SYNOPSIS SUBMITTED

FOR

PROJECT

FOR THE DEGREE OF

BACHELOR OF ENGINEERING

(Computer Science and Engineering)

IN

THE FACULTY

OF ENGINEERING & TECHNOLOGY

OF

SANT GADGE BABA AMRAVATI UNIVERSITY, AMRAVATI

TITLE:

"Credit Card Fraud Detection Using Machine Learning"

GROUP MEMBERS NAMES

1. **Adib Khan**
2. **Samarpit Nandanwar**
3. **Deven Malekar**
4. **Himanshu Kadu**
5. **Suraj Pawar**

GUIDE:

PROF. P. D. KAWARE

Department of Computer Science and Engineering

H.V.P.M's College of Engineering and Technology, Amravati

CONTENTS

ABSTRACT

Credit card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use.

In the present world, we are facing a lot of credit card problems. To detect fraudulent activities the credit card fraud detection system was introduced. This project aims to focus mainly on machine learning algorithms. The algorithm used is the Logistic Regression.

Logistic Regression is a popular machine learning algorithm used in various fields, including credit card fraud detection. In the context of credit card fraud detection, Logistic Regression can be employed to predict whether a given credit card transaction is fraudulent or not based on certain features.

The methodology involves preprocessing the dataset to handle missing values, outliers, and scaling numerical features. The data is split into training and testing sets for model evaluation. The trained Logistic Regression model is assessed using metrics such as accuracy, precision, recall, F1 score, and the ROC-AUC curve. Feature importance analysis is conducted to understand the contribution of each feature in predicting fraud.

The results demonstrate the effectiveness of Logistic Regression in credit card fraud detection, providing insights into its performance, interpretability, and potential for integration into broader fraud detection systems. The study concludes with recommendations for further research and the exploration of complementary techniques to enhance fraud detection capabilities.

The study emphasizes the importance of continuous monitoring and model updates to adapt to evolving fraud patterns. The classification threshold is adjusted based on the specific requirements and costs associated with false positives and false negatives.

1. INTRODUCTION

Recently, there has been a growing usage of the credit card payment method such that most people use credit cards instead of cash when they make normal payments in their daily lives. According to the MS Windows NT kernel description, credit cards are found in most American wallets. About 7 in 10 Americans have at least one credit card. The credit card allows customers to track their spending easily and they can know where the money goes. Also, the customers have no limits on their spending, unlike the cash method which is limited to the cash in your wallet.

In addition, most companies and institutions now tend to move their business toward online services due to the rapid increase of using modern technology in all fields. Thus, online transactions (e.g. booking a hotel) require a customer to have a credit card to access the services and complete the transaction in such an efficient way that it might be hard and time-consuming to perform while using cash payment. However, a credit card is susceptible to cybercriminals causing credit card fraud. The fraudsters perform fraudulent activities by making unauthorized access to credit card information and such activities cause a financial loss for both company and customer.

Thus, the challenges of fraudulent activities increased the demand for systems to detect credit card fraud. The researchers try to build fraud detection systems using machine learning, deep learning, and data mining techniques to detect the transaction whether it is fraudulent transactions or genuine based on datasets that include information about the transactions. However, credit card fraud detection is becoming more complex since the fraudulent transactions for the cards are more and more like legal ones.

To solve this issue, credit card providers must use more sophisticated techniques to detect fraudulent transactions. One of the biggest problems in this field is the lack of good datasets since the datasets available for this problem are imbalanced datasets and have a lot of unknown fields for private insurance. Which makes it harder for the programmers to understand the dataset and build the best model that solves this problem.

2. TECHNOLOGY USED

This Project will be using the following technology stack

- Technologies
  - Python
    - Numpy
    - Pandas
    - SkLearn

- Software Used
  - Google Collab
  - Jupyter Notebook

- Algorithm Used
  - Logistic Regression

3. RELATED WORK

In this section, we review some previous work related to the Fraud Detection. M. Zareapoor and P. Shamsolmoali Presented an application based on a bagging ensemble for credit card fraud detection problems. The ensemble approach is based on a decision tree algorithm that was used for the experimental step. Moreover, this paper includes a comprehensive study of methods used such as Naïve Bayes (NB), K-nearest Neighbor (KNN), and Support Vector Machines (SVMs). A real-world credit card dataset was obtained from the UCSD-FICO competition used to evaluate their experiments using 10-fold cross-validation techniques. The dataset contains 100,000 records of credit card transactions, including their labels (legitimate and fraudulent). The evaluation measure used for evaluating the system performance such as Fraud Catching Rate, False Alarm Rate, Balanced Classification Rate, and Matthews Correlation Coefficient. The experimental results show that the bagging classifier based on the decision tree achieved the best performance.

R.Patidar and L. Sharma introduced an artificial neural network (ANN) approach with a genetic algorithm to detect fraudulent transactions. The proposed system works when the holder of the credit card uses the card in an unauthorized way; the NN tends to check the pattern that a fraudster has used and compare it with the pattern of the original cardholder to ensure that both patterns are a match or not. When there is a big difference between the original pattern and the obtained one, it represents an illegal transaction that will happen. Several features were used by NN for each transaction that fed the network such as the current transaction descriptor, transaction history descriptor, payment history descriptor, and other descriptors. Moreover, Feed Forward Back Propagation was used as a Learning Algorithm; it is considered a standard learning technique that employs gradient descent in the error space, which plays an essential role in improving efficiency. Also, it helps to pick out the parameters for the network such as weight, network type, number of layers, and number of the node. Genetic Algorithm and Neural Network (GANN) as a proposed system aims to detect credit card fraud successfully.

H.Tran and K.P.Tran used anomaly detection techniques for credit card fraud detection based on the reasons that the detection of fraud must be very flexible to track the continuous evolution of fraud over time and the occurrence of unknown anomalies. Also, they proposed two data-driven methods which are one-class support vector machine OCVM with the optimal kernel parameter selection and T2 control chart. The performance of the method was tested on a large real-time data set of online e-commerce transactions from European credit card holders which contains a total of 284807 non-fraud transactions. Moreover, simulations were performed to generate fraudulent transactions, then 284000 transactions were used for training 200 fraudulent transactions, and 200 non-fraudulent ones for testing. To evaluate the results obtained from the methods, they used accuracy, F1-score, Recall (DR), FPR, and Precision matrices.

4. PROBLEM STATEMENT

Credit card fraud is a significant concern for banks and customers alike. It can result in substantial financial losses and damage to trust and credibility. The problem statement for credit card fraud detection is to develop models that can accurately predict fraudulent credit card transactions, thereby preventing customers from being charged for items they did not purchase.

The dataset used for this problem typically contains historical transaction data, including features such as transaction amount, time, and various anonymized features. The dataset is highly imbalanced, with a small number of fraudulent transactions compared to legitimate ones. Therefore, handling the class imbalance is an important step before building the models.

The goal is to develop machine learning models that can effectively identify fraudulent transactions and minimize false positives and false negatives. Several classification algorithms, such as support vector machines and logistic regression, can be used for this task.

To address this problem, researchers and practitioners have implemented various machine-learning algorithms and techniques to detect credit card fraud. The aim is to analyze the fraud patterns and develop models to accurately classify transactions as fraudulent or legitimate.

Overall, the problem statement for credit card fraud detection involves developing machine learning models that can accurately predict fraudulent credit card transactions, thereby helping banks and customers mitigate the financial and reputational risks associated with fraud.

5. PROPOSED WORK

Credit card fraud detection involves developing models and algorithms that can accurately identify fraudulent transactions and minimize false positives and false negatives. Here is a proposed approach for credit card fraud detection:

1. **Data Preprocessing**: The first step is to preprocess the dataset, which typically includes handling missing values, scaling numerical features, and encoding categorical variables. Additionally, addressing the class imbalance is crucial, as the number of fraudulent transactions is usually much smaller than legitimate ones. Techniques such as oversampling the minority class (fraudulent transactions) or undersampling the majority class (legitimate transactions) can be employed to balance the dataset.

2. **Feature Selection**: Selecting relevant features is essential for building effective fraud detection models. Feature selection techniques, such as genetic algorithms or recursive feature elimination, can be used to identify the most informative features for fraud detection.

3. **Model Selection**: Various machine learning algorithms can be employed for credit card fraud detection, including decision trees, random forests, logistic regression, artificial neural networks, and naive Bayes classifiers 1. It is recommended to compare the performance of different models and select the one that provides the best results.

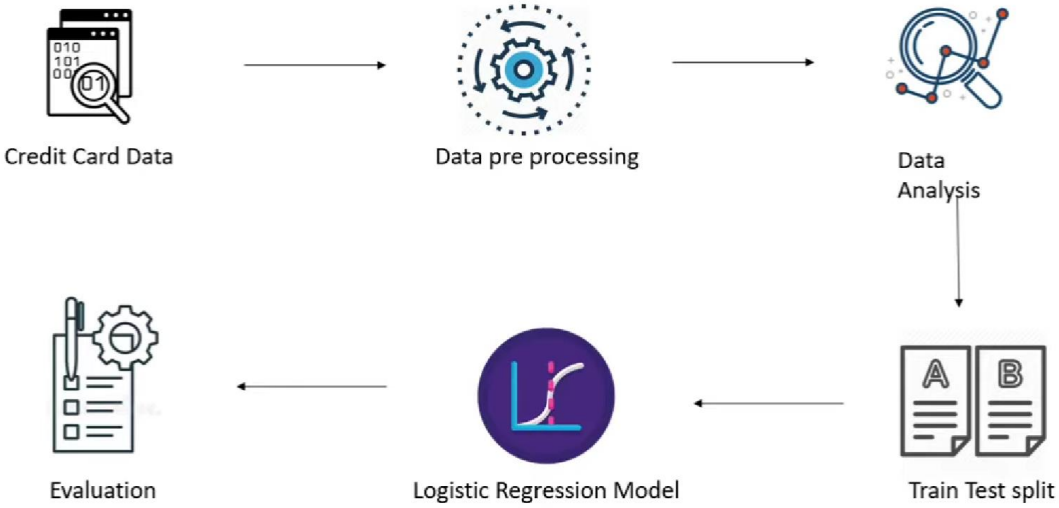
4. **Model Training and Evaluation**: The selected model is trained on the preprocessed dataset using appropriate training techniques, such as cross-validation. The model's performance is evaluated using evaluation metrics such as accuracy, precision, recall, and F1-score. It is important to consider the trade-off between false positives and false negatives, as both have different implications in credit card fraud detection.

5. **Model Optimization**: The model can be further optimized by tuning hyperparameters using grid or random search techniques. This helps in finding the best combination of hyperparameters that maximizes the model's performance.

6. **Real-time Monitoring**: Once the model is trained and optimized, it can be deployed in a real-time environment to monitor credit card transactions. The model can analyze incoming transactions and classify them as either fraudulent or legitimate. Real-time monitoring allows for immediate action to be taken in case of suspicious transactions.

7. **Continuous Improvement**: Credit card fraud patterns evolve over time, so it is important to continuously update and improve the fraud detection models. Regularly retraining the models with new data and monitoring their performance helps in adapting to changing fraud patterns and maintaining high accuracy.

Work Flow


Credit Card Data
Data pre processing
Data Analysis
Train Test split
Logistic Regression Model
Evaluation

6. CONCLUSION

In conclusion, Credit card fraud detection is a critical task in today's digital payment landscape. The proposed work for credit card fraud detection involves several key steps to effectively identify and prevent fraudulent transactions.

The process begins with data preprocessing, which includes handling missing values, scaling features, and addressing class imbalance. Feature selection techniques are then applied to identify the most informative features for fraud detection.

Next, a suitable machine learning model is selected, such as decision trees, random forests, logistic regression, or neural networks. The model is trained on the preprocessed dataset and evaluated using metrics like accuracy, precision, recall, and F1-score.

To optimize the model's performance, hyperparameter tuning techniques like grid search or random search can be employed. The optimized model is then deployed in a real-time environment to monitor credit card transactions and classify them as fraudulent or legitimate.

Continuous improvement is crucial in credit card fraud detection, as fraud patterns evolve. Regularly updating and retraining the models with new data helps in adapting to changing fraud patterns and maintaining high accuracy.

It is important to note that the specific implementation details may vary depending on the dataset, available resources, and chosen machine learning algorithms. However, the proposed approach provides a general framework for credit card fraud detection.

7. REFERENCES

1. "MS Windows NT kernel description," creditcards.com, accessed: 2010- 09-30.

2. D. Excell, "Bayesian inference–the future of online fraud protection," *Computer Fraud & Security*, vol. 2012, no. 2, pp. 8–11, 2012.

3. M. Zareapoor, P. Shamsolmoali, *et al.*, "Application of credit card fraud detection: Based on bagging ensemble classifier," *Procedia computer science*, vol. 48, no. 2015, pp. 679–685, 2015.

4. R. Patidar, L. Sharma, *et al.*, "Credit card fraud detection using neural network," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 1, no. 32-38, 2011.

5. P. H. Tran, K. P. Tran, T. T. Huong, C. Heuchenne, P. HienTran, and T. M. H. Le, "Real-time data-driven approaches for credit card fraud detection," in *Proceedings of the 2018 International Conference on E-Business and Applications*. ACM, 2018, pp. 6–9.

Submitted By –

Name	Sign
1. Samarpit Nandanwar	…….
2. Adib Khan	…….
3. Deven Malekar	…….
4. Himanshu Kadu	…….
5. Suraj Pawar	…….

Guide

Prof. P. D. Kaware

Head

Department of Computer Science & Engineering