Analysis and Prediction of Bankruptcy Identification

DATASET DESCRIPTION

The Company Bankruptcy Prediction dataset [1] that we will be using was obtained from Kaggle. The data itself was obtained from the Taiwan Economic Journal during 1999 to 2009. Bankruptcy was defined by the Taiwan Stock Exchange.

The dataset includes 95 attributes and one class label (bankruptcy). Many of the attributes are ratios and percentages displayed as a decimal, however there are a few attributes that are not, such as operating profit per person. The dataset does not include any missing values.

PROJECT IDEA

The goal of this project is to predict whether the attributes of company assets are related to a potential of company bankruptcy to help companies highly like to go bankrupt. Various pre-processing techniques and classification models will be used to analyze the company asset attributes and predict whether the company will go to bankruptcy or not. By using the classification model, companies will be able to objectively analyze the factors that could contribute to the chance of going to bankruptcy and prevent the tragedy.

We are going to perform data analysis to find out the correlation between company asset attributes and chance of company bankruptcy, In addition to this, we will look for the specific attributes that could most likely influence the chance of bankruptcy by using a random forest classification model. Random forest will provide the embedded dimensionality reduction and ensure high accuracy of the classification model we use.

SOFTWARE SPECIFICATIONS

Python3, JupyterHub, Zoom, NumPy, Scikit-Learn, Pandas, Matplotlib

PAPERS TO READ

[1] “Feature selection in bankruptcy prediction” 2009.

[2] “Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study" 2016.

[4] “ Machine learning models and bankruptcy prediction” 2017

WORK DIVISION

Data Pre-processing, Results Comparison, Model tuning will be done by the all member equally.

We are going to set the deadline together and work towards the deadline in the weekly zoom meeting which will be every Monday and Wednesday 2pm to 4pm.

Team member	Responsibility
TaeJoon	Analyzing Data / Data preprocessing / Model Selection and Creation
John	Model training / Model Testing
Nick	Model evaluation / Model tuning

MIDTERM MILESTONES

1. Cleaning and transform the data set into the form that is applicable to the goal of the project
2. Complete bankruptcy classification model by analyzing the data based on the random forest classifier.
3. Compare the fitness of random forest model to the project goal with other classification models generated by the data set

REFERENCES

[1] Chih-Fong Tsai. 2008. Feature selection in bankruptcy prediction. (August 2008). Retrieved March 23, 2021 from https://www.sciencedirect.com/science/article/abs/pii/S0950705108001536

[2] Deron Liang, Chia-Chi Lu, Chih-Fong Tsai, and Guan-An Shih. 2016. Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. (January 2016). Retrieved March 23, 2021from https://www.sciencedirect.com/science/article/abs/pii/S0377221716000412

[3] Fedesoriano. 2021. Company Bankruptcy Prediction. (February 2021). Retrieved March 23, 2021 from https://www.kaggle.com/fedesoriano/company-bankruptcy-prediction

[4] Flavio Barboza, Herbert Kimura, and Edward Altman. 2017. Machine learning models and bankruptcy prediction. (April 2017). Retrieved March 23, 2021 from https://www.sciencedirect.com/science/article/pii/S0957417417302415#ecom0001