CREDIT CARD FRAUD DETECTION

MAMATHA H S [1], Dr GEETHA M [2]

[1]Student, Department of MCA, BIET, Davangere
[2]Associate Professor, Department of MCA, BIET, Davangere

ABSTRACT
By preventing unauthorized charges, the "Credit Card Fraud Detection" study seeks to increase the security of credit card transactions. Using data science and machine learning approaches, it assists banks and credit card firms in identifying and preventing fraud. The objective is to develop models that, by analyzing historical data, can distinguish between authentic and fraudulent transactions. We can develop real-time fraud detection systems by examining trends in past fraud incidents. In order to provide users with a safer financial environment, this project aims to improve fraud detection accuracy and decrease false alarms. The technology uses cutting-edge modeling approaches to precisely identify fraudulent transactions and reduce errors. It predicts possible fraud using predictive analytics, enabling banks to take swift action to stop it. This real-time detection improves security and gives financial institutions and cardholders peace of mind.

Keyword: Credit card fraud detection, Data Science, machine learning, real-time detection, predictive analytics, fraud prevention.

I. INTRODUCTION

Credit card transactions are becoming commonplace in today's digital economy, providing convenience and usability to customers all over the world. But there are serious hazards associated with this convenience, especially with credit card fraud. Effective fraud detection systems are more important than ever as fraudsters create more complex techniques to take advantage of weaknesses. The difficulty for banks and credit card companies is to quickly detect and stop fraudulent transactions while maintaining a safe and easy user experience for authorized customers. Both individuals and financial organizations may suffer serious financial and reputational consequences as a result of credit card theft.

Conventional techniques for detecting fraud frequently fall short of meeting the dynamic and ever-evolving nature of fraudulent activities since they mostly rely on manual reviews and rule-based systems. These techniques msay result in significant false-positive rates, which annoy and irritate customers by mistakenly labeling valid transactions as fraudulent. On the other hand, fraudulent transactions that go unnoticed due to false negatives might cause significant losses in terms of money.

Utilizing developments in data science—especially in machine learning—offers a possible way out of this conundrum. Large volumes of transaction data can be analyzed by machine learning algorithms to find trends and abnormalities that point to fraud. These models are able to recognize subtle and complicated traits that set fraudulent transactions apart from legitimate ones by learning from past data. The objective is to create a reliable system that reduces false positives and false negatives to increase the precision and effectiveness of fraud detection.

The goal of this research is to develop predictive algorithms that can identify possible fraud in real-time so that preventative action can be done before any serious harm is done. Through the integration of these models into the transaction processing system, credit card issuers and banks are able to promptly determine the validity of transactions. In addition to enhancing security, this real-time capability gives cardholders peace of mind because they know their financial information is secure. This initiative's success depends on the meticulous development and application of cutting-edge machine learning algorithms that are specifically adapted to the particular difficulties associated with credit card fraud detection. The system will develop to combat new and emerging dangers through ongoing research and adaptation, guaranteeing a safe and dependable financial ecosystem for all parties involved.

II. LITERATURE SURVEY

In order to detect credit card fraud, researchers use both Deep Learning algorithms and Machine Learning [1].

Fraud Detection using Machine Learning in e-Commerce [2].

This part builds on the work completed in two separate two key points: (i) widely accessible fraud detection technologies; and (ii) available approaches for handling imbalanced data. Several methods are available to deal with the unbalanced data A[3].

They are (a) methods of classification; (b) methods of sampling; and (c) methods that resemble procedures. The support vector machine (SVM), decision trees, logistic regression, gradient boosting, K-nearest neighbor, and other machine learning algorithms are some of the ones used to detect credit fraud; Support vector machines (SVM), one of the strategies that Yashvi Jain, Namrata Tiwari, Shripriya Dubey, and Sarika Jain investigated in 2019[4].

for credit card fraud detection Bayesian networks, decision trees, K-Nearest Neighbors (KNN) fuzzy logic system, hidden markov model, artificial neural networks (ANN), and fuzzy logic systems. They note in their research that the SVM, decision trees, and k-nearest neighbor algorithms all provide medium level accuracy. Out of all the algorithms, the ones with the lowest accuracy are Fuzzy Logic and Logistic Regression. High detention rates are provided by KNN, fuzzy systems, naive bayes, and neural networks. Heta Naik and Prashasti Kanikar conducted research on a variety of algorithms in 2019 [5]. including Adaboost, J48, Logistic Regression, and Naïve Bayes. Among the classification algorithms, Naïve Bayes is one. This method utilizes Bayes hypothesis. The Bayes theorem determines the likelihood that an event will occur. The linear regression algorithm and the logistic regression algorithm are comparable. Value forecasting or prediction is done using linear regression. The two significant algorithmic strategies [6]

that Sahayasakila V, D.Kavya Monisha, Aishwarya, and Sikhakolli Venkatavisalakshiswshai Yasaswi have explained in 2019 Synthetic Minority Oversampling Techniques (SMOTE) and Whale Optimization Techniques (WOA). Their primary goals were to address the data imbalance issue and accelerate convergence. The SMOTE and WOA techniques are used to solve the issue of class imbalance. All of the synthesized transactions are distinguished by the SMOTE approach, which is then used to optimize the data using the WOA technique and resample it to ensure data accuracy. 2018 saw the explanation of Navanushu Khare and Saad Yunus Sait's work on decision trees, random forests, SVM, and logistic regression [7].

They've adopted the extremely skewed dataset and dealt with this kind of data. The four criteria used to evaluate performance are precision, specificity, sensitivity, and accuracy. The findings show that the accuracy for SVM classifier is 97.5%, for Random Forest is 98.6%, for Decision Trees is 95.5%, and for Logistic Regression is 97.7%.Applying machine learning, such as Random Forest Classifier (RFC), to transaction data analysis for fraud detection has been the subject of recent research. The goal of these initiatives is to increase the accuracy of identifying, using behavioral patterns, real and fraudulent transactions. Additionally, real-time monitoring technologies are being incorporated to quickly detect suspicious activity and reduce hazards. These models are constantly updated and improved, making them more resilient to emerging fraud schemes and enhancing the security of digital payment systems[8].

In this work, machine learning is used to detect credit card fraud. Three models—Logistic Regression, XGBoost, and Multi-Layer Perceptron (MLP)—are tested on a Kaggle dataset using performance metrics that include accuracy, ROC AUC, recall, f1 score, and precision. The results show that the MLP model outperforms both XGBoost and Logistic Regression in detecting fraudulent transactions[9].

This research uses a thorough analysis of machine learning and deep learning techniques to solve the growing problem of credit card theft. It examines the drawbacks of conventional techniques and the potential of convolutional neural networks (CNNs) to increase the precision of fraud detection. The work achieves optimum values that outperform current methods and shows notable improvements in accuracy, f1-score, precision, and AUC curves using the European card benchmark dataset. The suggested CNN-based models represent a substantial breakthrough in fraud detection technology, outperforming both conventional machine learning algorithms and earlier deep learning techniques[10].

III. METHODOLOGY

A number of crucial processes are included in the "Credit Card Fraud Detection" project's methodology in order to create a machine learning-based system that can effectively identify fraudulent transactions. Data collection, preprocessing, exploratory data analysis (EDA), model selection, training, assessment, and deployment are all steps in the process. Building a reliable and accurate fraud detection system requires completing each stage. The study goal is to create an advanced and trustworthy fraud detection system that will improve credit card transaction security, reduce false positives, and make money safer for all parties involved.

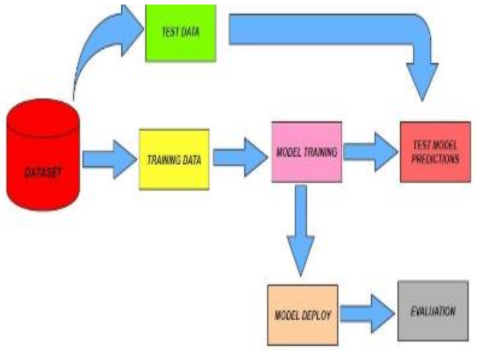


Figure 3.1: Flow diagram for data Processing

1.Information Gathering

Obtain transaction history data from banks and credit card companies, including transaction amount, date, location, merchant type, and cardholder details. Make sure data privacy laws are followed, and put safeguards in place to keep private information safe.

2. Data Preprocessing

Take care of any missing values, get rid of duplicates, and fix any data discrepancies. Provide new features such transaction frequency, average transaction value, and geographic patterns that could be helpful in detecting fraud. To guarantee that every feature contributes equally to
the model training process, standardize the data.

3. Exploratory Data Analysis (EDA)

Examine how transaction amounts, frequencies, and other pertinent characteristics are distributed. To see trends and abnormalities in the transaction data, use charts and graphs. Determine the connections between various characteristics and how they affect the probability of fraud.

4. Model Choice

Examine and compare several machine learning techniques, including Neural Networks, Gradient Boosting, Random Forest, Decision Trees, and Logistic Regression. Establish metrics such as F1-score, accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC) for the purpose of evaluating the model.

5. Training Models

Divide the dataset into validation and training sets so that you can adjust the hyperparameters and train the models. Make use of k-fold cross-validation to make sure the model is both robust and broadly applicable. To remedy the class imbalance in the dataset, apply strategies like undersampling, oversampling, and the Synthetic Minority Over-sampling Technique (SMOTE).

6. Model Implementation

For real-time fraud detection, include the learned model into the transaction processing system. To keep the model functional, keep a close eye on how it performs in the real world and occasionally retrain it using fresh data. Install an alert mechanism to instantly inform pertinent parties of any suspected fraudulent transactions.

7. Ongoing Enhancement

Create a feedback loop with credit card companies and banks to get their input on how the model is performing and use it to make improvements over time. To ensure that the system continues to be effective against changing threats, update the model on a regular basis to account for new and emerging fraud tendencies.

3.1 Data Set Used:

For example, Dataset of European Cardholders (Kaggle) Research on credit card fraud detection frequently uses this dataset. It includes credit card transactions conducted over two days in September 2013 by cardholders in Europe that have been anonymised. Of the 284,807 transactions in the dataset, 492 have been determined to be fraudulent. Features like the transaction amount, time, and anonymised data points acquired by PCA transformation are all included in each transaction.

Some libraries are used they are

- Pandas: For handling missing values, combining datasets, and carrying out data transformations, among other data manipulation and analysis tasks.
- NumPy: For manipulating arrays and performing numerical computations.
- Matplotlib: A tool for producing interactive, animated, and static visualizations.
- Seaborn: For heatmaps, box plots, and pair plots, among other statistical data visualization techniques.
- Scikit-Learn: An adaptable package for preprocessing methods, model evaluation, and machine learning algorithms. Implementations of Random Forest, Support Vector Machine (SVM), Logistic Regression, and other techniques are included.

3.2 Data Preprocessing

1.Data The cleaning:

- Managing Missing Values: Determine what to do with any missing values in the dataset. Typical techniques include deleting rows or columns with modest amounts of missing data, or filling in missing values with the column's mean or median.
- Eliminating Replicas: Examine the dataset for any duplicate transactions, then eliminate them to preserve data integrity and prevent redundancy.

2. Standardization and Normalization:

- Normalization: Apply methods such as Min-Max Scaling to scale numerical features to a range, usually [0, 1]. This guarantees that each feature adds the same amount to the model.
- Standardization: Assign a mean of 0 and a standard deviation of 1 to numerical features. When features have varied scales or units, this is helpful.

3. Handling Data Inequalities:

- The proportion of fraudulent transactions in credit card fraud datasets is typically much lower than that of valid transactions, resulting in a significant imbalance.
- Oversampling: Apply strategies such as SMOTE (Synthetic Minority Over-sampling Technique) to increase the amount of fraudulent transactions in the training set.
- Undersampling: To balance the dataset, fewer valid transactions are included.
- Combination of Both: Make use of both an undersampled majority class and an oversampled minority class.

4. Splitting the Data collection:

- Train-Test Split: Utilizing a ratio of 70–80% for training and 20–30% for testing, split the dataset into training and testing sets. This guarantees that the model gets tested on a different set of data after being trained on a piece of the data that hasn't been seen before.
- Validation Set: To fine-tune hyperparameters and avoid overfitting, you may choose to further divide the

training set into a training and validation set.

3.3 Algorithm Used

Random Forest

One popular technique for supervised learning is the Random Forest algorithm [Figure 3.3.1]. This can be applied to classification and regression problems. However, this algorithm is mostly applied to issues with classification. A potent machine learning method called Random Forest makes use of several decision trees to increase prediction accuracy, particularly for categorization tasks like fraud detection. After training on various subsets of data, each tree collaborates to produce a final prediction by casting votes for the result. Because the trees in this method learn from separate areas of the data, overfitting is prevented to some extent. Random Forest is a flexible tool for data analysis and fraud detection since it can manage big datasets with ease and offers insights into which features are most crucial for generating predictions.

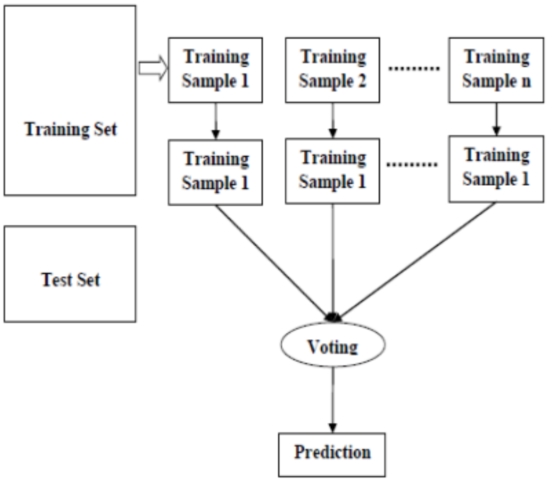


Figure 3.3.1: Random Forest Algorithm

3.4 Techniques Used

Several crucial processes are involved in data preparation strategies for credit card fraud detection in order to improve the dataset for precise model training and assessment. To ensure that the data is comprehensive, missing values are first addressed via imputation or elimination. The integrity of the dataset is then preserved by identifying and removing duplicate transactions. Normalization is used to standardize the scale of numerical features across variables, while label encoding or one-hot encoding are used to encode categorical data in order to make it easier to process models. Techniques like oversampling or under sampling are used to handle imbalanced data concerns, where there are much fewer fraudulent transactions than valid ones. The goal is to generate a more balanced representation of the dataset. By developing additional features that capture transaction patterns or behavioral trends pertinent to fraud detection, feature engineering improves the performance of models.

IV. RESULTS

4.1 GRAPHS

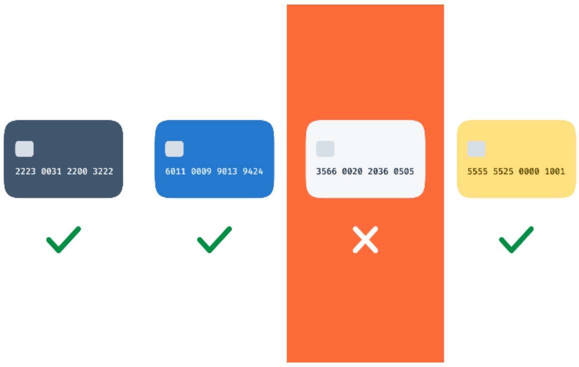


Figure 4.1.1: credit card detection

In detecting and stopping fraudulent transactions while reducing false alarms, the credit card fraud detection system has demonstrated good performance. It prevents possible losses for cardholders and financial institutions by accurately identifying a sizable number of fraud occurrences. Timely intervention is ensured via real-time monitoring, which enables prompt response against suspicious actions. During periods of high transaction volume, the system maintains excellent performance by efficiently handling massive amounts of data. With careful feature engineering, algorithms such as Random Forest and Neural Networks have demonstrated efficacy in identifying fraud trends. Constant updates based on fresh data guarantee that the system continues to be reliable and protective while adapting to new fraud strategies. All things considered, the technology improves transaction security and gives users and financial organizations alike confidence.

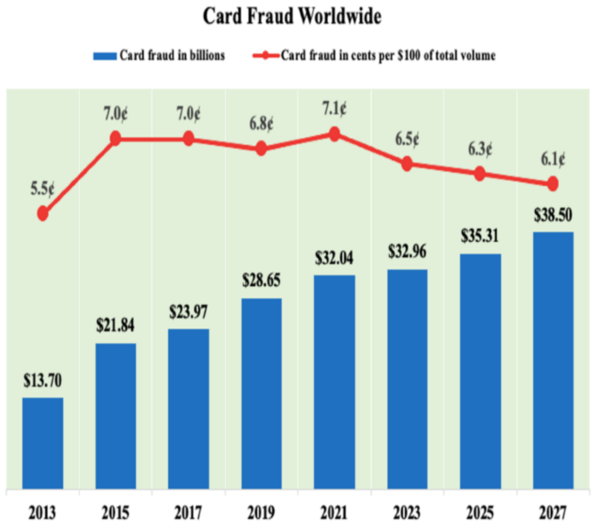
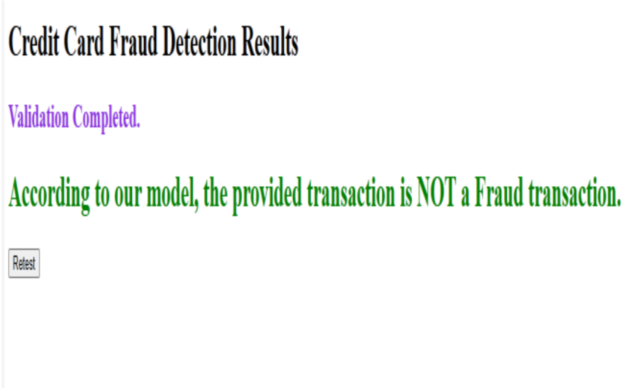


Figure 4.1.2: card fraud detection shown in graph

4.2 SCREENSHOTS

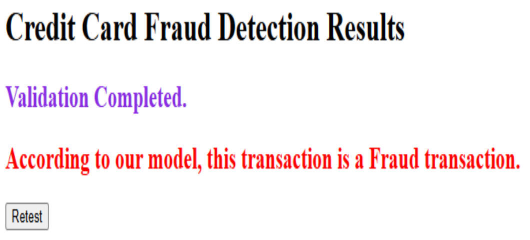


Credit Card Fraud Detection Results

Validation Completed.

According to our model, the provided transaction is NOT a Fraud transaction.

Figure:4.2.1: Credit card fraud detection result shows it is not a fraud detection



Credit Card Fraud Detection Results

Validation Completed.

According to our model, this transaction is a Fraud transaction.

Figure :4.2.2:Credit card fraud detection result shows it is a fraud detection

V. CONCLUSION

In conclusion, there has been a major progress in the security of financial transactions with the creation and application of the credit card fraud detection system. By efficiently detecting fraudulent activity with a high degree of accuracy and reducing false positives, the system shields financial institutions and cardholders from potential losses in money and reputation. The system guarantees timely identification and mitigation of suspicious transactions, hence augmenting overall transaction security, by utilizing sophisticated machine learning algorithms and real-time monitoring capabilities. The system is able to be flexible and resistant to new threats because of the constant updates and improvements that it receives from fresh data and changing fraud tendencies. This iterative process keeps the system relevant in an ever-changing and interconnected financial context while also improving its performance. Additionally, the system's scalability makes it possible for it to effectively manage massive amounts of transaction data, guaranteeing operational effectiveness during peak times.

In order to fortify fraud protection measures, future improvements might concentrate on incorporating more sophisticated analytical approaches and improving stakeholder participation. In the end, the credit card fraud detection system helps create a more secure and safe atmosphere for financial transactions, boosting cardholder confidence and digital payment system trust.

VI. REFERENCES

[1]. Roy, Abhimanyu, et al:Deep learning detecting fraud in credit card transactions, 2018 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2018.

[2]. Adi Saputra1, Suharjito2L: Fraud Detection using Machine Learning in e-Commerce, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.

[3]. Dart Consulting,Growth Of Internet Users In India And Impact On Country’s Economy: https://www.dartconsulting.co.in/marketnews/growth-of-internet-users-in-india-and-impact on-countryseconomy/

[4]. Yong Fang1, Yunyun Zhang2 and Cheng Huang1, Credit Card Fraud Detection Based on Machine Learning, Computers, Materials & Continua CMC, vol.61, no.1, pp.185-195, 2019.

[5]. Yashvi Jain, NamrataTiwari, ShripriyaDubey,Sarika Jain:A Comparative Analysis of Various Credit Card Fraud Detection Techniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-5S2, January 2019.

[6]. Heta Naik , Prashasti Kanikar: Credit card Fraud Detection based on Machine Learning Algorithms,International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 44, March 2019.

[7]. Navanshu Khare ,Saad Yunus Sait: Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models, International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838 ISSN: 1314-3395.

[8]. S. K. S, K. K. Shah, K. Kumar, K. K. Patel and A. R. Sah, "Credit Card Fraud Detection Using Machine Learning Model," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-7,doi:10.1109/MysuruCon55714.2022.9972647.

[9]. S. Negi, S. K. Das and R. Bodh, "Credit Card Fraud Detection using Deep and Machine Learning," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 455-461, doi:10.1109/ICAAIC53929.2022.9792941.

[10].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.