

Random Survival Forests in the Credit Risk Models for Small and Medium Enterprises

The credit risk is considered to be a key risk in the banking activity. Statistical and data mining models used during the assessment process of the SMEs' credit risk are mainly based on the financial data sourced from the financial statements. However, in the case of small and medium enterprises, the qualitative factors seem to play a significant role when assessing the credit risk and this is the reason why the most frequently used ones will be discussed. The purpose of this project is to check whether the inclusion of the qualitative factors (such as the age of the company, branch, location, legal form and number of employees) improves the prediction of the credit risk model. The combination of qualitative factors and financial ratios will be presented. During the model building process, the Random Survival Forests method will be applied. The results of the model will be compared with those received by the single survival model. We expect that Random Survival Forests should provide not only better results but also more stable ones than the single survival model. The following hypotheses will be verified: (1) The inclusion of the qualitative factors in the credit risk model for SMEs increases the prediction of the model and (2) The application of the Random Survival Forests method improves the learning performance (effectiveness) of the model in comparison with a single survival model. In the project we used a data sample consisting of 864 companies, including 331 bankruptcies. The data was provided for us as a courtesy of one financial institution operating on the Polish market.