IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 11 ISSUE : 02 Print / Issue Publication Date: June 2026

ISSN : 2455-2143

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2026
Vol. 11, Issue 02, ISSN No. 2455-2143, Pages 75-80
Published Online June 2026 in IJEAST (http://www.ijeast.com)

CREDITCARD FRAUD DETECTION USING MACHINE LEARNING

Sahana Vishwakarma, Khushi Nirale, Sheetal Parshetty, Aishwarya Reddy
Department of Computer Science and Engineering
Lingaraj Appa Engineering College Bidar, Karnataka, India

Dr. Vivek Jalade
Professor & Head of Department of Computer Science and Engineering
Linagaraj Appa Engineering College Bidar, Karnataka, India

Abstract— **Credit card fraud has become one of the major problems in the digital payment system due to the rapid increase in online transactions and electronic banking services. Fraudulent activities not only cause financial losses to banks and customers but also reduce trust in online payment systems. Traditional fraud detection methods are often unable to identify complex and evolving fraud patterns accurately. Therefore, there is a need for an intelligent and automated fraud detection system that can efficiently detect fraudulent transactions in real time.**

This project presents a Credit Card Fraud Detection System using Machine Learning techniques to identify fraudulent and genuine transactions effectively. The proposed system uses various machine learning algorithms such as Logistic Regression, Decision Tree, and XG Boost Classifier is applied to balance the dataset and improve prediction accuracy. The system also performs data preprocessing, feature extraction, model training, evaluation, and real-time transaction analysis

I. INTRODUCTION

This chapter of the project report isthe beginning of the content of this report. It contains the building up of the plot ofthisreport. The problem statement along with the main objectives of this project are discussed here. The significance of this project and the real motivation behind the intentions to take up this topic as our project are also listed in detail in this particular chapter.“Fraud” in credit card transaction is unauthorized and unwanted usage of an account by someone other than the owner of the account. Necessary prevention measures can be taken to stop this abuse and the behaviour of such fraudulent practices can be defined as a case where a person uses someone else’s credit card for personal reasons

while the owner and the card issuing authorities are unaware of the fact that the card is being used. In today’s era, with the widespread use of credit cards for online transactions, the risk of fraudulent activities has increased significantly. Addressing this challenge demands sophisticated methods that can swiftly and accurately detect fraudulent transactions to safeguard financial assets and uphold customer trust. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behavior, which consist of fraud, intrusion, and defaulting. Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent.

Some of the currently used approaches to detection of such fraud are:

- Artificial Neural Network (ANN)
- Fuzzy Logic
- Genetic Algorithm
- Logistic Regression
- Decision Tree
- Support Vector Machines (SVM
- Bayesian Networks
- Hidden Markov Model (HMM)
- K-Nearest Neighbour

1.1 BACKGROUND SURVEY

With the rapid growth of e-commerce, online banking, and digital payment systems, credit card usage has increased tremendously worldwide. Along with this growth, credit card fraud has also become a major challenge for financial institutions and customers. Fraudsters use advanced techniques to steal card information and perform unauthorized transactions, causing severe financial losses and reducing customer trust in online payment systems.

Earlier fraud detection systems mainly used manual verification and rule-based techniques. These traditional systems were unable to detect complex fraud patterns and produced a high number of false alerts. To overcome these limitations, researchers introduced Machine Learning techniques for fraud detection. Machine Learning algorithms XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE can automatically analyze transaction data, identify

unusual behavior, and detect fraudulent transactions with better accuracy.
Several research studies have been conducted in the field of credit card fraud detection. Researchers have used algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, Neural Networks, and XG Boost for fraud classification. Many studies also focused on handling imbalanced datasets using techniques like SMOTE, GANs, and oversampling methods to improve fraud detection performance.

1.2 PROBLEM STATEMENT

The rapid increase in online transactions and digital payment systems has led to a significant rise in credit card fraud activities. Fraudulent transactions cause major financial losses to banks, businesses, and customers. Existing traditional fraud detection systems are not efficient in detecting complex and continuously changing fraud patterns. These systems often generate high false positives and fail to identify fraudulent transactions accurately.
Another major challenge in fraud detection is the highly imbalanced nature of transaction datasets, where genuine transactions are much higher compared to fraudulent transactions. This imbalance reduces the performance and accuracy of machine learning models. Therefore, there is a need for an intelligent, accurate, and automated fraud detection system that can efficiently identify fraudulent credit card transactions in real time while minimizing false alerts. The proposed project aims to solve these problems using machine learning algorithms and data balancing techniques such as SMOTE to improve fraud detection accuracy and system performance

1.3 AIM AND OBJECTIVES

The main aim of this project is to develop an efficient Credit Card Fraud Detection System using Machine Learning techniques that can accurately identify fraudulent and genuine transactions, reduce financial losses, and improve security in online payment systems.

There are some proposed methods to develop a mechanismto determine that the upcoming transaction is fraud or not. The fraud transaction will be recognized with the help of location where the transaction took place, Frequency the interval of the time between two transactions, Amount what was the amount that was withdrawn from the transaction. And the comparison of different Machine Learning algorithms will be shown. The figure below shows the overall system framework.

The main objectives which we try to aim during the completion of this project are all listed below –
Get Credential Information.

- To balance the dataset which is unbalanced using SMOTE technique.
- To create a machine learning model using Logistic Regression, XG Boost, Decision Tree.
- Faster detection and higher accuracy

II. LITRATURE SURVEY

The aim of the Credit Card Fraud Detection Machine Learning (ML) project is to develop a reliable system that can detect credit card fraud. He acknowledged that financial fraud linked to electronic payments and e-commerce platforms is increasing and emphasized the need for effective detection methods. The limitations of existing security methods, such as tokenization and encryption, require the use of machine learning (ML) methods because these methods often fail to protect new information from fraud.
Overall, this paper presents research on various machine learning, challenges and new techniques to improve credit card fraud, detection systems. The plan will involve a group of cardholders, train different employees and using strategies to learn more about fraud. These studies aim to analyze the customer's details through the transaction, extract behavioral patterns in the cardholder group according to transaction costs, and then introduce different people to this group.

Credit card fraud detection using machine learning techniques A comparative analysis (2017): This article focuses on the challenges of credit card fraud, highlighting the vulnerability of credit card fraud as well as the everchanging nature of fraudulent behavior and fraud-related data. financial information fraud. It investigates the performance of three machine learning classifiers (Naïve Bayes, K-Nearest Neighbors (KNN), and logistic regression) on credit card fraud profiles obtained from residents of Europe(with284,807transactions).The results show the best accuracy achieved by Naive Bayes (97.92%), KNN (97.69%) and logistic regression (54.8%) classifiers. Comparative analysis shows that the K-nearest neighbor method outperforms Naive Bayes and logistic regression methods in terms of accuracy in credit card transactions.

Credit Card Fraud Detection Using Machine Learning (2022): This article addresses the problem of credit card fraud that has arisen due to the increasing use of credit cards around the world. The authors cite statistics from 2019 and 2020 that show an increase in credit card fraud due to the creation of new illegal accounts or unauthorized use of existing accounts. This warning led the authors to consider an analysis to address the problem, specifically using various machine learning (ML) methods to detect fraud in many credit card transactions. Overall, this article focuses on the use of machine learning techniques to solve the

International Journal of Engineering Applied Sciences and Technology, 2026
Vol. 11, Issue 02, ISSN No. 2455-2143, Pages 75-80
Published Online June 2026 in IJEAST (http://www.ijeast.com)

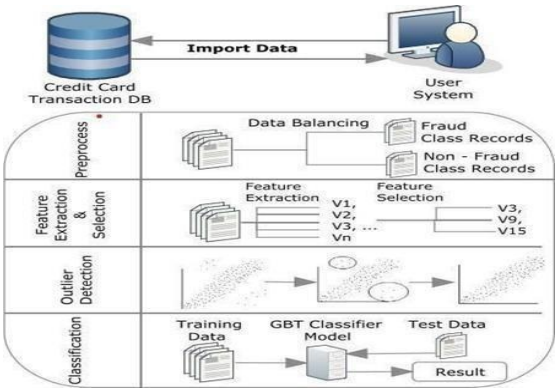
growing problem of credit card fraud to determine the most appropriate and effective methods for detecting fraud based on comparisons and insights from previous research.

Credit Card Fraud Detection using Machine Learning and Data Science (2017): This article centers around the utilization of information science and AI procedures to dissect charge cards. It accentuates the significance of recognizing fake strategic policies to forestall unlawful charges against purchasers. The objective is to foster a model that precisely distinguishes deceitful exchanges while limiting misclassification. The examination included investigating and focusing on informational indexes utilizing fair-minded search strategies, for example, residential areas backwoods prohibition from PCA exchange charge card exchanges.

III. SYSTEM DESIGN AND METHODILOGY

The proposed Credit Card Fraud Detection System is designed to identify fraudulent and genuine credit card transactions using machine learning algorithms. The system performs data collection, preprocessing, feature engineering, model training, evaluation, and real-time fraud prediction.
Initially, transaction data is collected from the dataset containing both fraudulent and non-fraudulent transactions. The collected data is then preprocessed by removing missing values, handling inconsistencies, and normalizing the dataset. Since the dataset is highly imbalanced, the SMOTE (Synthetic Minority Over-Sampling Technique) method is applied to balance fraudulent and genuine transactions.
After preprocessing, feature engineering and feature selection techniques are applied to extract important transaction attributes such as transaction amount, location, transaction time, and customer behavior patterns. These features help the machine learning models identify unusual transaction activities.

3.1 SYSTEM ARCHITECHTURE



The above figure shows the process of CCFDS. This system model accepts a real time customer credit card transaction database. It is more important to find the fraud rate of credit cards.

DFD (Data Flow Diagram)
The DFD used as communication system an user. It issimple representation of the complete project process. Transaction detection activity follows three phrases.

- Data Exploration
- Data Preprocessing
- Data Classifications

DATA PREPARATION
Below figures show the structure of the dataset where all attributes are shown, with their type, in addition to glimpse of the variables within each attribute, as shown at the end of the figure the class type is integer which needed to change to factor and identify the 0 as not- fraud to ease the process of creating the model and obtain visualizations.

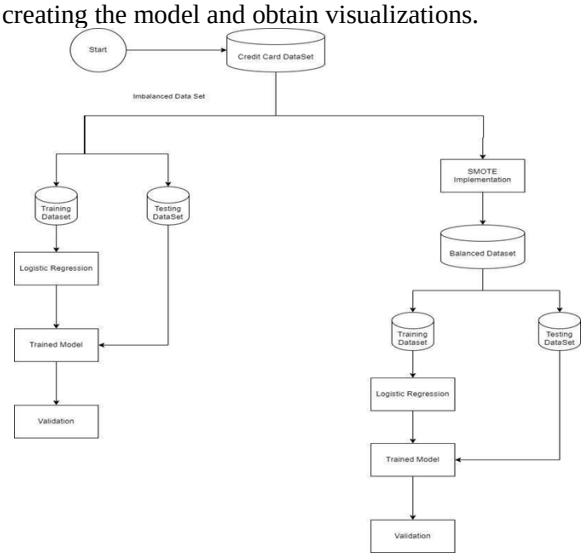


FIG 1-System Architecture and Overall System Flow

3.2 DATA FLOW AND SEQUENCE

The data flow of the Credit Card Fraud Detection System explains how transaction data moves through different stages of the system.

1. The user performs a credit card transaction.
2. Transaction details such as amount, location, time, and customer information are collected by the system.
3. The collected transaction data is sent to the preprocessing module for cleaning and normalization.
4. The SMOTE technique is applied to balance the dataset and improve fraud prediction accuracy.
5. Feature extraction and feature engineering are performed to identify important transaction characteristics.

6. The processed data is passed to machine learning models such as Logistic Regression, Decision Tree, and XG Boost Classifier.
7. The machine learning models analyze the transaction data and classify the transaction as genuine or fraudulent.
8. The prediction results are evaluated using performance metrics like Accuracy, Precision, Recall, F1-Score, and ROC-AUC.
9. If the transaction is identified as fraudulent, the system generates an alert and blocks or flags the transaction.
10. The results are stored in the database for future analysis and model retraining.

3.3 USE CASES

1. **User Transaction Processing**: The customer performs a credit card transaction through an online or offline payment system. The system receives transaction details for verification.
2. **Transaction Data Collection**: The fraud detection system collects transaction information such as amount, location, transaction time, and customer details for analysis.
3. **Fraud Detection:** The machine learning models analyze the transaction data and identify whether the transaction is genuine or fraudulent.
4. **Alert Generation:** If a fraudulent transaction is detected, the system generates alerts and informs the bank or customer immediately.
5. **Data Balancing and Model Training:** The administrator balances the dataset using SMOTE and trains machine learning models to improve fraud detection performance.
6. **Performance Evaluation:** The system evaluates the machine learning models using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics.
7. **Real-Time Monitoring:** The system continuously monitors transaction activities to detect suspicious behavior and new fraud patterns in real time.
8. **Report Generation:** The system generates reports containing fraud analysis results, model performance, and transaction statistics for stakeholders and financial institutions.

IV. IMPLIMENTATION

Tools and Technologies used:

- Matplotlib
- Google Colab
- Scikit
- Pandas

Algorithm Used:

Logistic Regression: Logistic regression is a simple and widely used technique in binary distribution problems. Unlike linear regression, which predicts a continuous outcome, logistic regression is suitable for situations where the variable is categorical and has two groups. The algorithm works by predicting the probability that a given entry falls into a particular category. It models the relationship between independent variables and the probability of a particular event occurring. The logistic regression model calculates the log difference of the probability and then converts it to the probability using the logistic function (also known as the sigmoid function). This function produces output 0 and 1, which is a list of different numbers for efficiency. Logistic regression can make it suitable for many fields such as finance, healthcare and business because it can control the relationship between different input and output distributions. Logistic Function:

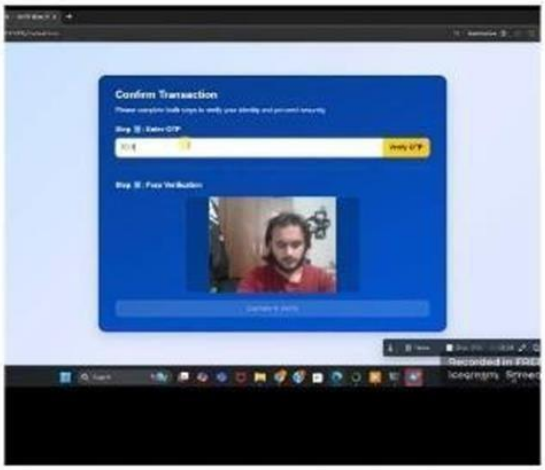
XG BOOST CLASSIFIER: XG Boost classifier is a robust machine learning algorithm that can help you understand your data and make better decisions. XGBoost is an implementation of gradient-boosting decision trees. It has been used by data scientists and researchers worldwide to optimize their machine learning models. XGBoost stands for “Extreme Gradient Boosting” and is has become one of the most popular and widely used machine learning algorithm due to its ability to handle large used machine learning algorithms due to handle large datasets and its ability to achieve state-of-the-art performance in many machine learning tasks such as classification and regression. One of the key feature of XG Boost is its efficient handling of missing values, which allows it to handle real-world data with missing values without requiring significant pre-processing. XGBoost has built-in-support for parallel processing, making it possible to train models on large datasets in a reasonable amount of time. To understand XGBoost classifier we first need to understand the following things:

V. PROJECT RESULT

The project is provided using various machine learning models for analysis to verify credit card transactions. It trains various classifiers such as logistic regression, decision tree, XG Boost and compares their accuracy. After running the code, it will display the accuracy scores and ROC curve of different learning models used in the fraud detection task, along with a comparison table showing their performance. The “Credit Card Fraud Detection Using Machine Learning” project was successfully implemented and tested using different machine learning algorithms such as Logistic Regression, Decision Tree, and XG Boost Classifier. The system was trained using credit card transaction datasets containing both fraudulent and non-fraudulent transactions. The dataset was highly imbalanced, so the SMOTE

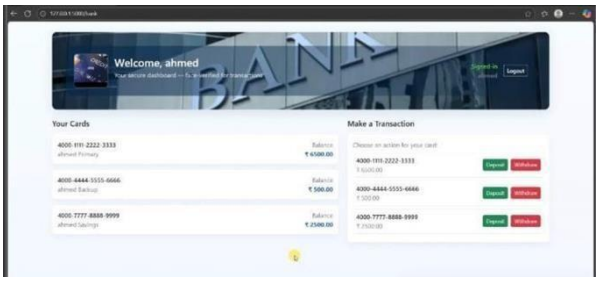
International Journal of Engineering Applied Sciences and Technology, 2026
Vol. 11, Issue 02, ISSN No. 2455-2143, Pages 75-80
Published Online June 2026 in IJEAST (http://www.ijeast.com)

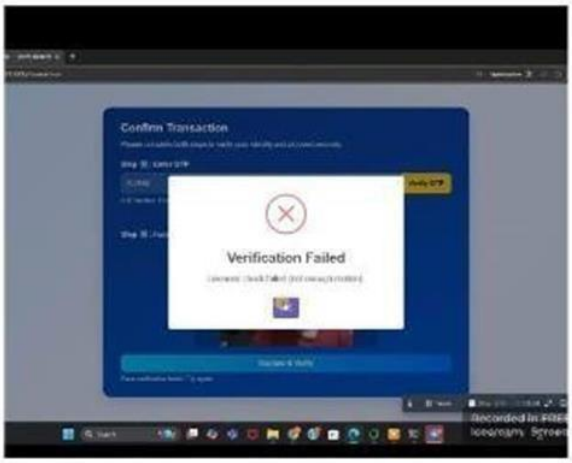
(Synthetic Minority Over-Sampling Technique) method was applied to balance the dataset and improve prediction performance.



After preprocessing and training, the machine learning models were evaluated using various performance metrics such as Accuracy, Precision, Recall, F1-Score, Sensitivity, Specificity, ROC Curve, and ROC-AUC score. The experimental results showed that the machine learning models were able to detect fraudulent transactions efficiently and reduce false transaction predictions.

Among all the algorithms used in the project, the XG Boost Classifier produced the best performance with the highest accuracy and ROC score. The XG Boost model achieved a ROC score of 1.0, indicating excellent fraud detection capability and accurate classification of fraudulent transactions. After balancing the dataset using SMOTE, the ROC score remained very high at 0.9999, which proves that the model performs efficiently even on balanced datasets.





VI. CONCLUSION

The conclusion is that the XG Boost classifier performs best in terms of accuracy and other parameters including ROC score on credit card identification test text.

XG Boost has the highest ROC score of 1.0, indicating that it accurately identifies the majority of fraudulent transactions while maintaining low cost.

After balancing the dataset the ROC score of 0.9999, which shows that the model perform quite perfect as compared to the other classifiers.

After the XG Boost, Decision Tree Classifier work will on balanced dataset with a roc score of 0.998. If the priority is to reduce negativity (misclassification is not fraud), a more accurate model such as XG Boost and Decision Tree will be preferred.

Consequently, considering the balance of accuracy and other parameters, it is recommended to use the XG Boost classifier as it performs best in testing card detection withdrawal patterns. In summary, although XG Boost showed the best performance in the test model, the most suitable model should be selected according to the specific needs, calculation needs and multi- purpose credit card fraud. Additional fine-tuning and rigorous testing is recommended before deploying the prototype in a production environment.

VII. REFERENCES

[1]. Altman, E.I., Marco, G., & Varetto, F.Corporate distress diagnosis comparisons using linear discriminant analysis and neural networks. Journal of Banking and Finance, 18(3), 505–529, 1994.

[2]. Chen J. Development and Application of Intelligent Transaction Scoring Model of the fraud risk. Credit card in China. 2006.

[3]. Dorronsoro, Ginel, Sgnchez and Cruz. Neural fraud detection in credit card operations. Neural Networks, IEEE Transactions. Volume: 8, Issue: 4: 827-834, 1997.

[4]. Flitman A.M. Towards analysing student failures: neural networks compared with regression analysis and multiple discriminant analysis. Computers & Operations Research, Volume 24, Issue 4, 367-377, 1997.

[5]. Ghosh, S.Reilly, D.L. Credit card fraud detection with a neural-network. Decision Support and KnowledgeBased Systems, Proceedings of the Twenty-Seventh Hawaii International Conference.Volume 3: 621-630, 1994.

[6]. Hanagandi, V.Dhar, A.Buescher, K.Density-based clustering and radial basis function modeling to generate credit card fraud scores. Computational Intelligence for Financial Engineering, 1996.

[7]. Hansen, J.V., McDonald, J.B., Messier, W.F., & Bell, T.B.A generalized qualitative- response model and the analysis of management fraud. Management Science, 42(7) 1022-1032, 1996.

[8]. Martin, D.Early warning of bank failure: A logestic regression approach. Journal of Banking and Finance, 1, 249–276, 1997.

[9]. Ohlson, J.A. Financial ratios and probabilistic prediction of bank ruptcy. Journal of Accounting Research, 18(1), 109–131, 1980.

[10]. QuinlanJ. R.C 4.5 Programs for Machine Learning, Morgan Kaufmann, SanMateo, CA. 1993.

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies

PEER REVIEWED
All submissions are rigorously peer reviewed to ensure quality.

OPEN ACCESS
Free and unrestricted access to research for all.

GLOBAL REACH
Connecting researchers and professionals worldwide.

TIMELY PUBLICATION
We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

editor@ijeast.com
www.ijeast.com
India

2455-2143