

Machine Learning for Financial Risk Management

Introduction

The financial sector faces significant challenges related to loan defaults and fraudulent activities. Traditional risk assessment methods often fail to accurately predict these events, leading to substantial financial losses. This project aims to explore machine learning decision tree models, specifically XGBoost, for predicting loan defaults and adapting these models for fraud detection and other financial risk management applications.

Objectives

The primary objectives of this project are:

- Assess the performance of XGBoost in predicting loan defaults.
- Adapt the trained model for detecting fraudulent activities within financial datasets.
- Analyze the effectiveness of machine learning techniques in enhancing financial risk management practices.

Available Resources/Datasets

We will utilize publicly available datasets relevant to loan defaults and fraud detection, including:

- Credit card transaction datasets with labeled fraudulent transactions.
- Historical loan data containing borrower information, payment history, and default status.

These datasets will provide the necessary information for training and evaluating our machine learning models.

Methodology

The methodology for this project includes several key steps:

1. **Literature Review:** Examine existing research on machine learning applications in financial risk management.
2. **Data Preprocessing:** Implement data cleaning processes such as sanity checks, data exclusions, and anomaly detection.
3. **Feature Engineering:** Utilize techniques including SHAP values for interpretation and PCA for dimensionality reduction.
4. **Model Development:** Implement XGBoost and compare its performance with Random Forest algorithms. Hyperparameter tuning will be conducted based on metrics like AUC (Area Under the Curve) and GINI index.
5. **Evaluation Metrics:** Assess model performance using precision, recall, F1-score, and ROC AUC.

Expected Outcomes

The intended final output of this project includes:

- A comprehensive report detailing methodology, findings, and implications of using machine learning models in financial risk management.
- A presentation summarizing key insights from the project for discussion with faculty and peers.
- The development of a predictive model that can effectively identify potential loan defaults and fraudulent transactions.