INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 4, April 2026 | ISSN: 2456-4184 | IJNRD.ORG

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING-BASED ANOMALY DETECTION

[1] **Kowshica.A.R,**
PG Student Department of CSE
Selvam college of Technology,Namakkal helenkowshica15@gmail.com

[2] **Tamilselvi.R,**
Head of the Department
Department of CSE
Selvam college of Technology, Namakkal.

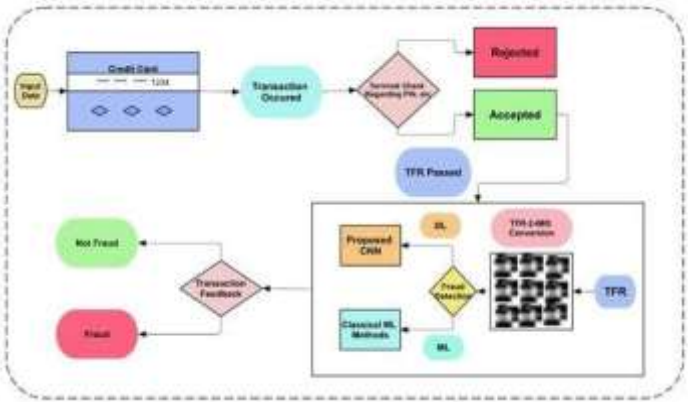
[3]**Ramya A.R,**
PG Student Department of Mechanical
Selvam College of Technology, Namakkal
ramyaar001@gmail.com

ABSTRACT
Credit card fraud poses a significant threat to financial institutions and customers as the volume of digital transactions continues to increase worldwide. Traditional fraud detection systems rely heavily on predefined rules, threshold checks, and manual verification,which often fail to detect new and evolving attack patterns. These systems also struggle with high false- positive rates and delayed detection, making them inefficient for modern real-time financial systems. To address these challenges, this project presents a machine learning-based anomaly detection framework capable of identifying fraudlent credit card transactions with improved accuracy and adaptability. The proposed system utilizes both supervised and unsupervised learning approaches to address the highly imbalanced nature of fraud datasets, where fraudulent transactions are extremely rare compared to legitimate ones. Key stages of the methodology include data preprocessing, missing-value handling, normalization, feature engineering, correlation analysis, and balancing techniques such as SMOTE to improve model learning. The system is further extended with a real-time fraud scoring mechanism capable of flagging suspicious transactions instantly. Overall, the project demonstrates that machine learning-based anomaly detection provides a robust and scalable solution for credit card fraud detection. It enhances security, supports faster decision-making, and offers high adaptability to evolving fraud behaviors. This work lays the foundation for deploying an intelligent fraud detection system in teal-time banking environments.

KEYWORDS: Credit Card Fraud Detection, Machine Learning, Anomaly Detection, Fraud Analytics, Transaction Monitoring, Predictive Modeling, Real-time Detection, Data Mining, Supervised Learning, Unsupervised Learning.

I.INTRODUCTION
In the modern digital era, credit cards have become one of the most commonly used financial instruments for making

online and offline purchases. Their convenience, speed, and global acceptance have made them an integral part of electronic commerce and personal banking. With the rise of internet-based financial systems, mobile banking, and cashless transactions, the volume of credit card transactions has increased dramatically across the world. As a result,banks and financial institutions every hour.However, this tremendous growth in digital payment infrastructure has attracted cybercriminals who continuously devise new techniques to exploit vulnerabilities in payment systems. Fraudsters use methods such as indetity theft, phishing, card skimming, account takeovers, fake merchant accounts, and automated bots to commit credit card fraud. These fraudulent activities result in huge financial losses for banks, merchants, and cardholders. According to global banking reports, credit card fraud losses amount to billions of dollars annually, and the number continues to rise as fraudsters develop more advanced techniques. Given this threat, credit card fraud detection has become a critical research area in the fields of financial technology (Fin Tech), cybersecurity, and machine learning. Traditional fraud detection systems based on manually defined rules are no longer sufficient against evolving fraud behaviors. Hence, artificial intelligence and machine learning offer an intelligent,dynamic, and scalable solution to detect anomalies in transaction patterns.



II. RELATED WORK

A literature survey provides an in-depth exploration of previous research, existing systems, methodologies, tools,and techniques used in credit card fraud detection. The purpose of this chapter is to understand how fraud detection methods have evolved over time and identify their limitations, strengths, and applicability to modern fraud scenarios. Reviewing earlier work also helps determine the gaps that this projects aims to address through machine learning-based anomaly detection.The research on fraud detection spans multiple domains including statistical modeling, artificial intelligence, pattern recognition, machine learning, and data mining. As credit card transactions have grown more complex, fradusters have adapted their techniques correspondingly. Therefore, researchers have continuously developed and improved computational models to detect fraudulent behavior effectively. This chapter provides a detailed examination of traditional rule-based solutions, classical machine learning techniques, anomaly detection methods, and hybrid approaches, laying the foundation for choosing the appropriate models. Existing fraud detection systems can be categorized into traditional techniques and modern intelligent techniques. Rule-based systems were among the earliest fraud detection mechanisms. They operate on manually defined rules such as:

- "block transactions above a certain amount"
- "Flag transactions from different countries within a short duration"
- "Trigger alert when high-value purchase is made suddenly" • "Allow transactions only from registered IP addresses"

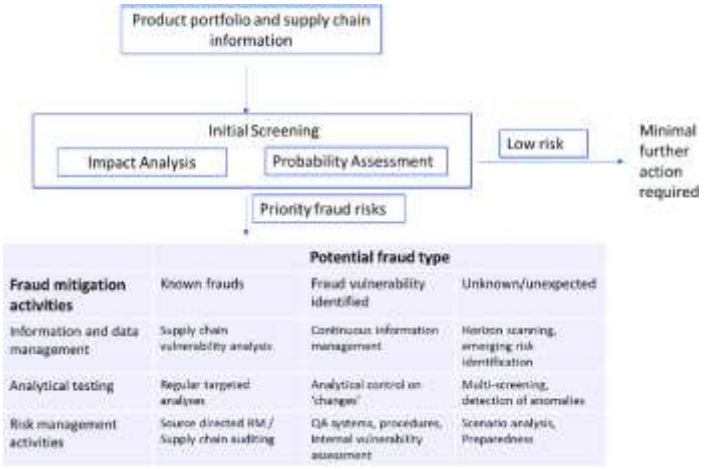
Advantages: • Easy to implement and understand • Suitable for simple fraud patterns • Low computational cost **Disadvantages:** • Inability to detect new/unseen fraud patterns • Requires frequent manual updates • High false positives • Poor scalability for large dataset
Manual fraud detection systems In traditional systems, fraud analysts manually review suspicious transactions.
Drawbacks: • Extremely time-consuming • Cannot handle millions of transactions • Analysts can overlook complex fraud patterns • Expensive and inefficient
Statistical models Earlier research also utilized statistical methods such as: • Logistic Regression • Bayesian Classification • Linear Discriminant Analysis These models assume that fraudulent behavior follows specific statistical distributions. However, fraud patterns constantly evolve and do not always adhere to fixed probability distributions. **Limitations:** • Sensitive to outliers • Poor at detecting nonlinear relationships • Require clean and balanced data Because of these limitations, researchers started exploring machine learning approaches.
Evolution of machine learning in fraud detection With the growth of digital payment systems, researchers began applying machine learning techniques to analyze complex transaction patterns. Supervised machine learning approaches Supervised models resquire labeled datasets containing both legitimate and fraudulent transactions Common supervised algorithms include: • Logistic

Regression • Random Forest • Decision Trees • Gradient Boosting Machines • Support Vector Machines(SVM) • Neural Networks

Advantages: • High accuracy when trained with good data

• Can classify transactions effectively • Learn nonlinear patterns

Disadvantages: • Require large labeled fraud datasets • Cannot detect new types of fraud easily • Highly imbalanced datasets reduce performance



Unsupervised anomaly detection methods: These methods identify anomalies without labeled data. Anomalies are rare patterns that differ significantly from normal user behavior. Popular algorithms include: • Isolation Forest • One-Class SVM • Autoencoders • Local Outlier Factor(LOF) **Advantage**: • No need for labeled data • Good at detecting unseen fraud patterns • Automatically adapts to unsual transactions

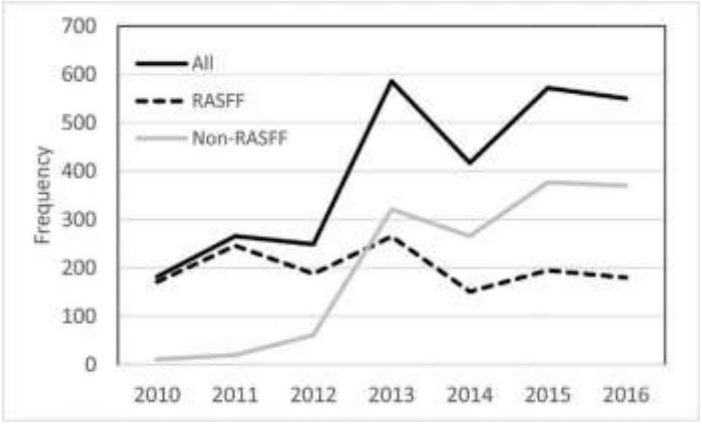
Disadvantage: • Sensitive to dataset noise • Accuracy depends on feature quality

Deep learning in fraud detection: With increasing computational capabilities, deep learning-based models are gaining popularity.

Recurrent neural networks(RNNS): RNNs model sequential transaction behavior of users. They detect: • Sudden spikes • Rapid unusual purchases • Time-based anomalies

Literature review of previous research studies: Below are some key research contributions in credit card fraud detection. Study 1:Dal Pozzolo et at.(2015) • Used European card transaction dataset. • Identified severe class imbalance problem. • Found Random Forest + undersampling gave good results. Study 2: J.West &M.Bhattacharya (2016) • Compared neural networks and decision trees. • Neural networks performed better on complex fraud patterns.

Research gap analysis: Based on the review of earlier systems, the following gaps are identified: Gap 1: High Data Imbalance Most datasets contain extremely few fraud cases. Gap 2: Evolving Fraud Patterns Fraudsters adapt quickly; rule-base systems cannot catch new patterns. Gap 3: Limited Real-Time capabilities Many models work offline but fail in real-time scoring. Gap 4: Lack of Hybrid Approaches Most studies focus on either supervised or unsupervised but not both.



III.SYSTEM DESIGN

Existing system Traditional fraud detection systems mainly rely on rule-based logic and manual inspection. For example, a transaction may be flagged if its amount exceeds a predefined threshold, originates from an unusual location, or is performed at an unexpected time. These rule-based systems are designed based on past fraud patterns and risk factors defined by fraud analysts. 3.1.1 Disadvantages ● Inability to Detect New Patterns – Fraudsters frequently change their techniques, making fixed rules ineffective. ● High False Positives – Legitimate transactions are often marked suspicious, disrupting customer experience. ● Manual Review Burden – Analysts must verify

flagged cases, which is inefficient for large-volume transaction streams.

● Lack of Real-Time Response – Many systems operate in batch mode, delaying fraud detection. ● Scalability Issues – with increasing transactions, traditional systems struggle to maintain accuracy and speed. Because of these drawbacks, existing systems are insufficient for handling modern fraud complexity, which motivates a more intelligent and adaptive solution.

SYSTEM REQUIREMETS:

4.1Hardware requirements

- Processor - Intel Core i5
- Speed - 2.3 GHz
- RAM - 8 GB
- Hard Disk - 500 GB
- Key Board - Standard Windows Keyboard
- Mouse - Three Button Mouse
- Monitor - 24‖ LED 4.2

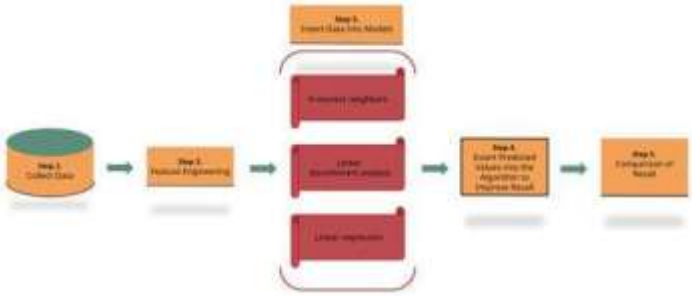
Software requirements

- Operating System : Windows 10
- OR Linux distributions (Ubuntu recommended)
- OR macOS(optional)
- Programming Language : Python

LANGUAGE SPECIFICATION

Python is a high-level, general purpose programming

language created by Guido van Rossun and first released in 1991. It emphasizes code readability, minimal syntax, and developer productivity. Python supports multiple programming paradigms including procedural, object- oriented, and functional programming, making it highly suitable for modern machine learning applications. It is widely used in industries such as: • Banking and Finance • Artificial Intelligence • Data analytics • Web development • Cybersecurity • Automation In the context of fraud detection, Python's capabilities make it an excellent choice for extracting make it an excellent choice for extracting meaningful patterns from large volumesof transaction data.

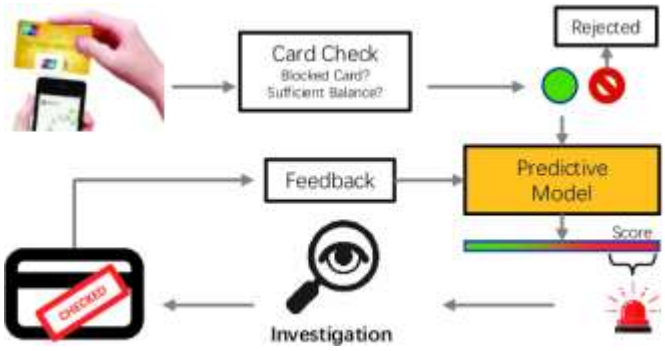


Key features of python Python : provides several features that make it particularly suitable for machine learning and data analysis projects. Some of the important features include: Python provides you with a certain level of freedom when composing a program, but there are some rules which must always be obeyed. One of these rules, which some people find very surprising, is that python uses indentation (that is, the amount of white space before the statement itself) to indicate the presence of loops, instead of using delimiters like curly braces ({}) or keywords (like

—begin‖ and —end‖) as in many other languages. The amount of indentation you use is not important, but it must be consistent within a given depth of a loop, and statements which are not indented must begin in the first column. Most python programmers prefer to use an editor like emacs, which automatically provides consistent indentation; you will probably find it easier to maintain your programs if you use consistent indentation in every loop, at all depths, and an intelligent editor is very useful in achieving this.

Simple and readable syntax: Python's syntax is clean and easy to understand, which accelerates development and reduces errors. This is crucial when dealing with large datasets and complex algorithms.

Extensive library support: Python has one of the richest ecosystems of libraries for data processing, visualization, machine learning, and deep learning. Popular libraries such as NumPy, Pandas, Matplotlib, and Scikit-Learn significantly simplify model implementation.



Object oriented: Python supports OOP concepts such as classes and objects, while also allowing functional programming styles. This flexibility enhances code reusability and organization. **Large community and documentation:** Python has one of the largest developer communities in the world. This ensures abundant tutorials, forums, libraries, and debugging support for machine learning projects. **Integration with AI / ML tools** Python integrates easily with machine learning toolkits such as: • TensorFlow • Scikit-learn • PyTorch • Keras • This makes Python the ideal choice for ML-based fraud detection systems**. Why python was chosen for this project** Python was selected as the primary development language for this credit card fraud detection project due to the following reasons**: Ease of Implementation** Python allows quick prototyping of ML models with fewer lines of code**. Rich Data Science Libraries** Libraries like Pandas and NumPy simplify data cleaning and preprocessing. **Strong ML Frameworks** Scikit-Learn provides ready-to-use funtions for classificationsanomaly detection, and model evaluation. 17 **Excellent Visualization Support** Libraries like Matplotlib and Seaborn help in generating EDA visualizations essential for Phase -1. **Efficiency with Large Datasets** Python handles large datasets well when optimized with vectorized operations. **Community Support** Any error or issue can be quickly addressed due to Python's massive online community. For these reasons, Python stands out as the most preferred language for machine learning research projects. **Python libraries used in the project** Python's power lies in its libraries. The following libraries were used in Phase-1 of this project: **Numpy (numerical python)** NumPy is a foundational package for numerical computation in Python. It supports: • Multi-dimensional arrays • Vectorized operations • Mathematical and statistical functions Usage in this project: • Handling numeric transaction data • Performing mathematical operations • Implementing scaling and normalization

Pandas Pandas is a data manipulation and analysis library widely used in data science. Key features: • DataFrames for structured datasets • Handling missing values • Grouping and sorting operations Usage in this project: • Loading the credit card transaction dataset • Cleaning and preprocessing

• Analyzing fraud vs.non-fraud attributes Critical bugs sometimes do not get fixed for long periods of time. An example is a bug with status critical existing since 2003. **Matplotlib / seaborn** The libraries are widely used for generating plots during Exploratory Data Analusis (EDA). Usage in this project: • Visualizing amount distribution • Time distribution patterns • Correlation heatmaps • Fraud vs.genuine comparison

Jupyter notebook Jupyter Notebook is an interactive development environment ideal for running machine learning experiments.

Python execution environment During the development, Python code was executed using: Jupyter notebook (for EDA and dataset analysis) Vs code (for python scripts) Anaconda Distribution (for environment management) 5.7. **Role of python in credit card fraud detection** Python played a critical role in executing the following tasks: **Handling Imbalanced** Datasets Using P+ scikit-learn preprocessing tools. **Identifying Baseline ML approaches** Python enabled rapid testing and comparison of: • Logistic Regression • Random Forest • Isolation Forest • One – class svm

IV EXISTING SYSTEM

Traditional fraud detection systems mainly rely on rule- based logic and manual inspection. For example, a transaction may be flagged if its amount exceeds a predefined threshold, originates from an unusual location, or is performed at an unexpected time. These rule-based systems are designed based on past fraud patterns and risk factors defined by fraud analysts.

DISADVANTAGES

- Inability to Detect New Patterns – Fraudsters frequently change their techniques, making fixed rules ineffective.
- High False Positives – Legitimate transactions are often marked suspicious, disrupting customer experience.

- Manual Review Burden – Analysts must verify flagged cases, which is inefficient for large-volume transaction streams.
- Lack of Real-Time Response – Many systems operate in batch mode, delaying fraud detection.
- Scalability Issues – with increasing transactions, traditional systems struggle to maintain accuracy and speed.

Because of these drawbacks, existing systems are insufficient for handling modern fraud complexity, which motivates a more intelligent and adaptive solution.

V. METHODOLOGY

We propose credit card fraud detection system using machine learning-based anomaly detection, this chapterpresents the step-by step methodology followed for the project, including dataset collection, preprocessing, exploratory data analysis (EDA), feature engineering, and baseline model selection strategy.**Dataset collection:** The dataset used for this project is the kaggle credit card fraud detection dataset. Which contains anonymized credit card transactions made by European customers. The dataset is ideal for anomaly detection becaused it represents a real- world imbalanced dataset where fraudulent cases are extremely rare. **Data preprocessing methodology:** Data preprocessing ensures that the
raw dataset is transformed into a meaningful and usable format for analysis. The following preprocessing tasks were completed: • Handling Missing Values • Duplicate and Noise Removal • Feature Scaling • Time Feature Ana

VI. CONCLUSION

The primary objective of this project was to study and analyze the feasibility of developing a machine learning- based anomaly detection system for identifying fraudulent credit card transactions. Throughout this phase, significant groundwork was completed in understanding the nature of fraud detection, examining existing systems, analyzing the dataset, identifying key features ,and outlining the appropriate machine learning techniques suitable for the project. The study revealed that traditional fraud detection approaches, especially rule-based and manual inspection systems, are insufficient for handling modern fraud patterns due to their inability to adapt to evolving fraud strategies. The literature survey and system analysis clearly demonstrated the importance of machine learning in detecting anomalies within massive transactional datasets. Machine learning-based detection offers improved adaptability,scalability ,and accuracy in recognizing unusual transaction behavior, making it an ideal approach for combating financial fraud. The outcomes from this phase ensure that the project is well-prepared for the next stage, where actual model training, evaluation, and tuning will be performed.

VII. REFERENCE

[1] Bhattacharya,S., Jha,S., Data Mining Approaches for Banking Fraud Detection ,2014.
[2] Dal Pozzolo,A., Adaptive Machine Learning Techniques for Credit Card Fraud Detection, 2015.
[3] West,J., Bhattacharya,M., Intelligent Fraud Detection Techniques for Financial,2016
[4] Carcilo,F., Real-Time Credit Card Fraud Detection Using Machine Learning, 2018.
[5] Patidar, R.,Sharma,L., Credit Card Fraud Detection Using Neural Networks, 2011.
[6] Sebastio,H., Pires, F., Classifier Ensemble Approaches for Credit Card Fraud Detection, 2017.
[7] Abdelrahim, A., Machine Learning Methods for Imbalanced Fraud Detection Datasets, 2019.
[8] Ahmed, M., Mahmood, A., Unsupervised Anomaly Detection for Fraudulent Transactions, 2016.
[9] Whitrow c., Hand, D., Transaction Aggregation for Fraud Detection,2009.
[10] Randhawa, k. Credit Card Fraud Detection Using Majority voting Ensembles, 2018.