

Credit card usage and default risk

Credit is an indispensable tool for individuals and corporations to achieve their financial goals. According to the federal reserve, at least 83% of adults in the US use at least one credit card. In Europe, this percentage can be as high as 68% in countries such as the UK. In order to guarantee their financial health, institutions such as banks and credit unions require tested methods to determine the credit risk associated with borrowers.

Credit risk assessment of individuals is traditionally determined using data such as age, income, education level, and FICO or Equifax credit scores. The increasing popularity of e-commerce and the large amount of bank card transaction data that can be collected and associated with clients these days suggest that a machine learning approach that incorporates this data can be useful for credit risk determination. Using data about bank card transactions and default history of other clients, we built a model to assess the credit risk of an applicant that requires only his bank card transaction history.

Using real bank data from Alfa Bank we were able to predict default on credit products with a high ROC AUC score. ROC AUC score (Area Under the Curve of the Receiving Operating Characteristic Curve) is an indicator between 0 and 1 that expresses how well the model is able to identify applicants likely to default on a product they are applying for. The closer the score is to 1, the better the model is able to separate those individuals likely to default from the rest of applicants. Our model achieved a score of 0.78 which represents a 56 percent improvement over the score of 0.50 of our baseline model. Our model also improved by 35 percent the ROC AUC score of a more elaborate second baseline model.

Alfa Bank is the largest private bank in Russia with international presence in seven other countries. The model we built can be used by leading financial institutions with significant international presence to devise lending strategies that minimize losses due to borrowers inability to pay loans back. It also provides a methodical way to make decisions about credit applicants and require only moderate computational resources. Because our model does not require comprehensive information about clients, it can be used as a quick early screening tool. This could result in reduced application processing times and an easier experience for clients.