security, and the integrity of financial ecosystems. This project illustrates how the fusion of machine learning and real-world transaction data can illuminate the dark corners where fraud hides. By analyzing anonymized features, implementing intelligent classification models like Random Forests and SVMs, and addressing data imbalance with methods like SMOTE, we build a vigilant system capable of identifying the rare and dangerous. The success of such models lies not just in accuracy, but in their ability to adapt—learning from new threats, evolving with trends, and responding in real time.

However, the journey doesn’t stop here. As technology advances, so do the tactics of cybercriminals. Future enhancements could involve deep learning architectures such as autoencoders or LSTM networks for sequential pattern detection, real-time fraud prediction engines powered by cloud computing, and integration with blockchain for enhanced transparency. Continuous feedback loops from banking systems and user behavior analytics could further sharpen detection with every transaction.

In a world where a single fraudulent transaction can ripple into massive loss, our work becomes more than code—it becomes a guardian. With vigilance, innovation, and ethical AI, we step closer to a world where digital trust isn’t just a goal—it’s a guarantee.

Credit card fraud is a persistent and evolving threat in today’s digital economy, and the need for intelligent, data-driven defense mechanisms has never been greater. Through the implementation of machine learning techniques on real-world transaction data, this project has demonstrated how algorithms can effectively learn to distinguish fraudulent behavior from legitimate activity. By leveraging PCA-transformed features, handling data imbalance with resampling techniques, and evaluating models using precision-focused metrics, we build not just a system—but a safeguard. While no model is perfect, this approach offers a powerful step forward in proactive fraud detection. As fraudsters grow smarter, so must our models—continuously learning, adapting, and protecting the flow of trust in financial systems. In the end, it’s not just about catching fraud—it’s about staying one step ahead of it.

Future scope:

The future of credit card fraud detection is shaping up to be a dynamic fusion of cutting-edge technology, behavioral insight, and intelligent automation—almost like giving the financial system a sixth sense. As cyber threats become more advanced, fraud detection systems are evolving from rule-based, reactive models to real-time, adaptive guardians of digital finance. Artificial Intelligence and Machine Learning will sit at the heart of this evolution, powering systems that not only detect known fraud patterns but also predict and adapt to new, never-before-seen anomalies. Deep learning algorithms will analyze massive amounts of transaction data to detect subtle deviations in spending behavior, while reinforcement learning will enable systems to self-improve based on feedback. In parallel, behavioral biometrics will become a key layer of security—monitoring how users type, swipe, and even how they hesitate before entering their PINs, making identity verification a continuous and almost invisible process. Meanwhile, the rise of blockchain and decentralized identity systems promises to reduce central points of failure, with smart contracts capable of autonomously flagging or blocking suspicious activity. As quantum computing looms on the horizon, quantum-resistant algorithms will become necessary to secure encryption and communication in fraud detection systems. Additionally, we'll see a tighter integration with cybersecurity frameworks, where fraud detection tools draw real-time insights from phishing attempts, data breaches, and social engineering attacks to anticipate fraud before it happens. Big data will enable hyper-personalized fraud detection, tailoring risk scores to each individual user based on their unique digital footprint, while edge AI will bring fraud detection to the device level—operating in real time even without an internet connection. This means a future where your phone or card can detect fraud on the spot. There's even the potential for gamification, where users are incentivized to participate in fraud reporting, making it more engaging and community-driven. Altogether, the future of credit card fraud detection is not just about stopping bad transactions—it's about creating an intelligent, responsive, and user-centric ecosystem that evolves as fast as the threats do, making finance safer and smarter with every transaction. In the ever-accelerating rhythm of the digital world, where transactions zip across borders in milliseconds and your financial fingerprint is scattered across clouds, devices, and apps, the battle against credit card fraud is becoming more poetic, more precise, and profoundly powerful. The future scope of credit card fraud detection isn’t just an upgrade—it’s a full-blown evolution, transforming static firewalls into intelligent, adaptive guardians of trust. Powered by artificial intelligence, fraud detection will no longer wait for red flags to rise after the fact; instead, it will become proactive, predictive, and personal. Machine learning models will evolve into hyper-intelligent entities that learn from each click, swipe, and second of hesitation, identifying anomalies not just in numbers, but in behavior, mood, and even biometric rhythms. These systems will evolve to understand the heartbeat of a user’s habits—how you usually shop, where your phone rests, how fast you type your name—creating a digital twin that fraudsters will find nearly impossible to impersonate. This future also embraces fusion—not just of tech, but of disciplines. Cybersecurity, psychology, finance, and data science will converge to build fraud detection models that understand not just the *what*, but the *why* behind transactions. Emotion-aware AI might soon detect stress or hesitation in voice commands for voice-authenticated payments. Augmented reality might bring transaction visualizations right before your eyes, allowing you to approve or deny in a snap. Furthermore, big data will empower hyper-personalization, tailoring security protocols uniquely for each individual based on lifestyle, culture, and environment. A student in Chennai will not have the same risk profile as a digital nomad in Berlin, and fraud detection will respect that uniqueness.There’s even a cultural shift on the horizon—where fraud prevention is no longer just a backend process, but a shared responsibility. Apps might start rewarding users for participating in fraud simulations, reporting suspicious activity, or training AI models. Think of a world where fraud prevention is part of daily interaction, gamified and community-driven, transforming vigilance into engagement. Businesses, too, will undergo a shift—focusing not only on detection but on seamless recovery, rapid customer reassurance, and transparent communication in the face of breaches. In this grand future, credit card fraud detection becomes more than a safety net. It becomes an intelligent, intuitive part of our lives—a silent sentinel that walks with us through every transaction, learning, adapting, and shielding us from the invisible hands that try to reach into our digital wallets.

It’s a space of limitless innovation, where careers will flourish, research will surge, and the pursuit of secure finance will become as elegant as it is essential. This is not just a future worth imagining—it's one worth building.

The future of credit card fraud detection lies in the integration of smarter, faster, and more adaptive technologies. As fraudsters evolve, detection systems must transform into intelligent guardians capable of real-time decision-making. Advanced deep learning models like autoencoders and LSTMs can analyze complex patterns and transaction sequences with greater accuracy, while adaptive AI systems will continuously learn from new data without needing manual retraining. Real-time fraud detection using cloud and edge computing will become the norm, ensuring that suspicious transactions are flagged instantly. The adoption of blockchain can bring transparency and security to transaction records, while behavioral biometrics—such as typing rhythm, device movement, or location habits—can offer personalized protection. In addition, global collaboration between banks and payment platforms could create shared fraud intelligence networks, enhancing early detection across borders. Multimodal data fusion, combining transaction, network, and user behavior data, will further strengthen the system’s ability to catch even the most subtle forms of fraud. In essence, the future scope envisions fraud detection not just as a technical tool, but as a living, learning, and constantly evolving shield for our digital transactions.

REFERENCES

[1].Sarker, A.,Yasmin, M. , Rahman, M. , Rashid, M. and Roy, B. (2024) Credit Card Fraud Detection Using Machine Learning Techniques. Journal of Computer and Communications, 12, 1-11. doi: 10.4236/jcc.2024.126001.

[2].Paul, Annu, and Varghese Paul. "Poor and rich squirrel algorithm-based Deep Maxout network for credit card fraud detection." International Journal of Wireless and Mobile Computing 28, no. 2 (2025): 123-134.

[3].Paul A, Paul V. Poor and rich squirrel algorithm-based Deep Maxout network for credit card fraud detection. International Journal of Wireless and Mobile Computing. 2025;28(2):123-34.s

[4].Veeru, B., Devender, N. and Shoban, K., 2025. Cryptographic Security in Credit Card Fraud Detection Using Homomorphic Encryption with CRO Based Hybrid BL-GRU Classification. In Sustainable Development Using Private AI (pp. 85-108). CRC Press.

[5].Al-Hagery, Mohammed Abdullah, and Abdalla I. Abdalla Musa. "Automated Credit Card Risk Assessment using Fuzzy Parameterized Neutrosophic Hypersoft Expert Set." International Journal of Neutrosophic Science (IJNS) 25, no. 1 (2025).

[6].Sizan, Mir Mohtasam Hossain, Anchala Chouksey, Nikhil Rao Tannier, Md Abdullah Al Jobaer, Jasmin Akter, Ashutosh Roy, Mehedi Hasan Ridoy, M. Saif Sartaz, and Dewan Aminul Islam. "Advanced Machine Learning Approaches for Credit Card Fraud Detection in the USA: A Comprehensive Analysis." Journal of Ecohumanism 4, no. 2 (2025): 883-905.

[7].Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. Journal of Big Data, 12(1), 6.

[8]. Zojaji, Zahra, Reza Ebrahimi Atani, and Amir Hassan Monadjemi. "A survey of credit card fraud detection techniques: data and technique oriented perspective." arXiv preprint arXiv:1611.06439 (2016).

[9].Al-Hagery, M. A., & Abdalla Musa, A. I. (2025). Automated Credit Card Risk Assessment using Fuzzy Parameterized Neutrosophic Hypersoft Expert Set. International Journal of Neutrosophic Science (IJNS), 25(1).

[10].Zojaji, Zahra, Reza Ebrahimi Atani, and Amir Hassan Monadjemi. "A survey of credit card fraud detection techniques: data and technique oriented perspective." arXiv preprint arXiv:1611.06439 (2016).

[11]. Singh, Amit, Ranjeet Kumar Ranjan, and Abhishek Tiwari. "Credit card fraud detection under extreme imbalanced data: a comparative study of data-level algorithms." Journal of Experimental & Theoretical Artificial Intelligence 34, no. 4 (2022): 571-598.

[12].Marco, Robert, Nur Aini, and I. Agastya. "A Hybrid Approach CNN-LSTM Based on Attention Mechanism for Credit Card Fraud Detection." International Journal of Intelligent Engineering & Systems 18.3 (2025).Marco, Robert, Nur Aini, and I. Agastya. "A Hybrid Approach CNN-LSTM Based on Attention Mechanism for Credit Card Fraud Detection." International Journal of Intelligent Engineering & Systems 18.3 (2025).

[13].Bharath, Gurram, and P. S. Priyadarsini. "Comparative analysis of ResNet50 over random forest in predicting credit card fraud detection with improved accuracy." AIP Conference Proceedings. Vol. 3252. No. 1. AIP Publishing, 2025.

[14].Btoush, Eyad Abdel Latif Marazqah, Xujuan Zhou, Raj Gururajan, Ka Ching Chan, Rohan Genrich, and Prema Sankaran. "A systematic review of literature on credit card cyber fraud detection using machine and deep learning." PeerJ Computer Science 9 (2023): e1278.

[15].Ding, Y., Kang, W., Feng, J., Peng, B., & Yang, A. (2023). Credit card fraud detection based on improved Variational Autoencoder Generative Adversarial Network. IEEE Access, 11, 83680-83691.

[16]. Strelcenia, E., & Prakoonwit, S. (2023). Improving classification performance in credit card fraud detection by using new data augmentation. AI, 4(1).

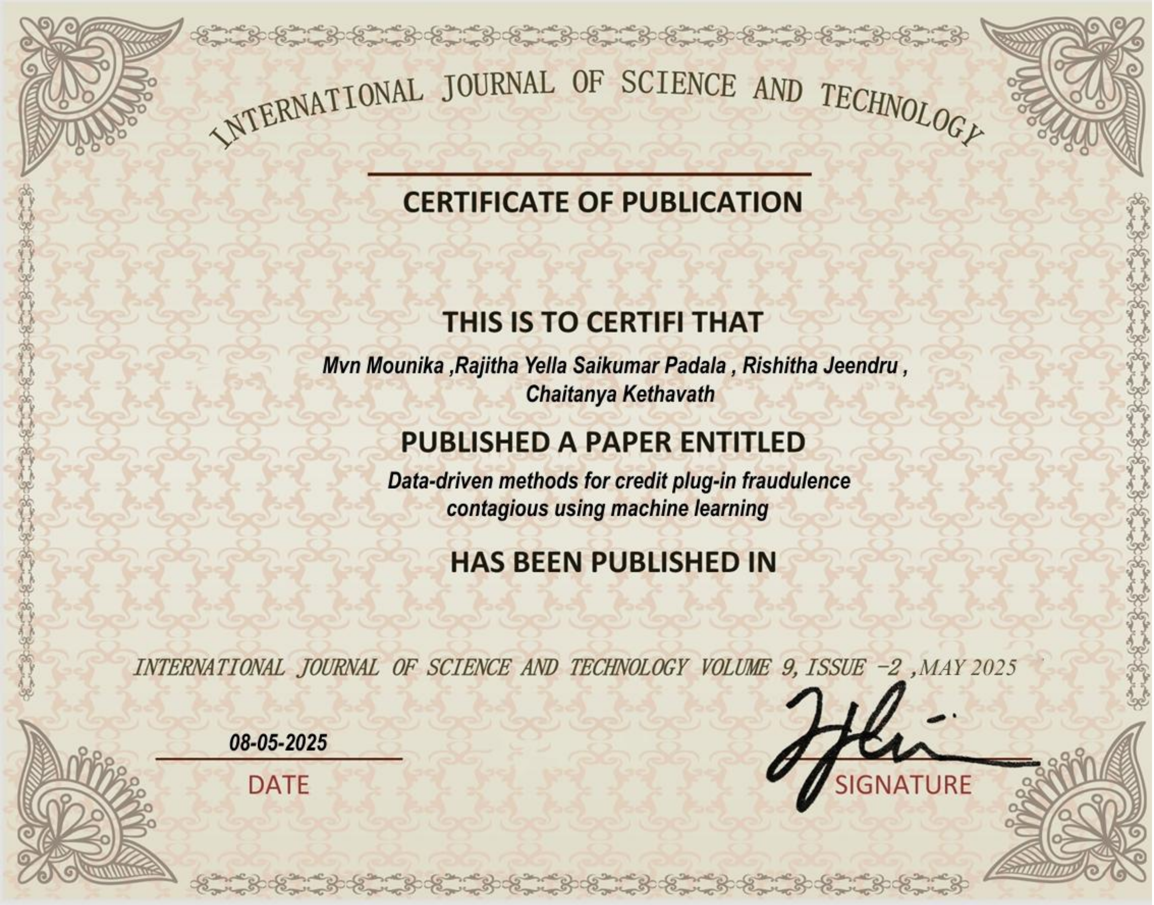
[17]. Rezapour, M. (2019). Anomaly detection using unsupervised methods: credit card fraud case study. International Journal of Advanced Computer Science and Applications, 10(11).

[18]. Mienye, Ibomoiye Domor, and Yanxia Sun. "A deep learning ensemble with data resampling for credit card fraud detection." Ieee Access 11 (2023): 30628-30638.

[19].Alamri, Maram, and Mourad Ykhlef. "Survey of credit card anomaly and fraud detection using sampling techniques." Electronics 11, no. 23 (2022): 4003.Alamri, Maram, and Mourad Ykhlef. "Survey of credit card anomaly and fraud detection using sampling techniques." Electronics 11, no. 23 (2022): 4003.

[20]. Mienye ID, Jere N. Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. IEEE Access. 2024 Jul 11.

PUBLICATION CERTIFICATE



INTERNATIONAL JOURNAL OF SCIENCE AND TECHNOLOGY

CERTIFICATE OF PUBLICATION

THIS IS TO CERTIFI THAT

Mvn Mounika ,Rajitha Yella Saikumar Padala , Rishitha Jeendru ,
Chaitanya Kethavath

PUBLISHED A PAPER ENTITLED

Data-driven methods for credit plug-in fraudulence
contagious using machine learning

HAS BEEN PUBLISHED IN

INTERNATIONAL JOURNAL OF SCIENCE AND TECHNOLOGY VOLUME 9, ISSUE -2 ,MAY 2025

08-05-2025
DATE

SIGNATURE