Bellarmine University

Company Bankruptcy Prediction Analysis

Ashley Ridley

aridley@bellarmine.edu

Jan 15, 2024

Bankruptcy can send a chill through financial markets and involve thousands of job losses. Having an accurate company default prediction model is vital for all firms. Traditional bankruptcy prediction modes typically focus on predicting the event of bankruptcy itself, and do not consider the socio-economic consequences of their prediction. Therefore, I will use machine-learning to conduct a prediction analysis to predict and consider different costs caused by bankruptcy.

For this project I will use predictive analytics to effectively predict bankruptcy utilizing data from the Taiwan Economic Journal for the years 1999-2009. Bankruptcy or business failures can have a negative impact on both the enterprise and the global economy. This has led to business practitioners, investors, and governments to identify the potential risk of business failures in order to reduce the economic loss caused by bankruptcy (Author links open overlay panel Deron Liang a et al.). The idea for this project is to use predictive analytics successfully and accurately to predict whether a company can go bankrupt or not. This will help companies to make appropriate decisions.

Effective bankruptcy prediction is critical for financial institutions to make appropriate lending decisions. Features such as financial ratios, and prediction techniques, such as statistical and machine learning are two of the most important factors affecting the prediction performance (Author links open overlay panel Deron Liang a et al.). In addition to financial ratios, corporate governance indicators have been found to be another important input variable. However, the prediction performance obtained by combining corporate governance and financial ratios has not been fully examined. Financial ratio analysis is a quantitative method that gains insight into a company's liquidity, operational efficiency, and profitability by studying its financial statements such as the balance sheet and income statement. This insight can demonstrate how a company is performing over time and can be used to estimate future performance (Bloomenthal). However, corporate governance creates transparent rules and controls, guides leadership, and aligns the interests of shareholders, directors, management, and employees. It promotes long term financial viability, and opportunity while reducing the potential for financial loss, risks, and corruption (Bloomenthal).

The data was found on Kaggle and collected from the Taiwan Economic Journal and the company bankruptcy was defined based on the business regulations of the Taiwan Stock Exchange. It contains 6,819 instances and 96 attributes, which 6599 are negative class and 220 are positive class. The dataset is highly imbalanced, which makes this problem challenging to analyze for final predictions. The target column "Bankrupt" has values '0' and '1' for not bankrupt and bankrupt. The dataset consists of multiple financial ratio features such as: return on assets, gross profits, cash flows, taxes, and debt.

For this project I will be using predictive models such as: random forest classifier, decision tree, and K-nearest neighbor. A random forest classifier is a group-decision making team that combines the opinions of many individual models to make better predictions, creating an accurate overall model. Since the algorithm's strength relies on its ability to handle complex datasets, this method makes it valuable for various predictive tasks. Although, this dataset is imbalanced, we can manage this by using bagging and feature selection (R, Sruthi E. "Understand Random Forest). I can also use the support vector machine by assigning weights to reduce errors, which can aid in balancing the dataset and improve the overall model performance.

The decision tree algorithm is a machine learning program that uses decision trees to make predictions. It follows a tree-like model of decisions and their possible consequences by splitting the data into subsets based on the most significant feature at each node of the tree (Aunalytics). Since we want to predict whether a company will go bankrupt, the decision tree algorithm is the perfect candite. Finally, I will use K nearest neighbor, which assigns a new observation to the class or value of its closest neighbors in a training dataset (Shi1). Like the random forest classifier, K nearest neighbor is also negatively impacted by imbalance within a dataset. However, we can give different weights to the training instances or perform data resampling. All of these algorithms listed are powerful models for this project because they have the ability to limit overfitting without increasing error due to bias.

For this project I will be using the programming language python. I will import numpy, pandas, matplotlib, and seaborn for data processing and manipulation. Seaborn is a library used for making statistical graphs in python. It corelates with matplotlib and pandas, which allows me to analyze and visualize large amounts of data quickly and efficiently. Next, I will import sklearn, where I can use logistic regression, K-neighbors classifier, random forest classifier and decision tree classifier to implement machine learning models and statistical modeling. From here, I can find out more information about the data including if there are any missing values. I will also use heatmaps as a graphical representation of the data to show behavior between the data. The support vector machine and min max scalar will be used to reprocess the data and make it balanced.

This semester project will consist of using predictive analysis to identify the potential risks of bankruptcy and successfully determine which categories tend to make a company bankrupt. There have been other prediction techniques such as financial ratio analysis and corporate governance, but they have not fully been examined on how accurately they can use prediction to help businesses make reasonable decisions. Using an imbalanced dataset from Kaggle, I will use random forest classifier, decision tree, and K-nearest neighbor for prediction analysis to implement machine learning and statistical models.

References

Author links open overlay panelDeron Liang a, et al. “Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study.” *European Journal of Operational Research*, North-Holland, 13 Jan. 2016, www.sciencedirect.com/science/article/abs/pii/S0377221716000412. Accessed 16 Jan. 2024.

Bloomenthal, Andrew. “Financial Ratio Analysis: Definition, Types, Examples, and How to Use.” *Investopedia*, www.investopedia.com/terms/r/ratioanalysis.asp.

R, Sruthi E. “Understand Random Forest Algorithms with Examples (Updated 2024).” *Analytics Vidhya*, 3 Jan. 2024, www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/.

Aunalytics. “Decision Trees: An Overview.” *Aunalytics*, 2 Nov. 2021, www.aunalytics.com/decision-trees-an-overview/.

Shi1, Zhan. “IOPscience.” *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 1 Jan. 2020, iopscience.iop.org/article/10.1088/1757-899X/719/1/012072#:~:text=The%20k%2DNearest%20Neighbors%20(k,imbalanced%20class%20distributions%20of%20data.