Kardelen Cicek

DAT6103 Data Mining Project: Company Bankruptcy

Individual Final Report

05.03. 2021

1. Introduction

In this semester for our DATS 6103 Data Mining course our group decided to work on Company Bankruptcy data set. This is very big and suitable data for Data Mining project. Also, as Applied Economics master students our background in economics match with the topic of data set.

To avoid financial uncertainty, publicly traded firms report their financial situation with annual files to financial authorities and regulatory agencies. In this reports firms clearly state their economic and financial standing by using financial ratios and statements. These ratios are important for financial analyses. Also, companies prepare corporate governance indicators (Shailer, 2004). These are important to permit shareholders to exercise their rights to oversee the operations of the firm and ensure that financial expectations are satisfied. Predicting and exploring company bankruptcy is important because it can prevent any potential bankrupt and negative spillover effect over market and financial sector. Predicting bankruptcy allow investors to take precautionary policies against any potential loss and risk. Investors can do necessary lending and trading activities against potential risks (Altman, 1968; Lee and Yeh, 2004; Lin et al., 2011).

The dataset, “Company Bankruptcy Prediction” which is available in Kaggle Competitions. This data was collected by the Taiwan Economic Journal from 1999 to 2009 and described the financial health of firms operating in Taiwan during this this period. The target variable is a binary variable, if company is bankrupted it is equal to zero otherwise 1. The definition of bankruptcy is based on the Taiwanese Stock Exchange regulations. The companies represented in this dataset are public firms with at least 3 consecutive years of public filings prior to the financial crisis. The companies included represent a wide variety of industries: manufacturing, service, and retail, among others.

This dataset contains 6,819 observations and 96 columns – one bankruptcy indicator variable and 95 variables that can be mostly characterized as either financial ratios or corporate governance indicators. The dataset is very well organized and without missing values.

First, we run explanatory data analysis. Then principal component analysis, Linear Probability model, and probit regression. Then we prepared predictive models. Our predictive models are four different models are: Decision Tree Classifier, Multi-Layer Perception Classifier, Histogram Based Gradient Boosting Regressor, and Logit Regression.

2. Description of Individual Work

I contributed coding, writing the report, and preparing the presentation. In coding my contribution is I prepare predictive machine learning models and I these machine learning models and their theories. I interpreted outcomes of models.

I explained background theory of PCA, LPM, Decision Tree Classifier, MLP, Histogram Based Gradient Boosting and Logit and probit regression.

Outcomes of the predictive models:

F1 Score	Estimator
0.75	Decision Tree Classifier
0.74	MLP Classifier
0.74	Histogram Gradient Boosting Classifier
0.61	Logistic Regression

Findings show that decision tree classifier had the best performance, followed by the MLP classifier and the Histogram Gradient Boosting Classifier. Logistic regression performance is relatively lower than our models. Decision Tree Classifier predicts company bankruptcy data with 75% accuracy. MLP classifier and Histogram Gradient Boosting Classifier predict the model with 0.74 accuracy. Logistic regression predicts company bankruptcy with 61 % accuracy.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, *23*(4), 589-609.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research* 16 (2002): 321-357.

Lee, Tsun-Siou, and Yin-Hua Yeh. "Corporate governance and financial distress: Evidence from Taiwan." (2004): 378-388.

Liang, Deron, Chia-Chi Lu, Chih-Fong Tsai, and Guan-An Shih. "Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study." *European Journal of Operational Research* 252, no. 2 (2016): 561-572.

Lin, Wei-Yang, Ya-Han Hu, and Chih-Fong Tsai. "Machine learning in financial crisis prediction: a survey." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, no. 4 (2011): 421-436.

Shailer, Gregory EP. *An introduction to corporate governance in Australia*. Pearson Education Australia, 2004.