Professor Fadi Alsaleem David, Amidou, & Vanessa

DSC 450 January 19th, 2025

**LOAN DEFAULT PREDICTION APP**

Loan Default Prediction App is a machine learning driven app that will show you whether a borrower will default or not. Users enter financial information like credit score, loan amount, income, work status, debt-to-income ratio, default history. The app then uses an XGBoost model to make the prediction which gives a categorical result (likely/unlikely to default) and a probability score representing the default likelihood.

We chose XGBoost because of the better performance on structured/tabular data, the fact that it supports imbalanced data, and the feature importance feature is built-in. The final model after a lot of model tuning and validation achieved AUC-ROC score of 0.94 with accuracy of 84% that can be used to predict loan defaults. feature importance function named credit score, previous defaults and debt-to-income ratio as the most important predictors.

In order to make the model interpretable we added SHAP (SHapley Additive Explanations), a powerful explainability tool which displays feature effect on predictions. The SHAP summary plot shows you how each feature impacts the model decisions. This allows lenders and borrowers to see why one prediction was made. Also, SHAP dependence plots can be used to probe features e.g., how a higher number of defaults in the past increases default risk.

There was a prescribed design: data preprocessing, feature engineering, hyperparameter calibration, model evaluation. The app was then designed using Dash which gives an easy-to-use UI to make predictions in real time. In testing, the scaling algorithm was tuned to tinker with probability outputs so that features such as Past Defaults were well weighted in the final prediction.

Looking forward, further enhancements include:

Database Integration – Stores the user feedback and forecasts for further processing and evolution of model.

Deployment to the Cloud – Enabling the app for practical financial use.

API integration in real time – Allows the model to be live-financial updates in real time for predictive updates.

Probability Calibration Optimization – Maintaining high-quality risk calculation with constant model tuning.

The main takeaway of this project is that feature synchronization and real-world scaling is very important for the financial modelling. The SHAP integration was a step towards transparency of the model, but there’s still a long way to go before it reaches that perfect usability/predictive accuracy threshold. The Loan Default Prediction App is an excellent start to actual credit risk analysis, that could also find further use in banking and financial technology (FinTech) verticals.

**Work Cited**

Fuentes, A. (2018). *Hands-On Predictive Analytics with Python: Master the complete predictive analytics process, from problem definition to model deployment*. Packt Publishing.

Sparsh, I. (n.d.). *Lending Club Loan Dataset (2007-2011)* [Data set]. Kaggle. Retrieved December 22, 2024, from <https://www.kaggle.com/datasets/imsparsh/lending-club-loan-dataset-2007-2011>