



Bankruptcy detection in the Taiwanese stock exchange



Noah Hunsicker



Data Sources

- The data comes from the Taiwan Economic Journal from the years of 1999 to 2009 and was accessed through Kaggle.

Business Question

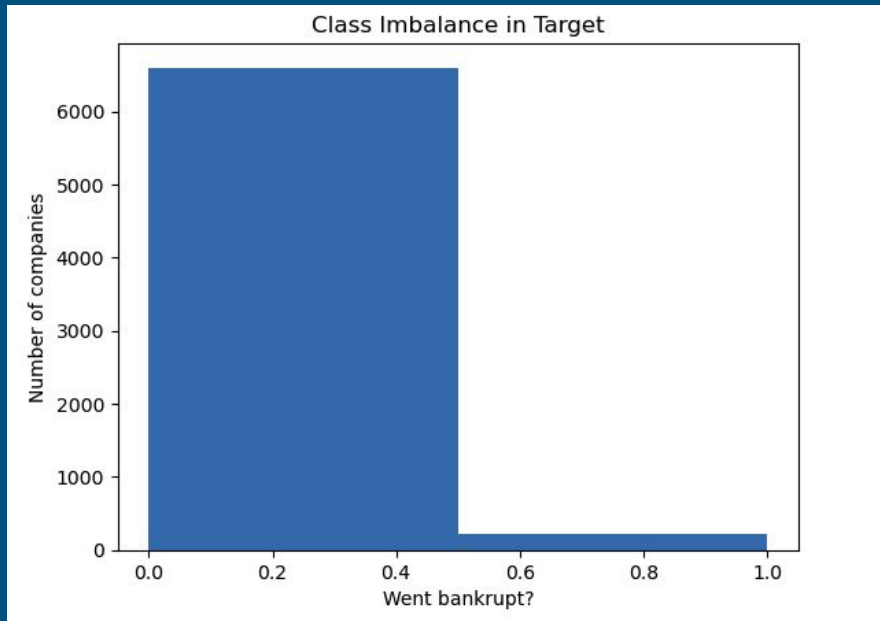
How well can we predict if a company is about to go bankrupt by just looking at publicly available financial disclosure data.

Criteria:

- If a company will go bankrupt based on their most recent financial disclosure

Distribution of data

- Very high class imbalance
- Treating this issue with BorderlineSMOTE
- Better at classifying boundary points and noise points



Choosing a metric

- Most important metric for this problem is recall
- High recall is good because it highly prioritizes alerting us when a company is going to go bankrupt, would rather be safe than sorry.

Recall



Accuracy and
precision

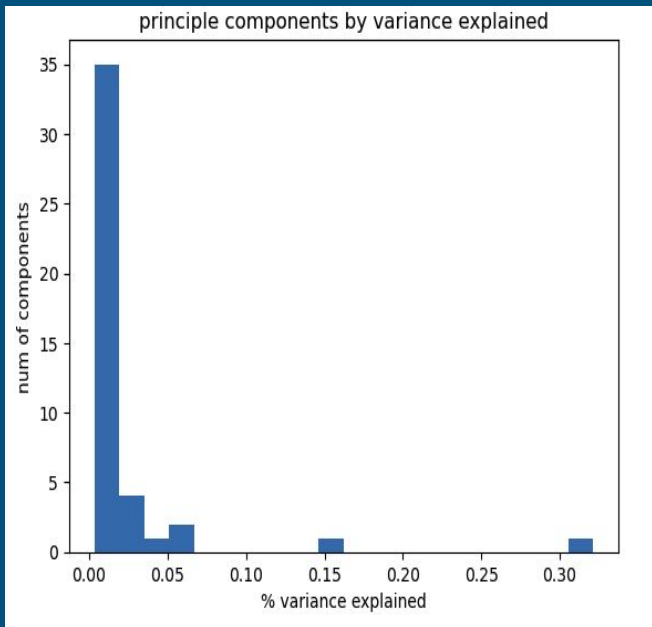


The features

[' 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'
' Continuous interest rate (after tax)' ' Operating Expense Rate'
' Research and development expense rate' ' Cash flow rate'
' Interest-bearing debt interest rate' ' Tax rate (A)'
' Net Value Per Share (B)' ' Net Value Per Share (A)'
' Net Value Per Share (C)' ' Persistent EPS in the Last Four Seasons'
' Cash Flow Per Share' ' Revenue Per Share (Yuan ¥)'
' Operating Profit Per Share (Yuan ¥)'
' Per Share Net profit before tax (Yuan ¥)'
' Realized Sales Gross Profit Growth Rate'
' Operating Profit Growth Rate' ' After-tax Net Profit Growth Rate'
' Regular Net Profit Growth Rate' ' Continuous Net Profit Growth Rate'
' Total Asset Growth Rate' ' Net Value Growth Rate'
' Total Asset Return Growth Rate Ratio' ' Cash Reinvestment %'
' Current Ratio' ' Quick Ratio' ' Interest Expense Ratio'
' Total debt/Total net worth' ' Debt ratio %' ' Net worth/Assets'
' Long-term fund suitability ratio (A)' ' Borrowing dependency'
' Contingent Liabilities/Net worth' ' Operating profit/Paid-in capital'
' Net profit before tax/Paid-in capital'
' Inventory and accounts receivable/Net value' ' Total Asset Turnover'
' Accounts Receivable Turnover' ' Average Collection Days'
' Inventory Turnover Rate (times)' ' Fixed Assets Turnover Frequency'
' Net Worth Turnover Rate (times)' ' Revenue per person'
' Operating profit per person' ' Allocation rate per person'
' Working Capital to Total Assets' ' Quick Assets/Total Assets'
' Current Assets/Total Assets' ' Cash/Total Assets'
' Quick Assets/Current Liability' ' Cash/Current Liability'
' Current Liability to Assets' ' Operating Funds to Liability'
' Inventory/Working Capital' ' Inventory/Current Liability'
' Current Liabilities/Liability' ' Working Capital/Equity'
' Current Liabilities/Equity' ' Long-term Liability to Current Assets'
' Retained Earnings to Total Assets' ' Total income/Total expense'
' Total expense/Assets' ' Current Asset Turnover Rate'
' Quick Asset Turnover Rate' ' Working capital Turnover Rate'
' Cash Turnover Rate' ' Cash Flow to Sales' ' Fixed Assets to Assets'
' Current Liability to Liability' ' Current Liability to Equity'
' Current Liability to Current Assets' ' Liability-Assets Flag'
' Net Income to Total Assets' ' Total assets to GNP price'
' No-credit Interval' ' Gross Profit to Sales'
' Net Income to Stockholder's Equity' ' Liability to Equity'
' Degree of Financial Leverage (DFL)' ' Interest Coverage Ratio (Interest expense to EBIT)' ' Net Income Flag'
' Equity to Liability']

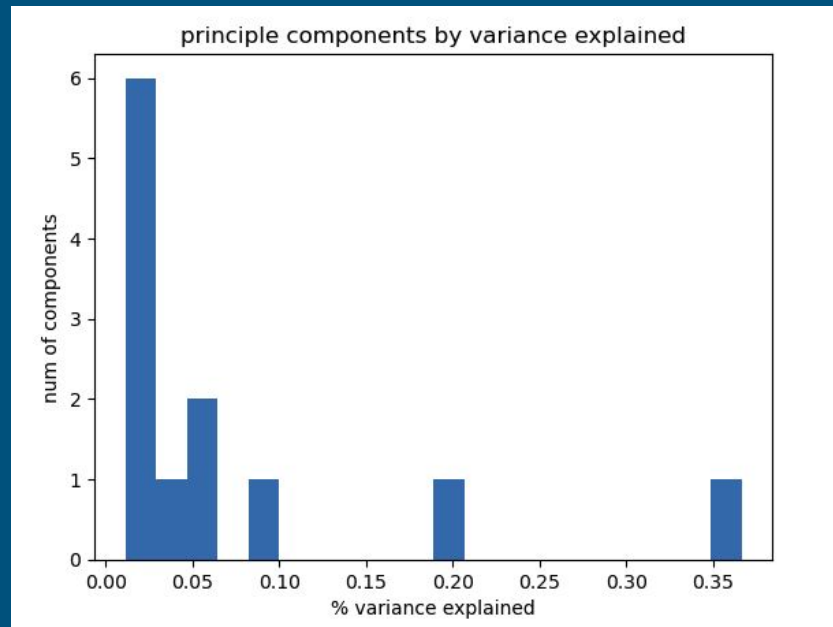


Feature reduction



<-before

after->



Modeling

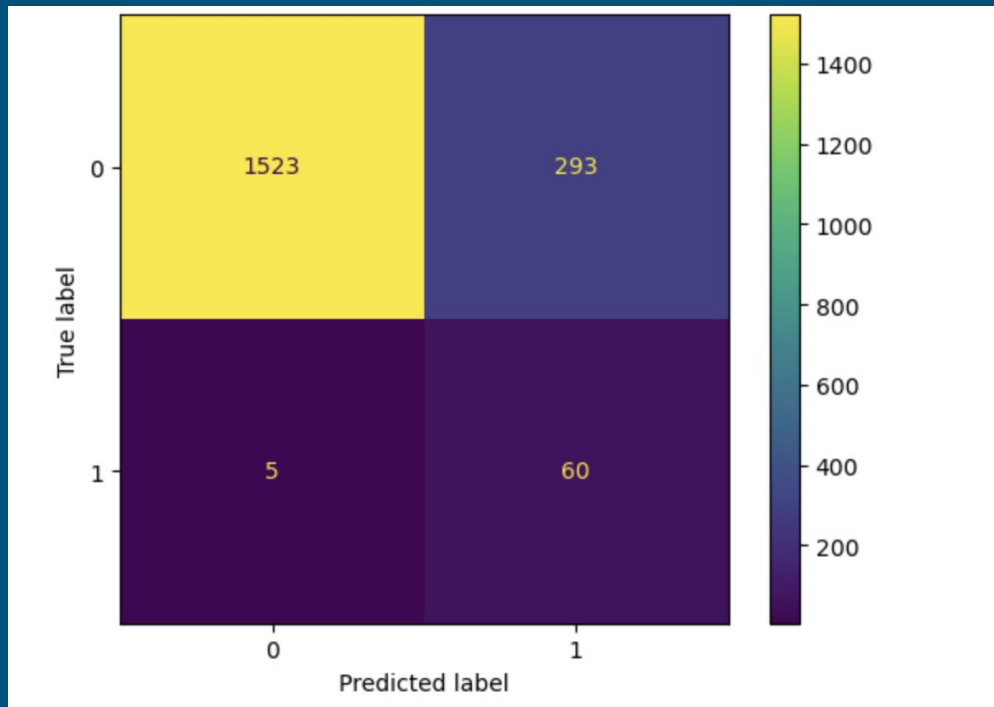
Model types used for problem:

- Logistic regression
- Random forest classifier
- Gradient boosted trees
- Neural network

Best Model

Random Forest Classifier

- Highest recall score class 1
- 99.6% precision score for class 0
- Faster than other models



Next Steps

- Implement a time series element to the data, change in financials is as important as current financials
- Get a larger dataset, class imbalance greatly hinders predictive power of model