

## **Master Thesis: Predicting Credit Risk of Companies Using Qualitative Data**

In recent years, companies have been required by regulators to disclose an increasing amount of textual and numerical data that is related to their financial situation. Moreover, companies share qualitative information with (potential) investors on their business situation via social media. Besides numerical data from, e.g., balance sheets, the usage of qualitative data is particularly promising to explain and predict the riskiness of a company. For instance, Donovan et al. (2018) show that information contained in companies' conference calls can be used to explain variation in companies' credit risk approximated by their credit default swap (CDS) spreads that has not been captured by common market-based measures of credit risk.

The goal of this thesis is to design, implement, and validate a machine learning (ML) model that predicts the spread of CDS instruments by using companies' qualitative (and possibly numerical) data. A further analysis might include how well the developed ML framework is able to anticipate the future downgrading (or default) of companies. Data input sources can be, for instance, firms' conference calls, data from annual reports, social media, and news feeds. The data has to be collected by the student, but will be actively supported by the supervisor. Besides the empirical part of the thesis, the student is expected to give a comprehensive and structured literature overview regarding studies analyzing determinants of companies' credit risk.

**Supervisor:** Timo Schäfer

### **Literature:**

- Donovan, J. and Jennings, J.N. and Koharki, K. and Lee, J.A. (2018), Determining Credit Risk Using Qualitative Disclosure. Available at SSRN: <https://ssrn.com/abstract=3149945>
- James, G. and Witten, D. and Hastie, T. and Tibshirani, R. (2013), An Introduction to Statistical Learning. New York: Springer.