

Executive Summary - Bondora P2P Lending

The business case that we have analyzed is a publicly available data set of a leading European Peer 2 Peer (P2P) lending platform, Bondora. P2P lending risk has been an attractive risk factor to assess a potential borrower's behavior. P2P lending is a way to raise debt for individuals which is becoming popular for developing countries. Our data set includes P2P data from Estonia. This data is an indicator for lenders and various financial institutions to understand trends of borrowers that may lead to them defaulting on a loan. These metrics and trends help researchers for lenders analyze the risk factors most important when selecting potential borrowers. What we are predicting using this dataset is the probability that someone will default on a loan, based on if there is a default date present in the data set. We have derived our recommended model based on the predictors in the data set that we have found through visualization and correlation that help derive our categorical target of default date. Our dataset initially consisted of 112 columns with 179,235 rows; it has been reduced to 22 columns and 30,591 rows based on pre-processing techniques. Variables that were removed were based on high volume of missing values, lack of predictive power to the target variable and strong correlation between columns through data visualization. Our dataset distribution of '1' for default date at present is 8,520 and 22,071 for '0' which represents no default date.

After running all of our predictive models the model with the best performance relative to our business case was Logistic Regression. On test data, our model had ~91% total accuracy with a misclassification rate of ~9%. Our model predicted 208 false positives meaning a default date was predicted when there wasn't one present, compared to 1364 true positives of accurately predicting our target. The inputs to this model that made the biggest contributions were Age, Credit Score, Interest, Number of Dependents, Income from Employer and Principal Payments made. These were the strongest characteristics for a borrower at the start of a loan. If a borrower has a history of a late penalty balance and how frequently they are paying the principal of the loan, have a strong correlation to the potential that a borrower would default on a loan currently. Which strong variables however, are not relevant to our business case.

What we would recommend to the business from our findings, Bondora should consider greater interest rates for applicants that have dependents. As borrowers with dependents have a higher likelihood of default, there should be further due diligence on such applicants. Bondora should also consider the impact of borrower age on likelihood of default. The majority of borrowers that default are between the age of 20-40. With further dependency on the education level when analyzing age, as additionally less educated (category 1,2) borrowers have 6% higher default rate on a loan.

Bondora should positively consider potential borrowers that already have an existing mortgage, as 80% of non defaulters have an existing mortgage as these borrowers have demonstrated previous credit worthiness and not showed default behavior with another lender. Finally, Bondora can consider interest rates based on a borrower's free cash on hand, as 75% of defaulters have <\$25 and even negative free cash available. Bondora could utilize the outputs from this model to identify specific financial characteristics of future borrowers who may be at risk of defaulting and not repaying their loan on time.