Title: Modelling Credit Default Using Open Banking and Psychometric Data; a Machine Learning Approach

Abstract:

Financial services providers often struggle to accurately model credit risk in thin-file customers. Our analyses have shown that 28% of accepted loan applicants from a single lender went into payment arrears, demonstrating the challenge of accurately predicting repayment outcomes in all customers. It is therefore important to explore alternative data sources that may help to model credit risk. Previous work has demonstrated the mental wellbeing of a loan applicant can influence repayment. For example, those with poor mental health are three times more likely to be behind on at least one key payment. In addition, Open Banking data can provide real-time metrics of how an individual manages their money. We therefore applied machine learning methods to Open Banking and psychometric data to create a model that provides additional information in the credit decisioning process.

To develop an alternative credit risk model, Open Banking and psychometric data were collected from applicants during a typical loan application. In addition, we sourced Open Banking and loan performance data from $\sim n$=5,000 applications across our commercial partners. Bank transactions were classified into spending categories and features pertaining to different aspects of financial behaviour were calculated. We then compared several machine learning approaches to ascertain which could accurately predict applicants going into payment arrears, using a stratified k-fold cross validation procedure.

Our results showed that XGBoost models returned significantly more accurate arrears predictions than L1-regularized logistic regression and GAM models according to AUC metrics (ps < .001). The model also identifies a proportion of accepted applicants with hidden risk not picked up by traditional scoring methods. To this end, we also found that models using Open-Banking features are improved with the addition of psychometric data when predicting the number of creditors an individual owes (BF = 6.81). This shows that while Open-Banking-only models can exceed existing decision accuracy benchmarks in some cases, they substantially benefit from the addition of psychometric predictors. Further results of a swapset analysis when using alternative credit risk models will be presented at the conference.

These results show the value of Open Banking and psychometric data as a source of unique information about credit repayment outcomes. We share how models utilizing

these data sources can provide a useful and scientifically-validated assessment of creditworthiness.