Find Default (Prediction of Credit Card Fraud)

Introduction

Credit card fraud is a significant concern for both credit card companies and consumers. In this project, we aim to develop a classification model to predict whether a transaction is fraudulent or not. The dataset used contains transactions made by European cardholders in September 2013. This report outlines the steps taken in developing and evaluating the model, as well as discussing future work.

Methodology

Exploratory Data Analysis (EDA)

- Conducted data quality checks, treated missing values, and outliers.
- Visualized data to identify patterns, relationships, and trends.

Data Cleaning

- There are no missing values so data cleaning was not required.

Dealing with Imbalanced Data

- Employed Synthetic Minority Oversampling methods to balance the highly imbalanced dataset.

Feature Engineering

- Created new features and transformed existing ones to enhance model performance.

Model Selection

- Selected the most appropriate model for the project based on its performance.

Model Training

- Split the data into train and test sets and estimated the best model parameters using the train set.

Model Validation

- Evaluated the model's performance on unseen data to assess its ability to generalize and identify any issues such as overfitting.

Model Deployment

- Developed a plan for deploying the trained model in a production environment.

Results and Performance Evaluation

Model Performance

- Achieved an accuracy of >75% on the test dataset.
- Implemented methods for hyperparameter tuning.
- Conducted model validation to ensure robustness.

Discussion of Future Work

- Addressed limitations of the current model.
- Explored avenues for improving model performance, such as incorporating additional features or experimenting with different algorithms.

Source Code

The source code used to create the pipeline is included in the attached zip file.

Conclusion

In conclusion, the developed model demonstrates promising performance in predicting credit card fraud. However, there are opportunities for further enhancement and refinement. By addressing the identified areas for improvement and continuing to iterate on the model, we can better mitigate the risks associated with fraudulent transactions.