ACROPOLIS INSTITUTE OF TECHNOLOGY AND RESEARCH

Department of Information Technology

Synopsis
On

Intelligent Fraud Detection System Using Machine Learning

1. Introduction

1.1 Overview

Fraudulent activities in bank transactions are an increasing concern in our quickly changing digital world, offering serious risks to both financial institutions and their clients. Traditional rule-based systems frequently struggle to recognize and stop these more complex fraud schemes. Machine learning has become a powerful technique to handle this problem, with logistic regression at its core. This study seeks to investigate the use of logistic regression in the context of detecting bank transaction fraud while highlighting its benefits, drawbacks, and practical use in the financial industry.

1.2 Purpose

The project seeks to create a reliable and accurate fraud detection model by utilizing logistic regression and sophisticated data analysis techniques. In comparison to the current system, this model is anticipated to show enhanced performance metrics, such as greater accuracy, precision, recall, and a lower false positive rate. Through feature importance analysis, the research also plans to discover essential characteristics and signs of fraud, giving valuable information for upcoming fraud detection techniques. The successful reduction of the class imbalance issue along with moral and legal observance is anticipated to guarantee that the fraud detection system complies with regulatory requirements and client privacy. As a result of the project, the financial institution should be better equipped to proactively identify and stop fraudulent transactions, safeguarding both the institution and its clients while promoting trust in banking transactions in a developing digital environment.

2. Literature Survey

2.1 Existing Problem

The study of the existing system in the context of a project on fraud detection using logistic regression in bank transactions involves a comprehensive examination of the current processes, technologies, and practices related to fraud detection within financial institutions.

Existing Systems and Technologies:

- **Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective**
  Authors: Samaneh Sorournejad, Zahra Zojaji, Amir Hassan Monadjemi.

  In this paper, after investigating difficulties of credit card fraud detection, we seek to review the state of the art in credit card fraud detection techniques, datasets and evaluation criteria.

  Disadvantages:
  • Lack of standard metrics

- **Detection of credit card fraud: State of art**
  Authors: Imane Sadgali, Nawal Sael, Faouzia Benabbau

  In this paper, we propose a state of the art on various techniques of credit card fraud detection. The purpose of this study is to give a review of implemented techniques for credit card fraud detection, analyses their incomes and limitless, and synthesize the finding in order to identify the techniques and methods that give the best results so far.

  Disadvantages:
  • Lack of adaptability

- **Credit card fraud detection using machine learning algorithm**
  Authors: Vaishnavi Nath Dornadulaa, Geetha S.

  The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyze the past transaction details of the customers and extract the behavioral patterns.

  Disadvantages
  • Imbalanced Data

2.2 Proposed Solution

Bank transaction fraud has grown to be a serious problem, endangering both financial institutions and their clients. By utilizing the capabilities of logistic regression, a machine learning technique, this research seeks to increase the effectiveness and accuracy of fraud detection within the financial industry. We will find areas for improvement in the present fraud detection methods by doing a thorough analysis of the current system. The development and application of a logistic regression-based fraud detection model will follow, and it will address challenges with feature engineering, class imbalance, and data preprocessing.

- **Data Preparation:** Collect, clean, and preprocess historical transaction data to create a high-quality dataset for model training. This includes feature engineering to extract meaningful information from raw data
- **Logistic Regression Model Development:** Build a logistic regression-based machine learning model to classify transactions as either legitimate or fraudulent. Train the model using historical data with labeled fraud cases.
- **Feature Importance Analysis:** Identify the most influential features contributing tofraud detection, providing insights into the key indicators of fraudulent activities.
- **Model Evaluation:** Evaluate the performance of the logistic regression model using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, ensuring its effectiveness in detecting fraudulent transactions.
- **Class Imbalance Mitigation:** Implement strategies to address the class imbalance problem inherent in fraud detection tasks, ensuring that rare fraudulent cases are not overlooked.

Planning of the Project work

To build a project like "Fraudulent Detection Using Machine Learning Algorithm " a well-structured methodology is essential. This project should primarily be developed using Python, a versatile programming language well-suited for machine learning and deep learning tasks.

Project Initiation Phase:

- Define project objectives, scope, and deliverables.
- Identify project stakeholders, including the project team, financial institution representatives, and data sources.
- Establish a project timeline and budget.
- Develop a project charter outlining the project's purpose, goals, and initial plan.

Data Collection and Preprocessing:

- Collect historical transaction data from various sources, ensuring data quality and completeness.
- Preprocess the data by cleaning, handling missing values, and transforming it into a format suitable for analysis.
- Perform feature engineering to extract relevant features for fraud detection.

Logistic Regression Model Development:

- Split the dataset into training and testing subsets.
- Develop a logistic regression-based machine learning model for fraud detection.
- Tune hyperparameters and optimize the model's performance using techniques such as cross-validation.

Feature Importance Analysis:

- Conduct feature importance analysis to identify the most influential variables for fraud detection.
- Visualize and interpret the results to gain insights into fraud indicators.

3. Theoretical Analysis

3.1 Software Designing

Technology Used:

Programming Language:
- Python: Python is a popular choice for machine learning and deep learning tasks. It offers a rich ecosystem of libraries and frameworks that are well-suited for image recognition, data preprocessing, and model development.

Data Management and Processing:
- Pandas: Pandas is a Python library used for data manipulation and analysis. It can be beneficial for handling datasets and preprocessing.
- NumPy: NumPy is another Python library for numerical computing. It's useful for performing operations on multidimensional arrays, which are common in image data.

4. Applications

Banking and Financial Services

- Detecting unusual patterns in banking transactions, such as large withdrawals or frequent transfers from multiple accounts, to prevent potential fraud.
- Supporting fraud detection across multiple channels (e.g., online banking, ATMs, in-branch transactions), enhancing the security of customers' accounts.
- Reducing the risk of financial loss for both banks and customers by identifying fraudulent credit and debit card transactions in real time.

E-commerce Platforms

- Identifying suspicious payment behaviour, such as repeated failed transactions, high-value purchases, or purchases from multiple accounts, which may indicate potential fraud.
- Ensuring that only legitimate transactions go through, improving user trust and platform security, especially during sales or high-traffic shopping periods.
- Preventing chargebacks and financial losses due to fraudulent transactions, which can protect the business's reputation and revenue.

Insurance Claims Processing

- Detecting fraudulent insurance claims by analysing transaction and claim patterns to identify anomalies.
- Enhancing the security of claim approvals, reducing payout errors, and minimizing the occurrence of false claims that impact insurance providers financially.

Healthcare Payments and Insurance Claims

- Analyzing healthcare claims for patterns of fraud, waste, or abuse.
- Identifying suspicious billing practices or patient behaviour that could indicate fraudulent insurance claims, thereby reducing unnecessary payouts.

REFERENCES

- https://www.itransition.com/machine-learning/fraud-detection
- https://www.tutorialspoint.com/scikit_learn/index.htm
- https://en.wikipedia.org/wiki/Data_analysis_for_fraud_detection
- https://en.wikipedia.org/wiki/Credit_card_fraud

Guided By:

Prof. Ankita Agrawal

Group Members:

PriyanshuYadav(0827IT211091)

Shubhra Jain (0827IT211109)

Shyamsundar (0827IT211111)

Yashika Rahinj(0827IT211134)