International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Bank Fraud Detection Using Machine Learning Algorithms

Mr. M.K.D.S.Prasadu[1], Mantina Harshitha[2], Anguluri Naveen Kumar[3], Koppula Kavya[4], Kota Samuel Finni[5], Akkanaboina Durga Vara Prasad[6]

[1]Assistant Professor, Cse Department, Vsm College Of Engineering, Ramachandrapuram, Dr.B.R.Ambhedkar Konaseema Dt, Andhra Pradesh - 533255
[2,3,4,5,6]Students, Cse Department, Vsm College Of Engineering, Ramachandrapuram, Dr.B.R.Ambhedkar Konaseema Dt, Andhra Pradesh - 533255

ABSTRACT
The banking sector is a very important sector in our present day generation where almost every human has to deal with the bank either physically or online. Financial fraud is a growing concern in the banking sector, leading to significant monetary losses and security threats. In banking the financial fraud includes transaction fraud, synthetic identity fraud, Loan and insurance fraud, Money laundering etc. Traditional rule-based fraud detection methods are often ineffective in identifying sophisticated fraudulent activities. Hence, implementing a machine learning (ML)-based fraud detection system is essential to improve accuracy, reduce false positives, and enhance security in financial transactions. This project aims to develop an intelligent system that can detect fraudulent banking transactions in real-time, ensuring a safer and more reliable banking experience.

Keywords: Machine Learning Algorithms, Fraud detection, logistic Regression, Random Forest

INTRODUCTION

With the increasing use of digital banking, fraud detection has become a critical challenge for financial institutions. Financial fraud is the act of gaining financial benefits by using illegal and fraudulent methods. Financial fraud in banks can be committed in different areas, such as transaction fraud, synthetic identity fraud, Loan and insurance fraud, Money laundering etc. Despite several efforts to reduce financial fraudulent activities in banks, its persistence affects the economy and society adversely, as large amounts of money are lost to fraud every day. Several fraud detection approaches were introduced many years ago. Most traditional methods are manual, and this is not only time consuming, costly, and imprecise but also impractical. With the advancement of the artificial intelligence approach, machine learning and data mining have been utilized to detect fraudulent activities in the financial sector. Both unsupervised and supervised methods were employed to predict fraud activities. The banking industry is experiencing a peak in fraudulent transactions, according to data from the Nigeria Inter-Bank Settlement System (NIBSS). From being committed by lone wolves to organized crime and fraud rings that employ advanced techniques to seize accounts and perpetrate fraud, fraud has changed over time. Javelin research estimates that 6 to 8 million Americans fell victim to card fraud in 2007. In 2007, such fraud on pre-existing accounts resulted in losses exceeding $3 billion. $4.84 billion is the estimated cost to the industry, according to the Nilson Report. In 2007, Javelin estimates that the losses were approximately $306 billion, more than six

times that amount. Naturally, since fraud occurs everywhere, it is not a domestic product. For example, in 2006, the UK economy lost GBP 423 million due to card fraud losses.

EXISTING SYSTEM

In current banking environments, fraud detection systems mostly rely on **rule-based logic and manual verification** methods. These systems operate by defining fixed rules or thresholds (e.g., maximum transaction amount, login location mismatches) to identify potentially fraudulent activities. Once such activities are flagged, a manual review is often required to confirm fraud. These systems define fixed rules such as:

Transactions above a certain amount are flagged.

Transactions from different geographical locations within a short time are considered suspicious.

A fixed number of failed login attempts leads to temporary account lockout.

While these methods work in basic scenarios, they lack the ability to adapt to new fraud techniques. Most existing systems are:

Manually intensive, requiring human verification after alerts are raised.

Easily bypassed by intelligent attackers who study the rules.

Unable to learn from new patterns or behaviors.

Additionally, mobile payments and online banking have created new attack vectors, such as SIM swap fraud, malware-based attacks, and unauthorized device access, which are not effectively handled by these older systems. Moreover, these systems are unable to differentiate between legitimate behavior variations and actual fraud attempts, leading to false positives and poor user experience.

PROPOSED SYSTEM

In proposed methodology, Detection of fraudulent activity is thus critical to control these costs. This paper hereby addresses bank fraud detection via the use of machine learning techniques; association, clustering, forecasting, and classification to analyze the customer data to identify the patterns that can lead to frauds. Upon identification of the patterns, adding a higher level of verification/authentication to banking processes can be added. These kinds of frauds can be credit card fraud, insurance fraud, accounting fraud, etc. which may lead to the financial loss to the bank or the customers. Thus, detection of these kinds of frauds are very important. Fraud detection in banking sector is based on the machine learning techniques and their collective analysis from the past experiences and the probability of how the fraudsters can steal from customers and banks. Therefore, this paper addresses the analysis of data mining techniques of how to detect frauds and overcoming it in banking sector.

YSTEM SPECIFICATIONS

1.SCOPE OF THE PROJECT

The scope of the Bank Fraud Detection using Machine Learning Algorithms project compasses Several key aspects**:**

Objective: To detect and prevent bank frauds using machine learning techniques. The system analyses user behavior and transaction patterns to identify suspicious activities in real time. It aims to enhance security, reduce financial loss, and support bank admins with timely fraud alerts and insight.

2. HARDWARE REQUIREMENTS

System: Pentium IV 2.4GHZ

Hard Disk: 40 Gb
Ram:512 Mb

3. SOFTWARE REQUIREMENTS

OS : Window 10(64 bit)
IDE: Jupiter notebook
Domain: Machine learning

SYSTEM ARCHITECTURE

The system architecture of the Bank Fraud Detection project is designed to automatically evaluate and classify banking transactions using machine learning. When a user initiates a transaction, the system extracts relevant features such as amount, time, location, transaction type, and device details. These features are passed to a trained machine learning model that analyses them in comparison to historical transaction data. The model then determines whether the transaction is fraudulent or non-fraudulent based on learned patterns and behavior .All the necessary historical data is stored in a secure transaction database, which is used for both training and real-time evaluation. The architecture ensures that transactions are assessed quickly and accurately, reducing the risk of financial fraud while maintaining a smooth user experience. This intelligent and automated setup improves security and enables proactive detection of fraudulent activities.

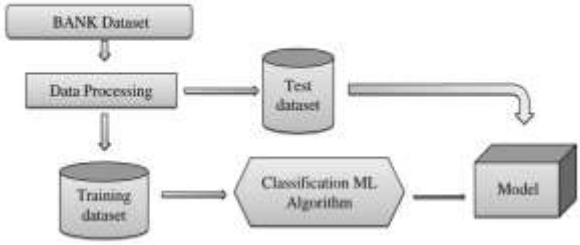

Fig:1 System Architecture diagram

PROCESS WORKFLOW

A process flow diagram (PFD) visually represents the sequence of steps in a system or process. It shows how data or tasks move from one step to another using standardized symbols like arrows, rectangles. In fraud detection, it outlines stages like data input, pattern matching, and decision-making.PFDs help in understanding system logic, identifying bottlenecks, and improving efficiency. They are commonly used in engineering, software development, and business Processes..

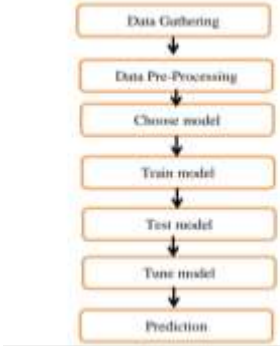

Fig:2 Process workflow diagram

IMPLEMENTED ALGORITHM

Logistic Regression

Logistic regression is a simple yet effective machine learning algorithm used for detecting bank fraud by classifying transactions as either fraudulent or legitimate. It works by analyzing various features of each transaction, such as the amount, time, location, and customer behavior, and then calculating the probability that a transaction is fraudulent. This probability is generated using a mathematical function called the sigmoid function, which outputs values between 0 and 1. If the probability exceeds a certain threshold, the model classifies the transaction as fraud. By training the model on historical transaction data, it learns to recognize patterns associated with fraudulent activity and can be used to make real-time predictions on new transactions, helping banks quickly identify and prevent fraud.

Logistic regression is a statistical technique used for binary classification jobs that require predicting one of two outcomes. Unlike linear regression, which predicts a continuous value, logistic regression forecasts the likelihood that a given input belongs to a specific class.

b)Random Forest

Random Forest is a powerful and widely used machine learning algorithm in fraud detection. It works by building multiple decision trees during training and combining their outputs to make a final prediction.A Random Forest classifier is initialized with 100 decision trees.The model is trained using the preprocessed training data.Accuracy is evaluated using confusion matrices and F1 score.A Classification report is generated to show precision ,recall and F1 score for fraud detection

WORKING OF MACHINE LEARNING ALGORITHMS

The first step is to obtain a dataset of raw data. This dataset is next subjected to preprocessing, a vital phase that entails cleaning the data by dealing with missing values, deleting duplicates, and normalising features for consistency. Next, the dataset is divided into two parts: training data and testing data. The training data is used to train the machine learning model, which means that the model learns from the data and adjusts its internal parameters to reduce prediction errors. The testing data, which the model did not view during training, is used to assess the model's performance and generalization. Creating the model entails choosing a good machine learning algorithm and configuring it with the required hyperparameters. The training phase then iteratively improves the model by reducing the discrepancy between its predictions and the actual outcomes in the training data. Finally, the model's accuracy and performance are evaluated against the testing data to ensure that it can make correct predictions on fresh, previously unseen data. This process can be repeated with other algorithms and preprocessing techniques to improve the model's performance.

RESULT ANALYSIS

Fig:3

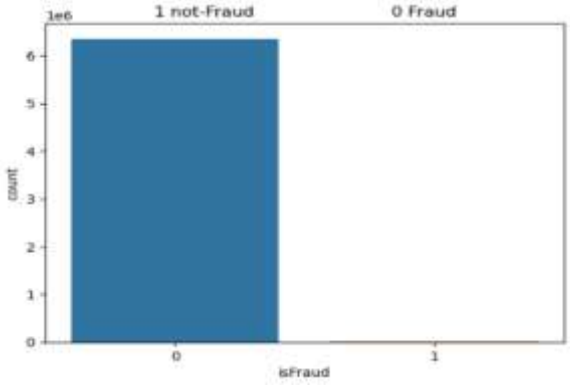

Fig:4

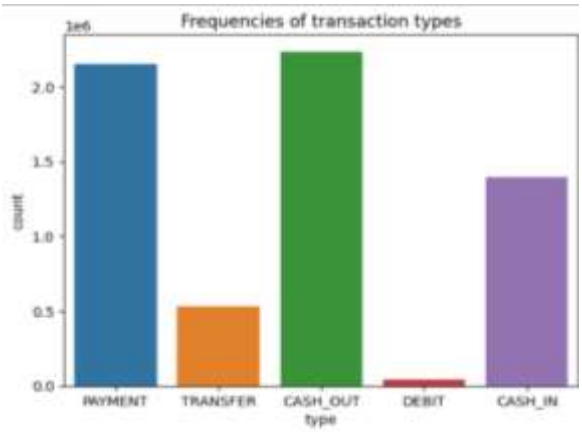

Fig:5

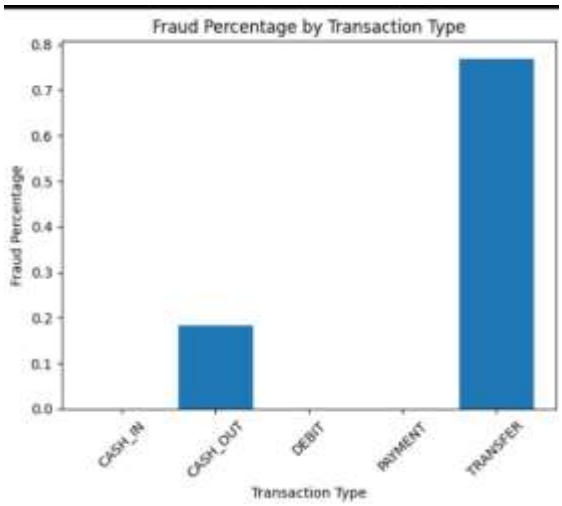

Fig:6

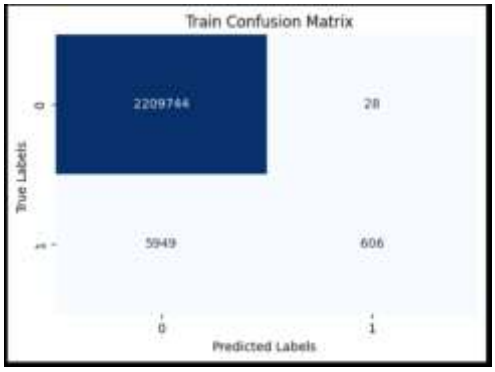

Fig:7

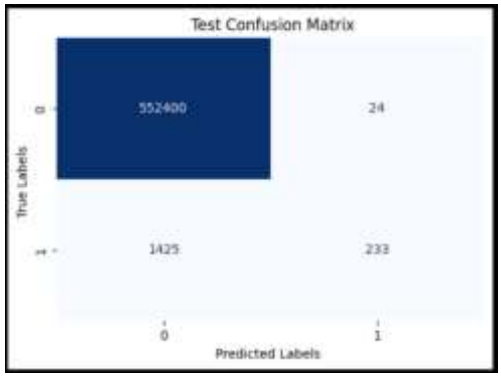

Fig:8

F1 score of train and test

```
from sklearn.metrics import f1_score

f1 = f1_score(y_train,y_train_pred)
print("F1 Score of train data:", f1)

f2 = f1_score(y_test,y_test_pred)
print("F1 Score of test data:", f2)

F1 Score of train data: 0.16859090276811797
F1 Score of test data: 0.24334203655352485
```

Fig:9

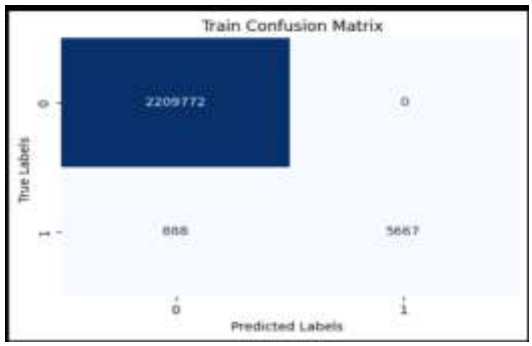

Fig:10

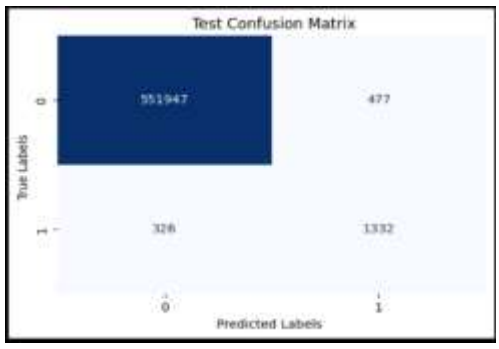

Fig :11

F1 score of train and test

```
from sklearn.metrics import f1_score
f1 = f1_score(y_train,y_train_pred_rf)
print("F1 Score of train data:", f1)

f2 = f1_score(y_test,y_test_pred_rf)
print("F1 Score of test data:", f2)

F1 Score of train data: 0.9273441335297005
F1 Score of test data: 0.7683876550331699

from sklearn.metrics import classification_report
report = classification_report(y_test, y_test_pred_rf)
print(report)
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	552424
1	0.74	0.80	0.77	1658
accuracy			1.00	554082
macro avg	0.87	0.90	0.88	554082
weighted avg	1.00	1.00	1.00	554082

Fig:12

CONCLUSION

Machine Learning is a technique used to extract vital information from existing huge amount of data and enable better decision-making for the banking and retail industries. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed and the information that is captured is used throughout the organization to support decision-making. Data Mining techniques are very useful to the banking sector for better targeting and acquiring new customers, most valuable customer retention, automatic credit approval which is used for fraud prevention, fraud detection in real time, providing segment based products, analysis of the customers, transaction patterns over time for better retention and relationship, risk management and marketing.

XI. REFERENCES

1. S. M. Darwish, "An intelligent credit card fraud detection approach based on semantic fusion of two classifiers," (in English), Soft Computing, Article vol. 24, no. 2, pp. 1243-1253, Jan 2020.
2. A. Eshghi and M. Kargari, "Introducing a new method for the fusion of fraud evidence in banking transactions with regards to uncertainty," (in English), Expert Systems with Applications, Article vol.
3. 121, pp. 382-392, May 2019.
4. S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using SpatioTemporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.
5. M. Zamini and S. M. H. Hasheminejad, "A comprehensive survey of anomaly detection in banking, wireless sensor networks, social networks, and healthcare," (in English), Intelligent Decision Technologies-Netherlands, Article vol. 13, no. 2, pp. 229-270, 2019.

6. I. Gonzalez-Carrasco, J. L. Jimenez-Marquez, J. L. Lopez-Cuadrado, and B. Ruiz-Mezcua, "Automatic detection of relationships between banking operations using machine learning," (in English), Information Sciences, Article vol. 485, pp. 319-346, Jun 2019.
7. M. Pohoretskyi, D. Serhieieva, and Z. Toporetska, "The proof of the event of a financial resources fraud in the banking sector: problematic issues," (in English), Financial and Credit Activity-Problems of Theory and Practice, Article vol. 1, no. 28, pp. 36-45, 2019.