Predicting Company Bankruptcy

Introduction and problem formulation

The proposed project seeks to predict company bankruptcy based on data collected on bankruptcies from the Taiwan Economic Journal between 1999 and 2009.

The key question we wish to answer in the project, is whether it is possible to accurately predict whether a company will go bankrupt. Prediction of bankruptcy is crucial in the financial industry to make better credit and loan assessments. In general these decisions take up a lot of resources for financial institutions and the largest bankruptcies impacts both the business owners, also financial institutions and potentially the entire society and economy. Therefore, the value of a more accurate prediction of bankruptcies will have a very large value for several stakeholders.

These decisions have traditionally been reliant on fundamental financial theory, but the increase in the amount of available data makes the potential application of machine learning more obvious. The given data set makes it possible to see whether patterns in the financial data would make it possible to predict which of these companies are at risk of going bankrupt, which will of course also be of interest to the company itself or other stakeholders.

Some specific questions that may of interest are how we balance the ratios of false positive and false negatives in our predictions to achieve the best results that create the most valuable predictions. We are mainly interested in identifying high-risk companies, which means we must consider the value of identifying a company with a high bankruptcy risk compared to falsely categorising a company as being at the risk of bankruptcy (false positives). We will most likely also be interesting to gain insights into the most valuable predictors of bankruptcy. This may allow us to better understand what financial indicators that give the best indication of bankruptcy risk, which could add to the wider understanding of financial statement analysis.

Data and Features

The data set contains a total of 6819 records, each of which corresponds to a company in Taiwan. The bankruptcy records were obtained from Taiwan Stock Exchange in accordance with their definition of bankruptcy and official filings. For each company, 96 features are observed which cover a long list of financial information and financial ratios that are often used to perform financial valuations and credit risk assessments. This will allow us to model the risk of bankruptcy and investigate how the features help predict potential bankruptcies. The dependent variable is a binary variable indicating whether the company went bankrupt or not.

Potential approaches and methods

The data is well suited for classification methods due to the bankruptcy being a binary outcome. Several of the methods that will be covered later in the course such as support vector machines, logistic regression or classification trees could be applied to solve the problem. Besides different models, it is expected that feature engineering methods will be used.

Data source:
https://archive.ics.uci.edu/ml/datasets/Taiwanese+Bankruptcy+Prediction