19Z610 Machine Learning Laboratory Project - Report-1

FOR

Bankruptcy Prediction

COIMBATORE – 641 004

Prepared by

TEAM 10

20Z213 - Deepak S G
20Z229 - Krithik R
20Z237 - Pranav P
20Z242 - S Pranav
20Z259 - Viraj Agarwal

Problem Statement:

Bankruptcy prediction is the problem of detecting financial distress in businesses which will lead to eventual bankruptcy. Bankruptcy prediction is an important problem in finance, since successful predictions would allow stakeholders to take early actions to limit their economic losses.

Bankruptcy prediction has been studied since at least the 1930s. The early models of bankruptcy prediction employed univariate statistical models over financial ratios. The univariate models were followed by multivariate statistical models such as the famous Altman Z-score model. The recent advances in the field of Machine learning have led to the adoption of Machine learning algorithms for bankruptcy prediction. Machine Learning methods are increasingly being used for bankruptcy prediction using financial ratios.

Bankruptcy prediction is important for modern economies because early warnings of bankruptcy help not only the investor but also public policy makers to take proactive steps to minimize the impact of bankruptcies. The reasons that add to the significance of bankruptcy prediction are as follows:

Significance of bankruptcy prediction

(1). Better allocation of resources

Institutional investors, banks, lenders, retail investors are always looking at information that predicts financial distress in publicly traded firms. Early prediction of bankruptcy helps not only the investors and lenders but also the managers of a firm to take corrective action thereby conserving scarce economic resources. Efficient allocation of capital is the cornerstone of growth in modern economies.

(2). Input to policy makers

Accurate prediction of bankruptcies of businesses and individuals before they happen gives law makers and policy makers some additional time to alleviate systemic issues that might be causing the bankruptcies. Indeed, with bankruptcies taking center stage in political discourse of many countries, the accurate prediction of bankruptcy is a key input for politicians, bureaucrats and in general for anyone who is making public policy.

(3). Corrective action for business managers

The early prediction of bankruptcy is likely to highlight business issues thereby giving the company's manager additional time to make decisions that will help avoid bankruptcy. This effect is likely to be more profound in public companies where the management has a fiduciary duty to the shareholders.

(4). Identification of sector wide problems

Bankruptcy prediction models that flag firms belonging to a certain sector are likely to be a leading indicator of an upcoming downturn in a certain sector of an economy. With robust models, the business managers and government policy makers would become aware and take corrective action to limit the magnitude and intensity of the downturn in the specific sector. Industry groups in turn have been shown to significantly affect forecasting models.

(5). Signal to Investors

Investors can make better and more informed decisions based on the prediction of bankruptcy models. This not only forces the management of firms to take corrective action but also helps to soften the overall economic fallout that results from the bankruptcies. Empirical studies have shown that investment opportunities are significantly related to likelihood of bankruptcy

(6). Relation to adjacent problems

Bankruptcy prediction is often the first step used by ratings agencies to detect financial distress in firms. Based on the predictions of bankruptcy models, ratings agencies investigate and assess credit risk. Getting flagged by bankruptcy prediction models is often the first step that triggers the process of revising credit ratings. A literature survey covering 2000–2013 demonstrates the close relation between bankruptcy prediction and credit risk.

Dataset Description:

Link: https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction

The Dataset consists of 6819 rows and 96 columns. The column “Bankrupt?” is a boolean column(0/1), where:

0 indicates that the situation has not led to Bankruptcy.
1 indicates that the situation has led to Bankruptcy.

The columns contain financial information, some of which greatly describe the course of events.

For example, the feature “Current Ratio” is an important indicator of the progress of the company.

Current Ratio = Current Assets/Current Liability

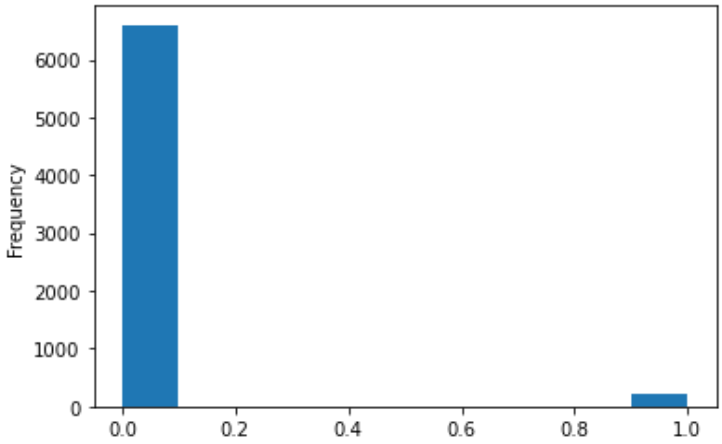

Fig: Histogram Depicting the “Bankrupt?” Column

```
In [21]: df.describe()
```

Out[21]:

	Bankrupt?	ROA(C) before interest and depreciation before interest	ROA(A) before interest and % after tax	ROA(B) before interest and depreciation after tax	Operating Gross Margin	Realized Sales Gross Margin	Operating Profit Rate	Pre-tax net Interest Rate	After-tax net Interest Rate	Non-industry income and expenditure/revenue	...	Net Inc to T As
count	6819.000000	6819.000000	6819.000000	6819.000000	6819.000000	6819.000000	6819.000000	6819.000000	6819.000000	6819.000000	...	6819.000
mean	0.032263	0.505180	0.558625	0.553589	0.607948	0.607929	0.998755	0.797190	0.809084	0.303623	...	0.807
std	0.176710	0.060686	0.065620	0.061595	0.016934	0.016916	0.013010	0.012869	0.013601	0.011163	...	0.040
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000
25%	0.000000	0.476527	0.535543	0.527277	0.600445	0.600434	0.998969	0.797386	0.809312	0.303466	...	0.796
50%	0.000000	0.502706	0.559802	0.552278	0.605997	0.605976	0.999022	0.797464	0.809375	0.303525	...	0.810
75%	0.000000	0.535563	0.589157	0.584105	0.613914	0.613842	0.999095	0.797579	0.809469	0.303585	...	0.826
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000

8 rows × 96 columns

Fig: Description of data Frame

Tools to be used:

1. **Coding environment -** Google Colab/ VSCode
2. **Language -** Python
3. **Libraries:**
    a. **Scikit-learn:** Scikit-learn is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via an interface in Python. It can be used in bankruptcy prediction by providing necessary tools for feature scaling, model training and evaluation, data splitting and cross validation.
    b. **Pandas:** Pandas is a Python library for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series.
    c. **Numpy:** Numpy is a Python library for mathematical computation of large, multidimensional arrays and matrices. It offers a large collection of high-level mathematical functions to operate on these arrays.
    d. **Matplotlib:** A plotting library for creating static, animated, and interactive visualizations in Python. It can be used to create histograms, scatter plots, line charts, bar charts, and many other types of visualizations.
    e. **Seaborn:** Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing statistical graphics.

Contributions:

Roll number	Name	Assigned work
20Z213	Deepak S G	Encoding categorical data, filling missing numbers.
20Z242	S Pranav	Feature scaling.
20Z229	Krithik R	Training various models, ensemble models, applying hyperparameter tuning,cross validation.
20Z259	Viraj Agarwal	Training various models, ensemble models, applying hyperparameter tuning,cross validation.
20Z237	Pranav P	Evaluation metrics, graphical representation using matplotlib.

References:

1) https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction
2) https://www.aimspress.com/article/doi/10.3934/DSFE.2021010?viewType=HTML
3) https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html
4) https://scikit-learn.org/stable/modules/model_evaluation.html
5) Barboza F, Kimura H, Altman E (2017) Machine learning models and bankruptcy prediction. Expert Syst Appl 83: 405-417. doi: 10.1016/j.eswa.2017.04.006
6) Cao B, Zhan D, Wu X (2009) Application of svm in financial research, In: CSO international joint conference on computational sciences and optimization, 2: 507-511.
7) Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system, In: Proc. of KDD'16, 785-794.
8) Altman EI (1968) The Prediction of Corporate Bankruptcy: A Discriminant Analysis. J Financ 23: 193-194.
9) https://www.udemy.com/course/complete-machine-learning-and-data-science-zero-to-mastery
10) https://www.kaggle.com/competitions/bluebook-for-bulldozers/overview