International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Credit Card Fraud Detection Using Machine Learning

G. Harika[1], V. Naveen Sai[2], V. Guru Babu[3], S. Divya[4], S. Siva Sankar[5]

[1,2,3,4,5]Department of CSE, Tadipatri Engineering College, Tadipatri.

Abstract

Nowadays, we are using credit cards a lot for online and offline payments instead of carrying money. Because of this, fraud transactions using credit card has also increased. Fraud means using someone's credit card details wrongly, which causes big money losses to card owner, banks, companies, and customers. So, it is very important to detect fraud quickly before losing more money. In this project, we use machine learning to check whether the transaction is real or fake. Here, we use real credit cards data is used for testing. Since fraud transactions are very few and different format compared to normal transactions. Different machine learning models are used to get better results and the system is tested using accuracy, precision. The results show that this method can detect fraud transactions better. Our research also explains that old rule-based systems are slow and not suitable for today's large amount of transaction data. So, we used algorithms like SVM, Logistic Regression, Random Forest, XG Boost, Decision Tree methods to improve fraud detection. These models can find hidden patterns and suspicious activities more accurately. By using these algorithms, we build a machine that checks credit card transactions using trained data and helps to detect fraudulent transactions in order to prevent financial losses. In this Project, Machine learning techniques are used to identify whether a transaction genuine or fraudulent. Real transaction data is used for testing the system. Since fraudulent transactions are very few compared to normal transactions, detecting them is challenging. To handle this problem, different machine learning models are applied and compared. The system studies transaction details such as amount, time, location and spending behaviour. Algorithms Decision Tree, Logistic project Regression, Support Vector Machine, Random Forese and XG Boost are used to analyse the data. These methods perform better than traditional rule-based systems because they can learn patterns from past transactions. The results show that the proposed system can detect fraud more accurately and help reduce financial losses.

Keywords: SVM, Logistic Regression, Random Forest, XG Boost, Decision Tree, Credit Card, Transactions, Payments, Banks, Fraud Transactions.

INTRODUCTION

Credit card fraud is increasing & causes money loss for people and organizations. Fraud mainly happens when stolen card details are used for wrong transactions. Fraud cases are few, finding them is very difficult. Most of the fraud transactions are just like normal ones, that's why finding them is difficult. Machine Learning helps computers to learn patterns from past card transactions. These models look that how the users, cards and shops are connected to find strange behaviour. Computer can learn from past card transactions to find differences between normal and fake transactions. Different Machine Learning methods like SVM, Logistic Regression, Random Forest, etc on the past transactions data. The method

focuses on being fast and accurate at the same time. It can handle different types of transaction behaviours easily. Each model checks the data in a different way. The system improves itself by learning from past fraud cases. When a payment looks strange it gets extra attention. Credit card refers to unauthorized and illegal transactions carried out without the knowledge or permission of the cardholder. Detecting such fraud using traditional manual methods is difficult because these methods rely on fixed rules and human intervention.

Since millions of transactions occur every day, manual verification of each transaction is particularly impossible. To overcome, these challenges, machine learned has emerged as a powerful and effective solution for credit card fraud detection. Machine learning techniques analyse large volumes of historical transaction data. Credit card fraud happens when someone uses another person's credit card details without permission to make transactions. With the growth of online shopping and digital payments, Credit card frauds has increased rapidly. To reduce these losses, automated systems are required that can identify suspicious transactions quickly and accurately. Different machine learning algorithms examine transaction features such as transaction amount, time of payment, purchase location and spending habits of the card holder.

Machine learning provides an effective solution to these problems by automatically learning from past transaction data. Instead of using fixed rules, machine learning models analyse historical transaction patterns and understand normal behaviour. Traditional fraud detection systems depend on fixed rules, such as blocking a card after a certain number of failed attempts. Theses methods are slow and not effective for large amounts of transaction data. Machine learning provides a better solution by learning patterns from previous transactions and identifying unusual behaviour. Machine learning models analyse how users normally spend money and compare new transactions with past behaviour. If a transaction looks unusual, it is marked as suspicious. This approach helps banks and financial institutions detect fraud faster and improve transaction security. Overall, machine learning based fraud detection systems play an important role in making digital systems safer and more reliable.

LITERATURE REVIEW

[1] The Authors M.Adil, Z.Yinjun, M.M.Jamjoom and Z.Ullah worked on the project "optdevnet:A optimized deep event-Based network framework for credit card fraud detection" with the help of SVM, Risk Management, deep learning. This project helps to detect fraudulent transactions balanced datasets using neural networks, machine learning algorithm with 89% of accuracy. [2]The Authors C.Hout,S.Heng,T.-k.kim and Y.Han worked on the project "Quantum Autoencoder for enhanced fraud detection in imbalanced credit card dataset with the help of imbalanced dataset, quantum Autoencoder, quantum Machine learning. This project helps foundational approach they worked on is based on quantum machine learning with 99% of accuracy. [3] The Authors S.N. Kalid, K.-C.Khor, K.-H,Ng and G.-K Tong wprked on "detecting frauds and payment defaults on credit card data inherited with imbalanced class distribution and overlapping class problems" with the help of ensemble learning and sampling methods. This project means the authors did not develop a new model based on one specific theory with 98% of accuracy.

[4] The Authors F.khaled Alarfaj and S. Shahzadi, "Enhancing fraud detection in banking with deep learning: Graph Neural Networks and Autoencoders foe Real-Time Credit Card Fraud Prevention" with the help of detection algorithms, deep learning, Graph Neural Network. This Project worked on improving credit card fraud detection with of accuracy. [5] The Authors F.K. Alarfaj, I. Malik, H.U. Khan,

N.Alamusallam M. Ramzan and M. Ahmed, "Credit Card Fraud detection using State-of-the-Art Machine learning and deep learning algorithms" with the help of support vector machines, prediction algorithms, machine learning algorithms. This project helps the theoretical foundation of the work is based on applying and comparing advanced machine learning and deep learning techniques with 99.72% of accuracy. [6] The authors E.lleberi, Y.Sun and Z.Wang, "performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost" with the help of SVM, boosting, Random Forests. This project helps built their work on machine learning theory with data imbalance handling and ensemble boosting techniques with 99.99% of accuracy.

[7] The Authors T.-T.-H. Le, Y.Hwang, H.Kang and H.Kim, "Robust credit card fraud detection Based on Efficient Kolmogorov-Arnold Network Models" with the help of Machine learning algorithms, Feature Extraction. This project helps theory and Foundational Approach they worked on is centred around a Kolmogorov-Arnold Network with 99.96% of accuracy. [8] The Authors J.Jemai, A.Zarrad and A.Daud, "identifying credit card Transactions using Ensemble Learning" with the help of Data Models, Data Integrity. This Project helps built their Fraud detection approach on the theory of ensemble learning with 99% of accuracy. [9] The Authors A.Mniai, M.Tarik and K.jeberi, "A Novel framework for credit card fraud detection" with the help of credit cards, SVM, Feature Extraction prediction algorithms. This project helps built their fraud detection system on multiple theoretical concepts from machine learning with 99% of accuracy.

[10] The Authors E. lleberi and Y. Sun, "A Hybrid Deep learning ensemble model for credit card fraud detection" with the help of Feature Extraction, deep learning, CNN. This project helps in a structured framework to improve fraud detection performance with 99% of accuracy. [11] The Authors C.sekhar Nama and K.S. Banu, "credit card fraud detection Using Deep learning Techniques and Handling Unbalanced Class Distribution with AGGSS" with the help of Nearest neighbour methods, classification algorithms. This helps data imbalance handling theory combined with deep learning theory. [12] The Authors H.S. Alsagri, "Hybrid Machine learning-Based Multi-stage Framework for credit card Anomalies and Fraud" with the help of Machine learning, Deep learning, Feature selection, multi-stage classification. This project helps combining more than one machine learning method instead of using only a single algorithm.

[13] The Authors B. Lebichot, T. Verhelst, Y. –A. Le Borgne, L. He-Guelton, F. Oble and G. Bontempi, "Transfer learning strategies for credit card fraud detection" with the help of Feature extraction and transfer learning. This project helps transfer learning theory using knowledge learned from one dataset or task with 99% of accuracy. [14] The Authors A. A Almazroi and N. Ayub, "Online payment Fraud detection Model using Machine learning techniques, "with the help of Data models and Neural networks. This project helps past online payment data that is already labelled as fraud or normal with 98" of accuracy. [15] The Authors R. San Miguel Carrasco and M. –A. Siciilia-Urban, "Evaluation of deep neural networks for reduction of credit card fraud alerts" with the help of SVM, decision trees neural networks, deep learning. This project helps is not just to find fraud, but to reduce unnecessary fraud alerts with 91.79% of accuracy.

PROPOSED METHODOLOGY

In this proposed methodology, credit card fraud detection is done using state-of-the-art machine learning in a clear and easy way. First, a large amount of past credit card transaction data is collected. This data contains both normal transactions and fraud transactions. The data is then cleaned by removing errors,

handling missing values, and converting details like date, time, and location into numbers so the system can understand them properly. Because fraud transactions are very few compared to normal ones, special balancing methods are used so the system does not ignore fraud cases and learns from them correctly. Next, useful information is taken from the data, such as transaction amount, time gap between payments, place of purchase, type of merchant, and spending behaviour of the cardholder. These details help the model understand what is normal behaviour and what looks unusual. After this, advanced machine learning models like random forest, gradient boosting, support vector machines, and deep learning models are trained. These models learn hidden patterns and relationships in the data and become good at spotting unusual transactions that may indicate fraud. Once the training is completed, the system is tested using new data to check how accurately it can detect fraud. When a real-time transaction happens, the trained model quickly analyses it and decides whether it is safe or suspicious. If the transaction looks risky, an alert is generated for manual verification or automatic blocking. Over time, the system keeps learning from new data, which helps it adapt to new fraud techniques. This proposed method improves fraud detection accuracy, reduces financial loss, and provides better security in a simple and effective manner.

Decision Tree Algorithm: -

A Decision Tree is a supervised machine learning algorithm that works by splitting data into smaller subsets based on feature values. In credit card fraud detection, past transaction data is first collected. Each transaction includes features such as transaction amount, time, location, merchant type, and whether the transaction is fraudulent or genuine. The algorithm builds a tree-like structure where each internal node represents a decision based on one feature each branch represents the outcome of that decision, and each leaf node gives the final classification as fraud or non-fraud. The splitting is done using measures like Gini Index or Information Gain, which help choose the feature that best separates fraudulent transactions from genuine ones. Decision Trees are easy to understand and interpret, which makes them very useful in financial systems where transparency is important. They can handle both numerical and categorical data and do not require complex data preprocessing. However, a single decision tree can easily overfit the training data, especially when fraud cases are very few compared to normal transactions. Despite this limitation, Decision Trees are often used as base models and form the foundation for more powerful ensemble methods like Random Forest and XG Boost.

Support Vector Machine (SVM): -

Support Vector Machine (SVM) is a supervised learning algorithm mainly used for classification problems. In credit card fraud detection, SVM aims to separate fraudulent transactions from legitimate ones by finding an optimal boundary, called a hyperplane, between the two classes. Each transaction is represented as a point in a high-dimensional space, where dimensions correspond to features such as amount, time, and transaction location. SVM tries to maximize the margin, which is the distance between the hyperplane and the closest data points from both classes, known as support vectors. A larger margin generally leads to better generalization and improved accuracy. One of the key strengths of SVM is its ability to handle high-dimensional data and complex patterns. Using kernel functions SVM can transform data into higher dimensions, making it possible to separate non-linear fraud patterns. This is especially useful because fraudulent transactions often follow hidden and complex behaviours. However, SVM can be computationally expensive for very large datasets, which is common in real-world banking systems. It

is also sensitive to parameter selection and class imbalance. Despite these challenges, SVM provides strong performance and is effective when detecting subtle and rare fraud patterns.

Logistic Regression:

Logistic regression learns the pattern from these past transactions. The result is a value between 0 and 1. Logistic Regression is a simple yet powerful supervised machine learning algorithm widely used for binary classification problems like credit card fraud detection. It predicts the probability that a transaction is fraudulent or genuine based on input features such as transaction amount, time, and location. Unlike linear regression, logistic regression uses a sigmoid function to map predicted values between 0 and 1. The output is interpreted as a probability. If the probability is above a chosen threshold (for example, 0.5), the transaction is classified as fraud otherwise, it is considered normal. Logistic regression works by learning the relationship between input features and the target variable using labelled historical transaction data. Each feature is assigned a weight and these weights indicate how strongly that feature contributes to fraud detection. This makes the model highly interpretable, which is important in financial applications where decisions must be explained clearly. The algorithm is fast, efficient, and performs well when the relationship between features and output is approximately linear. However, it may struggle with complex and non-linear fraud patterns.

Random Forest Algorithm: -

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting. In credit card fraud detection, the algorithm uses historical transaction data where each transaction is labelled as fraud or non-fraud. Instead of building a single decision tree, Random Forest creates many trees using different random subsets of data and features. Each tree independently classifies a transaction and the final decision is made by majority voting. This collective decision-making process makes the model more robust and accurate than a single decision tree. One major advantage of Random Forest is its ability to handle large datasets with high-dimensional features. It can capture complex interactions between variables, which is essential for identifying fraud patterns that are not obvious. It also handles missing values well and is less sensitive to noise in the data. Random Forest reduces overfitting by averaging the results of many trees, making it suitable for imbalanced datasets like credit card transactions where fraud cases are rare. However, the model is less interpretable compared to a single decision tree and requires more computational resources. Despite this, Random Forest is widely used in real-world fraud detection systems due to its strong performance and reliability.

XG Boost: -

XG Boost (Extreme Gradient Boosting) is an advanced and highly efficient machine learning algorithm based on boosting techniques. It is one of the most powerful models used in credit card fraud detection due to its high accuracy and speed. In XG Boost, decision trees are built sequentially, where each new tree focuses on correcting the errors made by the previous trees. This allows the model to learn complex fraud patterns more effectively. The algorithm uses gradient descent optimization to minimize loss and improve prediction performance at each step. XG Boost includes several regularization techniques that help prevent overfitting, which is a common problem in fraud detection. It also supports parallel processing, making it suitable for large-scale transaction data. The model efficiently handles missing values and imbalanced datasets, which are typical in credit card fraud problems. Another advantage of XG Boost is its flexibility

and ability to fine-tune parameters such as learning rate, tree depth, and number of estimators. Although the model is complex and less interpretable compared to simpler algorithms, its superior performance makes it a top choice in financial fraud detection systems. XG Boost is widely used in industry and has shown excellent results in detecting rare and sophisticated fraud transactions.

SYSTEM ARCHITECTURE

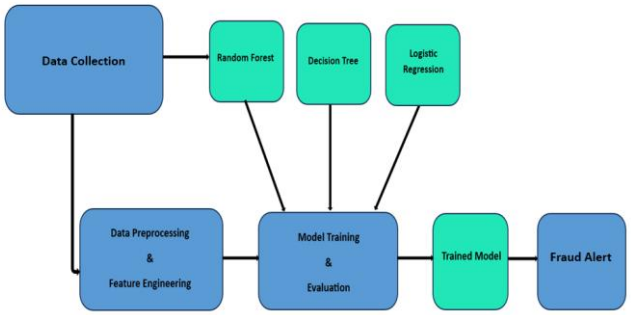

FIG 1.SYSTEM ARCHITECTURE

This Architecture is a typical system architecture for a machine learning pipeline. In this architecture random Forest, Decision Tree, and Logistic regression Algorithms are used. The process starts with Data Ingestion. In This data Preprocessing and Feature Extraction Raw Data undergoes Preprocessing and Feature Engineering to prepare it for training. Based on the Algorithms like Random Forest, Decision Tree, Logistic Regression data is used to Train and evaluate by using Various Machine learning Models.

Mathematical Concepts: -

Principal Component Analysis (PCA):

Used to anonymize and reduce dimensionality of transaction features. The transformation is defined as:

$Z = XW$

where **X** is the original data matrix, **W** is the matrix of principal components, and **Z** is the transformed data.

Correlation Coefficient:

Measures linear relationship between two variables:

$r = \sum[(x_i - \mu_x)(y_i - \mu_y)] / [n \cdot \sigma_x \cdot \sigma_y]$

where **μ** is the mean and **σ** is the standard deviation.

Confusion Matrix & Metrics:

For binary classification (fraud/non-fraud), metrics are:

Precision: Precision = TP / (TP + FP)

Recall: Recall = TP / (TP + FN)

F1 Score: F1 = 2 · (Precision · Recall) / (Precision + Recall)

ROC Curve & AUC:

The ROC curve plots True Positive Rate (TPR) vs. False Positive Rate (FPR):

- TPR=TP/(TP+FN)
- FPR = FP / (FP + TN)

The Area Under the Curve (AUC) quantifies overall model performance.

Anomaly Detection:
Unsupervised methods often use statistical thresholds:
$z = (x - \mu) / \sigma$
Transactions with high $|z|$ scores may be flagged as anomalies.

RESULTS & DISCUSSION: -
For this Credit card fraud detection project, we have used machine learning models. In the dataset fraud transactions are very less that's why they observed class imbalance problem very clearly. We used Precision, recall for analysing results. As per Results, linear models don't completely capture fraud patterns. Then compare to linear model, Non-linear model, Random Forest and XG Boost gives better performance. These models can easily learn complex transactions and identify fraud correctly. These metrics give a clear understanding of how reliable and practical each mode is in real-world situations. From the analysis it was observed that similar linear models, such as logistic regression, were not able to fully capture the complex behaviour of fraudulent transactions. On the other hand, advanced models like Random Forest and XG Boost showed much better performance. As a result, we were able to correctly identify most fraudulent transactions marked as fraud This balance is very important because fewer false alerts improve customer trust and reduce unnecessary transaction blocks. Overall, the result clearly, show that ensemble-based models are more suitable and effective for credit card fraud detection.

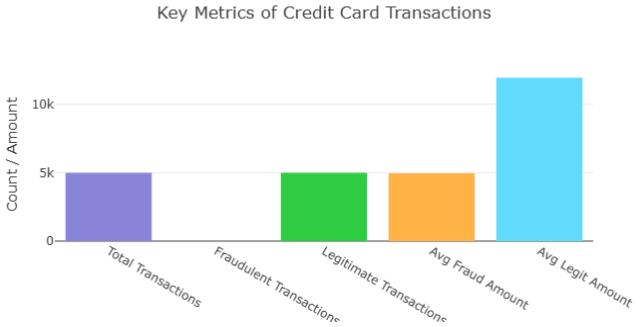

FIG 2.BAR GRAPH

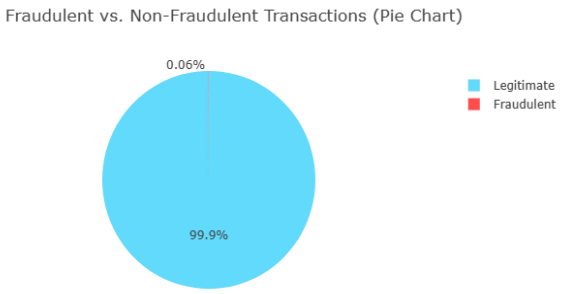

FIG 3.PIE CHART

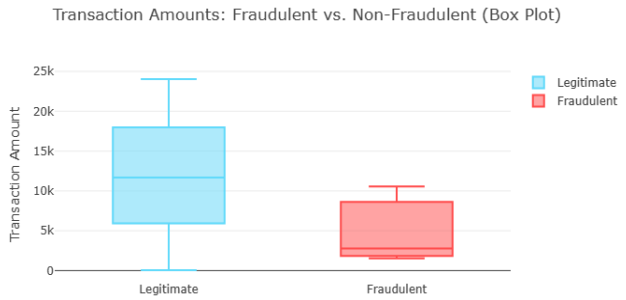

FIG 4. COMPARSION

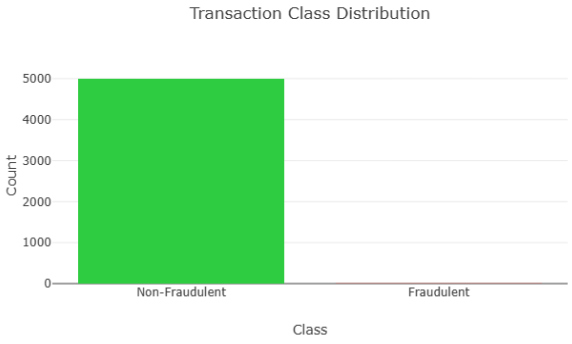

FIG 6.GRAPH

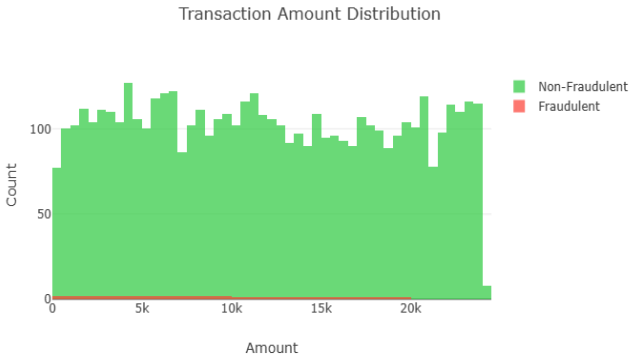

FIG 7.HISTOGRAM

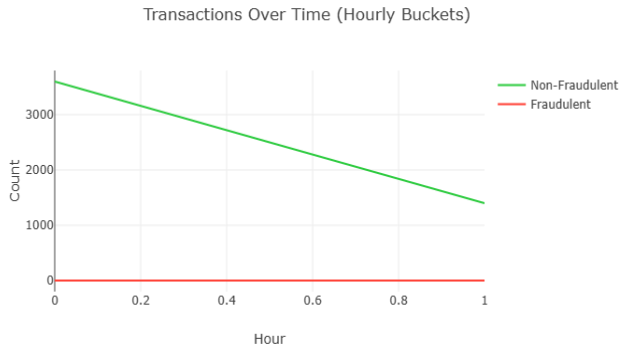

FIG 8.LINE GRAPH

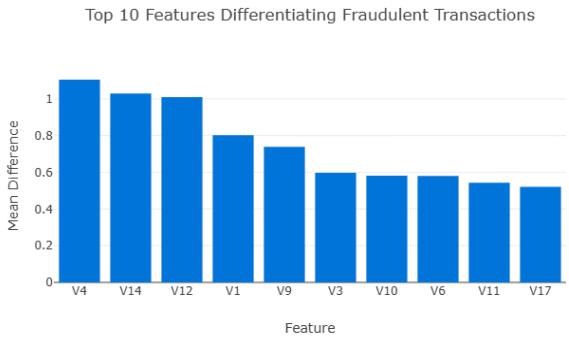

FIG 9.BAR GRAPH

SCREENSHOTS

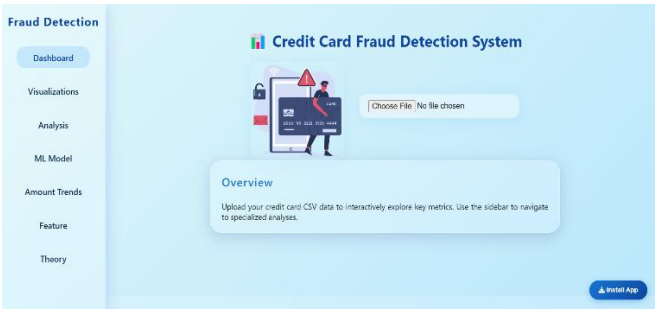

FIG 10.DASHBOARD PAGE

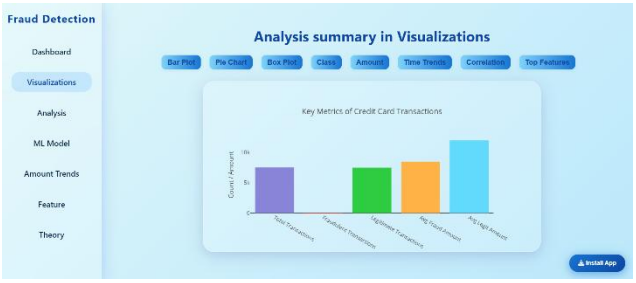

FIG 11.VISUALIZATION

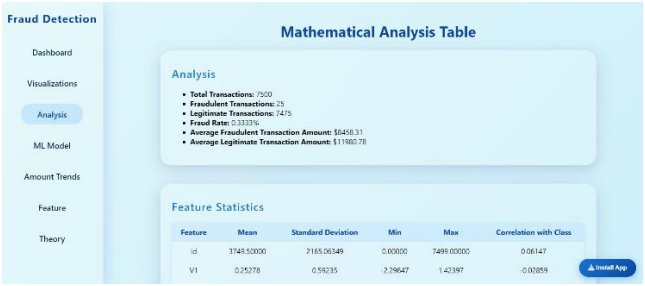

FIG 12.ANALYSIS PAGE

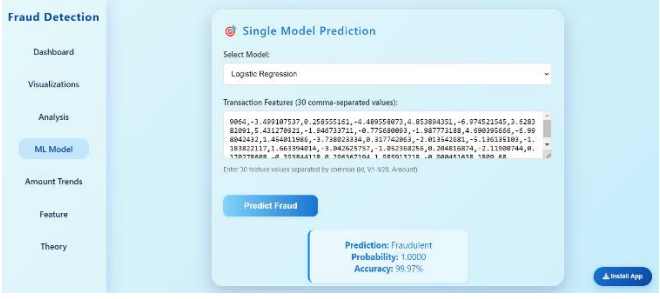

FIG 13.PREDICT PAGE

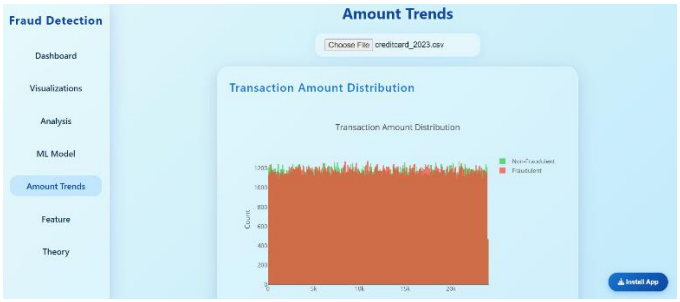

FIG 14.AMOUNT TRENDS PAGE

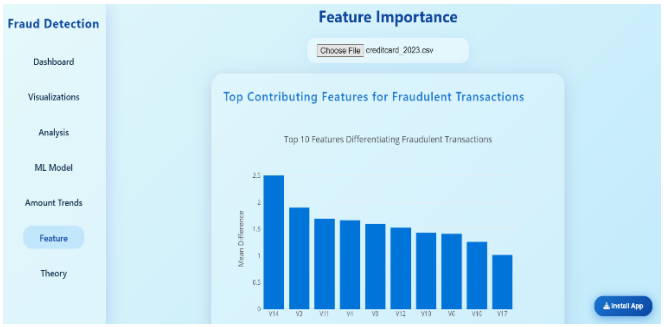

FIG 15.FEATURE PAGE

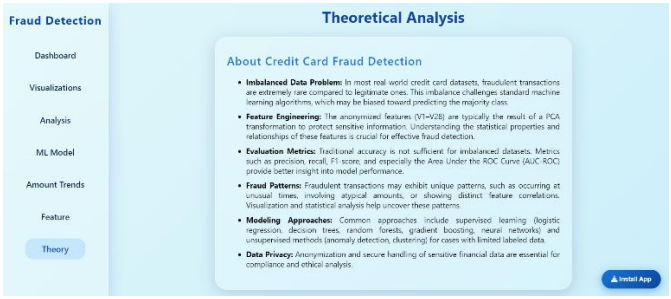

FIG 16.THEORY PAGE

CONCLUSION & FUTURE SCOPE

This project successfully implemented using machine learning techniques. Class imbalance problem was major challenge for this because of fraud transactions are very low in the dataset. That's why Precision, recall F1-Score etc. Metrics used in performance evaluation. As per Result Random Forest and XG Boost models are performed better than linear models. These models learned complex and Non-linear transactions patterns easily. Still this one classified some genuine transactions as fraud but this doesn't miss any fraud transactions. Overall, the combination of Forecast and XG Boost provided strong and reliable results on highly imbalance datasets. Finally, this one provided that machine learning based fraud detection systems are suitable for real-world applications. In the future, deep learning techniques like neural networks can be used to achieve much better results. Real-Time fraud detection systems can be developed to instantly analyse live transactions. Accuracy will improve further by utilizing large datasets and the last transaction patterns. This system can be implemented across online Payments, mobile wallets. Overall, in the future, this system will become smarter, faster, and more secure.

REFERENCES

1. M.Adil, Z. Yinjun, M. M. Jamjoom and Z. Ullah, "OptDevNet: A Optimized Deep Event-Based Network Framework for Credit Card Fraud Detection," in IEEE Access, vol. 12, pp 132421-132433,2024,doi: 10.1109/ACCESS.2024.3458944.
2. C.Huot, S.Heng, T.-K.Kim and Y. Han, "Quantum Autoencoder for Enhanced Fraud Detection in Imbalanced Credit Card Dataset," in IEEE Access, vol. 12, pp. 169671-169682, 2024, doi: 10.1109/ACCESS.2024.3496901.
3. S. N. Kalid, K. -C. Khor, K. -H. Ng and G. -K. Tong, "Detecting Frauds and Payment Defaults on Credit Card Data Inherited With Imbalanced Class Distribution and Overlapping Class Problems: A Systematic Review," in IEEE Access, vol. 12, pp. 23636-23652, 2024, doi: 10.1109/ACCESS.2024.3362831F.
4. F. Khaled Alarfaj and S. Shahzadi, "Enhancing Fraud Detection in Banking With Deep Learning: Graph Neural Networks and Autoencoders for Real-Time Credit Card Fraud Prevention," in IEEE Access, vol. 13, pp. 20633-20646, 2025, doi: 10.1109/ACCESS.2024.3466288.
5. F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," in IEEE Access, vol. 10, pp. 39700-39715, 2022, doi: 10.1109/ACCESS.2022.3166891.
6. E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," in IEEE Access, vol. 9, pp. 165286-165294,2021,doi: 10.1109/ACCESS.2021.3134330.
7. T. -T. -H. Le, Y. Hwang, H. Kang and H. Kim, "Robust Credit Card Fraud Detection Based on Efficient Kolmogorov-Arnold Network Models," in IEEE Access, vol. 12, pp. 157006-157020,2024,doi: 10.1109/ACCESS.2024.3485200.
8. J. Jemai, A. Zarrad and A. Daud, "Identifying Fraudulent Credit Card Transactions Using Ensemble Learning," in IEEE Access, vol. 12, pp. 54893-54900, 2024, doi: 10.1109/ACCESS.2024.3380823.
9. A. Mniai, M. Tarik and K. Jebari, "A Novel Framework for Credit Card Fraud Detection," in IEEE Access, vol. 11, pp. 112776-112786,2023,doi: 10.1109/ACCESS.2023.3323842.
10. E. Ileberi and Y. Sun, "A Hybrid Deep Learning Ensemble Model for Credit Card Fraud Detection," in IEEE Access, vol. 12,pp.175829175838,2024,doi:10.1109/ACCESS.2024.3502542.
11. C. Sekhar Nama and K. S. Banu, "Credit Card Fraud Detection Using Deep Learning Techniques and Handling Unbalanced Class Distributions With AGSS," in IEEE Access, vol.14,pp.18471864,2026,doi:10.1109/ACCESS.2025.3649833.
12. H. S. Alsagri, "Hybrid Machine Learning-Based Multi-Stage Framework for Detection of Credit Card Anomalies and Fraud," in IEEE Access, vol. 13, pp. 77039-77048, 2025, doi: 10.1109/ACCESS.2025.3565612
13. B. Lebichot, T. Verhelst, Y. -A. Le Borgne, L. He-Guelton, F. Oblé and G. Bontempi, "Transfer Learning Strategies for Credit Card Fraud Detection," in IEEE Access, vol. 9, pp. 114754-114766,2021,doi: 10.1109/ACCESS.2021.3104472.
14. A. A. Almazroi and N. Ayub, "Online Payment Fraud Detection Model Using Machine Learning Techniques," in IEEE Access, vol. 11, pp. 137188-137203, 2023, doi: 10.1109/ACCESS.2023.3339226.

15. R. San Miguel Carrasco and M. -Á. Sicilia-Urbán, "Evaluation of Deep Neural Networks for Reduction of Credit Card Fraud Alerts," in IEEE Access, vol. 8, pp. 186421-186432,2020,doi: 0.1109/ACCESS.2020.3026222.