**Find Default: Prediction of Credit Card Fraud**

**Introduction**

Credit card fraud poses a significant threat to both credit card companies and consumers. This project aims to develop a classification model to predict fraudulent transactions. The dataset comprises transactions made by European cardholders in September 2013. This report outlines the steps taken to develop and evaluate the model and discusses future work.

**Methodology**

Exploratory Data Analysis (EDA)

Conducted data quality checks, treated missing values, and handled outliers.

Visualized data to identify patterns, relationships, and trends.

Data Cleaning

Confirmed no missing values were present, hence no data cleaning was required.

Dealing with Imbalanced Data

Employed Synthetic Minority Oversampling Technique (SMOTE) 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 based on its performance.

Model Training

Split the data into training and testing sets.

Estimated the best model parameters using the training set.

Model Validation

Evaluated the model's performance on unseen data to assess its ability to generalize.

Identified and addressed 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 over 75% on the test dataset.

Implemented methods for hyper parameter tuning.

Conducted extensive model validation to ensure robustness.

Discussion of Future Work

Address limitations of the current model.

Explore 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**

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