CSE4062

Introduction to Data Science and Analytics

2023-2024 Spring

Group 5

Company Bankruptcy Prediction

Ahmet Onat Özalan - 150118054 - Computer Engineering
Email: o171141@gmail.com

Selin Zeydan - 150322823 - Industrial Engineering
Email: selinzeydan9@gmail.com

Mediha Ecem Polat - 150820822 - Bioengineering
Email: medihaecempolat@gmail.com

Fırat Bakıcı - 150120029 - Computer Engineering
Email: firat143@gmail.com

Kardelen Kubat - 150118056 - Computer Engineering
Email: kardelenkubatcse@gmail.com

Berfin Ege Yarba - 150321036 - Industrial Engineering
Email: berfinegeyarba@gmail.com

Project Description

In this project, we will develop a machine learning model that predicts if a company will go bankrupt or not. For this purpose, we will use a dataset titled as “Company Bankruptcy Prediction” on Kaggle.

Our dataset consists of data that is collected from Taiwan Economic Journal between the years 1999 and 2009. Company bankruptcy is defined based on the business regulations of Taiwan Stock Exchange.

Our reasons for choosing this dataset are as follows:

- It provides us an attribute (named as “Bankrupt?”) to make a classification of the data instances. We will use this attribute to classify the data instances into two groups: Will go bankrupt and won’t go bankrupt.
- Other than the attribute above, it has 95 attributes and we believe this is a good amount of attributes for this project.
- All 95 attributes are numeric attributes. We believe this is a very nice trait of this dataset.
- The subject area of the dataset is business and we think our group members from the Industrial Engineering department can provide insights during the development of the project.

Dataset Statistics

Our dataset consists of 6819 instances (rows) and 96 attributes (columns).

There is 1 target attribute that assigns class labels to the instances. As we described above, the name of the target attribute is “Bankrupt?” and it is an integer type attribute. The value of “1” means the company went bankrupt, “0” means the company didn’t go bankrupt.

The remaining 95 attributes are features. They are all numeric type features. Some of the important features are as follows:

- Operating gross margin: The gross profit of the company divided by the net sales.
- After-tax net interest rate: The net income of the company divided by the net sales.
- Operating expense rate: The operating expenses of the company divided by the net sales.
- Net worth/assets: The equity of the company divided by the total assets.