FIND DEFAULT: PREDICTION OF CREDIT CARD FRAUD

Introduction

Project Overview

Credit card fraud poses a significant threat to financial institutions and their customers. Detecting fraudulent transactions swiftly and accurately is crucial in mitigating these risks. This project, titled "Find Default," aims to develop a predictive model to identify potential credit card fraud cases using machine learning techniques. The goal is to enhance the efficiency and accuracy of fraud detection, thereby reducing financial losses and improving customer trust.

Objective

The primary objective of this project is to create a reliable model that can predict fraudulent credit card transactions. This involves:

- Analysing and preprocessing the provided dataset
- Building various machine learning models
- Evaluating the performance of these models
- Selecting the best-performing model for deployment

Data Analysis

Data Description

The dataset used for this project consists of credit card transactions made over a specific period. The dataset includes features such as transaction amount, time, and several anonymized variables representing other transaction details. The target variable indicates whether a transaction is fraudulent (1) or not (0).

Data Preprocessing

Data preprocessing is a critical step to ensure the accuracy and efficiency of the machine learning models. The steps taken in this project include:

- **Handling Missing Values**: Checking for and addressing any missing values in the dataset.
- **Data Normalization**: Scaling the features to ensure uniformity and improve model performance.

- **Class Imbalance**: Addressing the imbalance between fraudulent and non-fraudulent transactions using techniques such as SMOTE (Synthetic Minority Over-sampling Technique).

Methodology

Model Selection

Several machine learning algorithms were considered for this project, including:

- Logistic Regression
- Decision Tree
- XGBoost
- Support Vector Machine (SVM)

Model Tuning

Hyperparameter tuning was performed to optimize the performance of the selected models. Techniques such as Grid Search and Random Search were employed to find the best combination of parameters for each model.

Evaluation Metrics

The models were evaluated based on the following metrics:

- **Accuracy**: The proportion of correctly classified transactions.
- **Precision**: The ratio of true positive predictions to the total predicted positives.
- **Recall**: The ratio of true positive predictions to the total actual positives.
- **F1 Score**: The harmonic mean of precision and recall, providing a balanced measure of model performance.

Results

Model Performance

The performance of each model was evaluated, and the results were compared to identify the best-performing model. The XGBoost demonstrated superior performance with a high accuracy, precision, recall, and F1 score.

Feature Importance

The feature importance analysis revealed that certain variables had a more significant impact on predicting fraudulent transactions. These insights can be valuable for further enhancing fraud detection strategies.

Conclusion

Summary

The Find Default project successfully developed a predictive model for credit card fraud detection. Through rigorous data analysis, preprocessing, model selection, and tuning, the XGboost classifier emerged as the best model, demonstrating high accuracy and robustness in identifying fraudulent transactions.

Future Work

Future improvements to this project could include:

- Incorporating additional features to capture more transaction details.
- Exploring advanced techniques such as deep learning for potentially better performance.
- Continuously updating the model with new data to maintain its accuracy and relevance.

Impact

Implementing this model can significantly enhance the ability of financial institutions to detect and prevent credit card fraud, thereby protecting customers and reducing financial losses

---

THANK YOU!