Predicting Credit Risk and Building a Credit Risk Score to detect Fraud

Chamika Hapuarachchi[1], Purani Annasiwatte[2]

1 University of Sri Jayawardenapura,196/1, Orutota, Gampaha,1100, hchamikadilshan@gmail.com *
2 University of Sri Jayewardenepura, 17, Kandahenawatte Road, Depanama, Pannipitiya, 10230, annasiwatte55@gmail.com

The financial industry faces significant risks associated with lending, particularly in the form of fraudulent activities. Fraud occurs when individuals exploit lending systems to secure credit with no intention of repayment, resulting in substantial losses for financial institutions. A key indicator of such fraudulent behaviour is the charge-off, where loans are written off as uncollectible. This study introduces a predictive modelling approach designed to assess credit risk and identify potential fraud by assigning a credit risk score to customers. Through an extensive exploratory data analysis (EDA), we explored the relationships between various predictor variables and the target variable, Charge-Off Status. Key factors influencing charge-off risk were identified, including FICO Score, Debt-to-Income Ratio, Number of Delinquent Accounts, and Number of Credit Applicants. The analysis revealed challenges such as missing data and class imbalance, which were addressed by applying data imputation techniques and resampling methods to balance the distribution of charge-off and non-charge-off customers. Two predictive models Logistic Regression and XGBoost were assessed, with performance metrics indicating comparable results. Given the importance of model transparency, the Logistic Regression model was ultimately chosen. This model offers interpretability in its predictions. Further studies will focus on refining the model by implementing advanced imputation strategies and further addressing class imbalance to enhance the model's predictive accuracy and reliability in fraud detection.

Keywords: Logistic Regression, XGBoost, Machine Learning Explainability, Predictive Modelling