

Small Business Borrower Profiling in Predicting Default



Summary: For my project I partnered with a company which acts as a matching platform between small businesses and prospective lenders. The setup is such that the lenders have to make their decision without access to credit scores, and loans don't have collateral. Right now, the lenders don't have enough information when deciding whether to give out the loan, and often have to act on their 'gut feel'.

Business Challenge: Currently, only 15% of all the loan applications get approved, and yet 30% of borrowers default. The company asked me to try to

identify factors, which can be used as default predictors.

Data and Modeling: The data was far from perfect. I had to roll up my sleeves and do a lot of data cleaning, as well coming up with creative ways of normalizing data, feature engineering, and combining time series with static features. I explored the Time Series, built a Logistic Regression and Random Forest models.

Best predictors: I discovered that there are some industry clusters which are much more likely to default. There was also a somewhat surprising negative relationship between default and bank account 'swings' - the larger changes in the daily bank balances (both positive and negative) were linked to the lower chance of default. I explain why this can make sense in my presentation.

Outcome: As a result, I am working with the company to implement a new model, where borrowers with a higher risk of default will have to pass more strict criteria to get accepted for the loan.



Adel Sabirova

<https://www.linkedin.com/in/adel-sabirova/>

<https://github.com/adelchik/>