Credit Card Fraud Detection Using Machine Learning

This presentation details a machine learning approach to detect fraudulent credit card transactions. With the rise of digital payments, fraud detection has become a critical area. This project employs data preprocessing, analysis, and machine learning models to identify and flag suspicious transactions, offering a robust solution to combat credit card fraud.

by Sankari C

Project Goals and Workflow

The primary goal of this project is to accurately detect fraudulent transactions within a credit card dataset. Due to the imbalanced nature of the dataset, the Under Sampling method was employed to address this challenge.

1

Data Preprocessing

Cleaning and formatting the data for analysis.

2

Data Analysis

Exploring the data to identify patterns and anomalies.

3

Train/Test Split

Dividing the data for model training and validation.

4

Model Training

Training the various models and compare the prediction.

5

Evaluation

Assessing the model's performance.

Data Preprocessing Steps

Data preprocessing is a crucial step in preparing the dataset for analysis and model training. This involves loading the dataset, handling time-based data, converting features to numeric representations, and understanding the distribution of normal and fraudulent transactions.

1 Loading Data

Loading the dataset from a CSV file into a Pandas dataframe.

2 Feature Conversion

Converting features to numeric for security purposes.

3 Missing Values

Checking for and handling any missing values in the dataset.

Addressing Data Imbalance

The dataset exhibits a significant imbalance between normal and fraudulent transactions. If we create the ML model directly from it, then it can't recognize the fraudulent data because it is too small. To mitigate this, an Under Sampling technique is applied to create a more balanced dataset.

Normal Transactions

Analyzing statistical measures for normal transactions.

Fraudulent Transactions

Analyzing statistical measures for fraudulent transactions and comparing them to normal transactions.

Under Sampling Methodology

Under Sampling involves creating a balanced dataset by reducing the number of instances in the majority class (normal transactions) to match the minority class (fraudulent transactions). This helps prevent the model from being biased towards the majority class.

1. **Random Sampling**

   Take the random sample of the data.

2. **Building Dataset**

   Build a sample dataset containing similar distribution of normal(fraud) distributions.

3. **Concatenation**

   Concatenating the two dataframes.

4. **Verification**

   Cross verify the values in two distributions.

Model Training and Selection

After preparing the dataset, the next step is to define the feature and target variables, split the data into training and testing sets, and train the machine learning models. The following models were used:

Logistic Regression

A linear model for binary classification.

SVC

Support Vector Classification, a powerful and versatile classification model.

Decision Tree

A tree-based model for classification.

Evaluation and Results

The trained models are evaluated using the testing data to assess their performance. Accuracy score is a key metric used to measure the effectiveness of the models in detecting fraudulent transactions.

0.974

Accuracy

The accuracy score for testing data using Logistic Regression.

0.974

Accuracy

The accuracy score for testing data using SVC.

0.964

Accuracy

The accuracy score for testing data using Decision Tree.

Key Takeaways and Next Steps

The project demonstrates the effectiveness of machine learning in detecting credit card fraud. The Under Sampling method helps address the issue of imbalanced data, and various models can be used to achieve high accuracy. Future work could explore more advanced models and feature engineering techniques.

- Machine learning is effective for fraud detection.
- Data imbalance can be addressed using Under Sampling.
- Further improvements can be achieved through advanced techniques.