CREDIT CARD FRAUD DETECTION

PROJECT ID (PCSE25-67)

PROJECT SYNOPSIS

OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY

CSE

SUBMITTED BY

MAYANK MISHRA [Roll No:2100290100095]

YASH KUMAR JHUNJHUNWALA[Roll No:2100290100198]

VANSHIKA MITTAL [Roll No:2100290100182]

Project Guide

Prof./Dr. NEHA YADAV

KIET Group of Institutions, Delhi-NCR,
Ghaziabad (UP)
Department of Computer Science and Engineering

October 2023

Title: Credit Card Fraud Detection System

INTRODUCTION

In the contemporary landscape of digital transactions, ensuring the security and integrity of financial operations is paramount. The development of robust credit card fraud detection systems has become a critical necessity in the face of increasingly sophisticated fraudulent activities. This project aims to leverage advanced data processing techniques and machine learning algorithms to detect and prevent fraudulent credit card transactions in real-time, thereby safeguarding the financial interests of both consumers and financial institutions.

Technology Used:

This project will harness the power of cutting-edge technologies such as Python, R, and various machine learning libraries including scikit-learn, TensorFlow, and Keras. The software will be developed using a combination of advanced data processing techniques, statistical analysis, and machine learning models to effectively identify fraudulent patterns and anomalies within credit card transaction data.

Field of Project:

The project resides at the intersection of financial technology and data science, integrating principles from both domains to create a comprehensive solution for credit card fraud detection. By employing sophisticated algorithms and data analysis methodologies, the project seeks to contribute to the ongoing efforts to enhance the security of digital financial transactions.

Special Technical Terms:

This project will involve the utilization of specialized technical terms such as anomaly detection, supervised and unsupervised learning, feature engineering, neural networks, precision, recall, F1 score, and model interpretability. Additionally, terms related to credit card fraud detection, such as chargeback, skimming, phishing, and identity theft, will be central to understanding the context and significance of the project's objectives.

Through the implementation of these advanced technologies and the utilization of specialized terms, this project aims to establish an effective and robust credit card fraud detection system that can adapt to evolving fraudulent techniques and ensure the continued security and trustworthiness of electronic financial transactions.

OBJECTIVE

- Financial Security: Credit card fraud poses a significant threat to the financial security of individuals and organizations. Detecting and preventing fraudulent transactions helps safeguard the hard-earned money of consumers and prevents potential financial losses for banks and other financial institutions.

- Consumer Trust: Maintaining consumer trust is crucial in the financial sector. Effective fraud detection systems reassure customers that their financial transactions are secure, fostering a sense of confidence in the use of credit cards and digital payment methods.

- Reduced Liabilities: Implementing robust fraud detection mechanisms can significantly reduce the liabilities of financial institutions. By identifying and preventing fraudulent activities promptly, banks can minimize the financial losses associated with unauthorized transactions and fraudulent activities.

- Compliance and Regulations: Compliance with industry regulations and standards is imperative for financial institutions. A comprehensive fraud detection system helps institutions adhere to regulatory requirements and maintain transparency in their financial operations.

- Mitigating Risks: Fraud detection systems play a crucial role in mitigating various risks associated with financial transactions. By identifying potential fraudulent activities in real time, these systems help mitigate the risks of identity theft, unauthorized transactions, and other fraudulent practices.

- Preventing Reputation Damage: Instances of credit card fraud can significantly damage the reputation of financial institutions. Effective detection and prevention mechanisms help in preventing reputation damage, ensuring that customers perceive the institution as trustworthy and reliable.

- Technological Advancements: With the advancement of technology, fraudsters continuously develop new and sophisticated techniques to bypass traditional security measures. The rationale for credit card fraud detection is to stay ahead of these fraudulent activities by leveraging advanced technologies and data analytics to detect anomalies and patterns indicative of fraudulent behavior.

Literature Review

A Research Paper on Credit Card Fraud Detection

BORA MEHAR SRI SATYA TEJA1, BOOMIREDDY MUNENDRA2, Mr. S. GOKULKRISHNAN3

03 MAR 2022

This research paper focuses on credit card fraud detection using different machine learning algorithms. The authors highlight the increasing threat of online frauds due to the rise in online payment modes. They aim to categorize transactions into different groups and apply various machine learning algorithms to identify fraud transactions. The dataset used in the study consists of online transactions made using credit cards, with a large portion of genuine transactions and a small number of fraud transactions.The proposed system utilizes machine learning algorithms such as decision trees and random forests to predict fraud transactions. The authors evaluate the performance of these algorithms by calculating accuracy scores and creating confusion matrices. They also address the issue of class imbalance in the dataset and propose oversampling as a method to resolve it.

A machine learning based credit card fraud detection using the GA algorithm for feature selection

Emmanuel Ileberi1*, Yanxia Sun1 and Zenghui Wang2

2022

The article titled "A machine learning based credit card fraud detection using the GA algorithm for feature selection" proposes a credit card fraud detection engine that utilizes machine learning and the genetic algorithm (GA) for feature selection. The authors aim to improve upon existing systems by addressing the challenges of class imbalance and feature selection.The research utilizes a dataset of credit card transactions made by European cardholders over a two-day period in September 2013. The dataset contains 30 numerical features, including time, amount, and anonymous variables (V1 to V28) for data security reasons. The dataset is highly imbalanced, with only 0.172% of transactions being fraudulent.To address the class imbalance, the researchers apply the Synthetic Minority Oversampling Technique (SMOTE) in the data preprocessing phase. They then use the GA algorithm to select the most optimal features for fraud detection. The GA algorithm is chosen for its ability to handle a large number of input variables and handle missing values.

LITERATRATURE REVIEW 3

A good amount of research work has been undertaken in the field of imbalance of class issues [2],[4],[6], where a number of algorithms involving ML and data analysis and mining, were used to reduce the affect of imbalance. Yet, it remains to be a huge challenge. Secondly, face recognition is also a tough spot. It is highly specific and a tough case of object recognition [8]. Maintaining accuracy of this model is also challenging due to noise, camera distortion, etc. Developing a model for the same is heavy computationally. With newer methodologies like OpenCV, camshaft algorithm, pillow and face_recognition library coming up the problem of lower accuracy still exists [7]. Face recognition has a great upcoming advantage in terms of verification processes and has therefore become an active topic for research purposes [9-11]. [1] focuses on imbalance problem and conducts rigorous experimental study on various classifiers and sampling techniques. [2] shows a very creative method to detect fraud, in skewed data distributions using Minority class sampling. The process works by bringing the C4.5 DT algorithm, backpropagation (BP), and NB together. [3] Highly sophisticated methods like false - positive, for bettering the rates of detection and driving the point of identification are proposed as an instrument to measure the fraud detecting and preventing strategies' performances.

LITERATURE REVIEW 4

analyses a transaction behavior and makes predictions. Discusses ML methodologies to detect frauds in credit card and reduce manual labor. Techniques used include Bayes' belief and NN. Main idea is to input the algorithm with some features and train it on those inputs. [5] analyses various machine learning algorithms, like LR, NB, RF and MLP to find out if the classifier is suited for detecting frauds depending on the given dataset. Compares suitability for the above stated algorithms for fraud detection in credit card. [6] aims to provide improved and advantageous features for saving processing time and memory and time cost. The paper focuses

on regression methodologies to detect fraud. Also, identifies the different types of fraud.
[7] proposes an approach leading to detection of outliers in given dataset by using a scoring mechanism for each point in data while forming clusters. A methodology that showcases great and high potential for identifying outliers and pure inliers. This detection of outlying data greatly helps. Also, it was observed that Precision-Recall curves showcased better performance analysis compared to ROC curves. [8] recognizes whether a new transaction is a fraud or not. Focuses on analyzing and pre-processing data as well as the use of detection by multiple anomaly algorithms. Proposed the use of the latest ML algorithms to find outliers. [9] the paper forcused on discussing ways of setting up facial recognition model using CNN algorithms and how things can be improved in the same. Focus on the DeepFace implementation developed in Facebook's AI department.

FEASIBILITY

- Technical Feasibility: Evaluate whether the required technology, including hardware, software, and data processing capabilities, is readily available or can be feasibly developed within the project constraints. Consider whether the existing infrastructure can support the implementation of the fraud detection system and the integration of advanced data processing and machine learning algorithms.

- Financial Feasibility: Analyze the costs associated with developing, implementing, and maintaining the credit card fraud detection system. Consider the costs of acquiring necessary technologies, hiring skilled professionals, and implementing security measures. Assess whether the potential benefits of fraud prevention outweigh the initial and ongoing investment required for the project.

- Legal and Compliance Feasibility: Examine the legal and regulatory requirements related to the implementation of a credit card fraud detection system. Ensure that the system complies with industry standards, data protection regulations, and privacy laws. Assess the potential legal implications and ensure that the system aligns with all relevant regulations and requirements.

- Operational Feasibility: Evaluate whether the organization has the necessary resources, expertise, and capacity to operate and maintain the fraud detection system effectively. Assess the impact of the system on current operations and processes, and identify any potential challenges or requirements for restructuring existing workflows to accommodate the new system.

- Security Feasibility: Assess the security measures required to protect sensitive financial data and ensure the confidentiality, integrity, and availability of the system. Evaluate the potential risks and vulnerabilities associated with implementing the system, and develop a comprehensive security plan to mitigate these risks effectively.

- Market Feasibility: Conduct a market analysis to determine the demand for a credit card fraud detection system within the financial industry. Identify potential competitors, understand market trends, and assess the system's potential to meet the evolving needs of the market. Determine the potential market share and the expected return on investment for the project.

- Risk Analysis: Identify and assess potential risks and challenges that could impact the success of the credit card fraud detection project. Develop a risk management plan that outlines strategies to mitigate and address these risks effectively, ensuring the successful implementation and operation of the system.

METHODOLOGY

Data Collection:

Obtain a comprehensive dataset of credit card transactions, including both legitimate and fraudulent transactions, from reliable sources.Ensure that the dataset is representative of various transaction types, including different transaction amounts, locations, and times.

Data Preprocessing:

Clean the dataset by handling missing values, outliers, and inconsistencies.Normalize or standardize the data to bring all features to a common scale, facilitating effective analysis.Handle imbalanced data by employing techniques such as oversampling, undersampling, or generating synthetic data points.

Feature Engineering:

Extract relevant features such as transaction frequency, amount, geographical location, time of day, and transaction type.Create additional features based on domain knowledge or insights derived from the data to improve the discriminatory power of the model.

Model Selection:

Choose appropriate machine learning models, such as logistic regression, decision trees, random forests, support vector machines, or deep learning models.Consider the specific characteristics of the dataset and the problem at hand when selecting the most suitable model for fraud detection.

Model Training and Validation:

Split the dataset into training and validation sets to train the selected model.Utilize appropriate cross-validation techniques to assess the model's performance and ensure that it generalizes well to unseen data.

Hyperparameter Tuning:

Fine-tune the model's hyperparameters using techniques like grid search, random search, or Bayesian optimization to optimize its performance.Optimize the hyperparameters to achieve the best possible trade-off between bias and variance in the model.

Facilities required for the proposed work

Facilities required for credit card fraud detection include a robust computing infrastructure capable of handling large-scale data processing and complex machine learning algorithms. This entails access to high-performance servers or cloud computing resources with sufficient processing power, memory, and storage capacity to accommodate the data-intensive nature of fraud detection tasks. Additionally, specialized software tools for data preprocessing, feature engineering, and model development are essential. This may include programming languages such as Python or R, along with relevant libraries for machine learning and data analysis, such as scikit-learn, TensorFlow, or Keras. Furthermore, access to secure data storage facilities and encryption mechanisms is crucial to ensure the confidentiality and integrity of sensitive financial information throughout the development and deployment phases.

SIGNIFICANCE

Credit card fraud poses a significant financial risk to both cardholders and financial institutions. A robust fraud detection system can mitigate these risks, protect cardholders from financial losses, and enhance the trustworthiness of digital payment systems.

Expected Outcomes

A functioning credit card fraud detection system capable of real-time monitoring.
High accuracy in detecting fraudulent transactions while minimizing false positives.
A user interface that allows authorized personnel to review and manage flagged transactions.
Improved security measures to safeguard cardholder data.

REFERENCES

- A machine learning based credit card fraud detection using the GA algorithm for feature selection | Journal of Big Data | Full Text (springeropen.com)
- IRJET- International Research Journal of Engineering and Technology
- https://developer.mozilla.org/en-US/docs/Glossary/Algorithm