Company Bankruptcy Prediction

Presented by
Artem Ramus

Introduction and Background

The goal of this project is to spot bankrupt companies from 95 financial book attributes.

Bankruptcy data from the Taiwan Economic Journal for the years 1999–2009

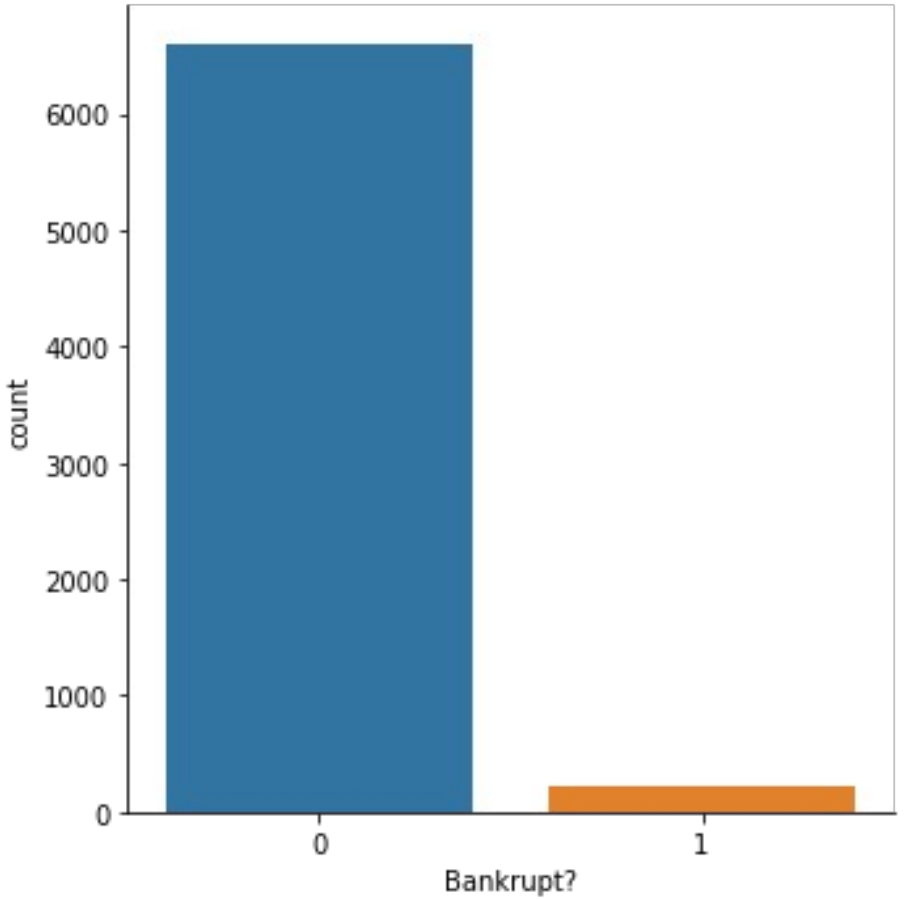
The data were collected from the Taiwan Economic Journal for the years 1999 to 2009. Company bankruptcy was defined based on the business regulations of the Taiwan Stock Exchange.

Source: Deron Liang and Chih-Fong Tsai, deronliang@gmail.com; cftsai@mgt.ncu.edu.tw, National Central University, Taiwan The data was obtained from UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/Taiwanese+Bankruptcy+Prediction

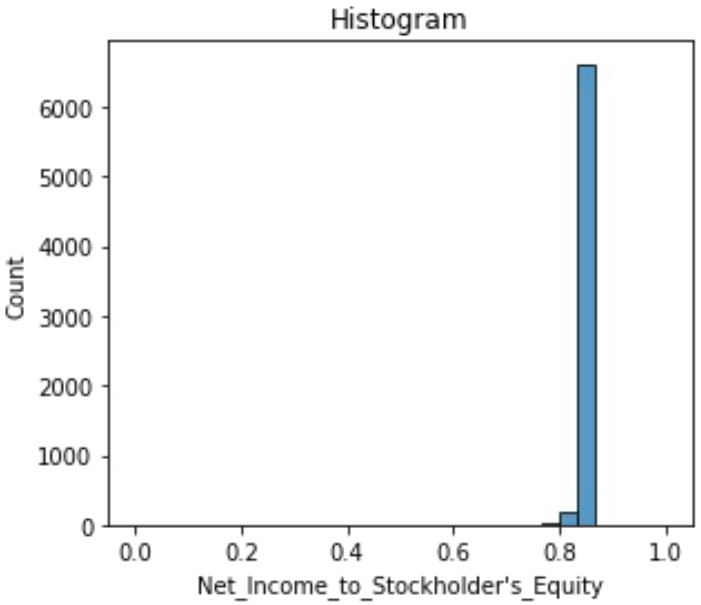
Exploratory Data Analysis - target

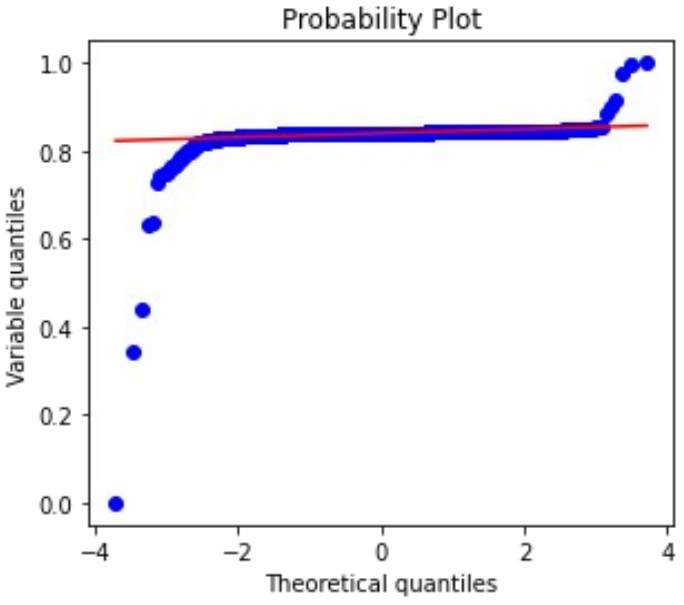
The target is imbalanced with underrepresented class 1, bankrupt.

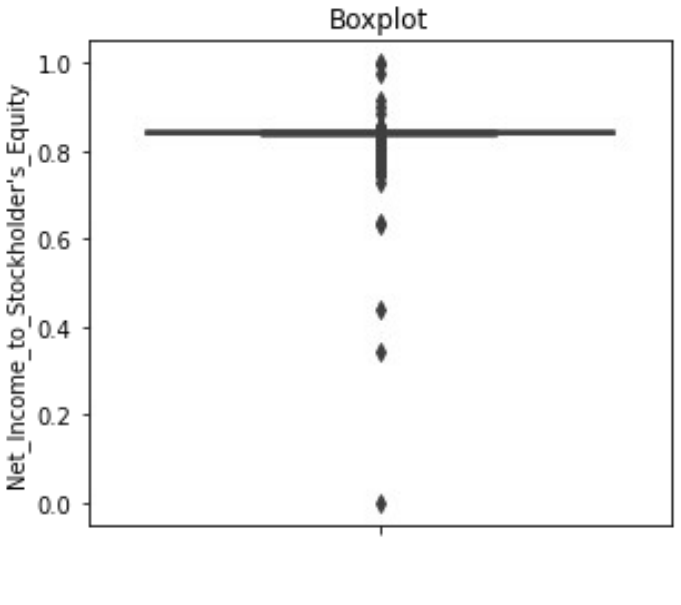
class	0	1
count	6599	220



Exploratory Data Analysis







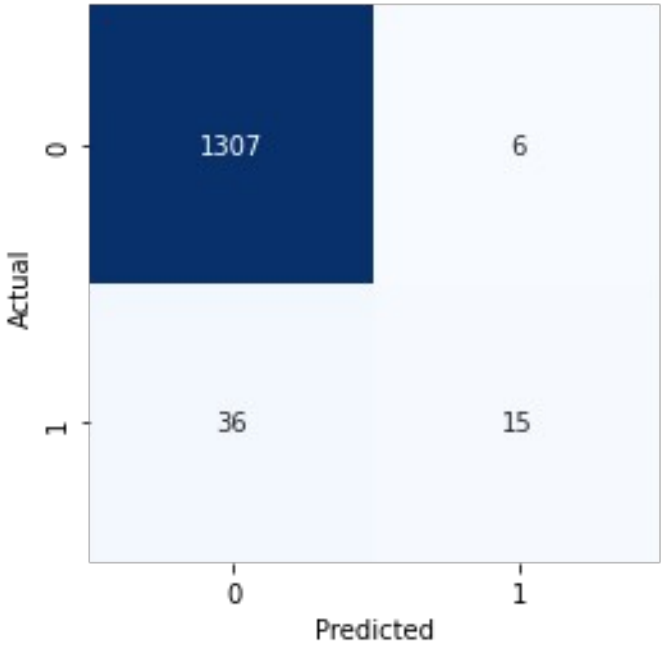
Most variables are skewed or/and have high kurtosis. Some of them have outliers on one or both sides. All the numerical variables have mean that differs between target classes.

Feature Engineering and Selection

- Original features’ distribution
- Shuffle feature selection didn't give any insight because all the features reduced the score by only 0.009 to 0.023.
- Ordering features by single feature prediction and then iteratively selecting k best features gave ROC AUC score of 0.9581 and reduced number of features to 10.

Performance

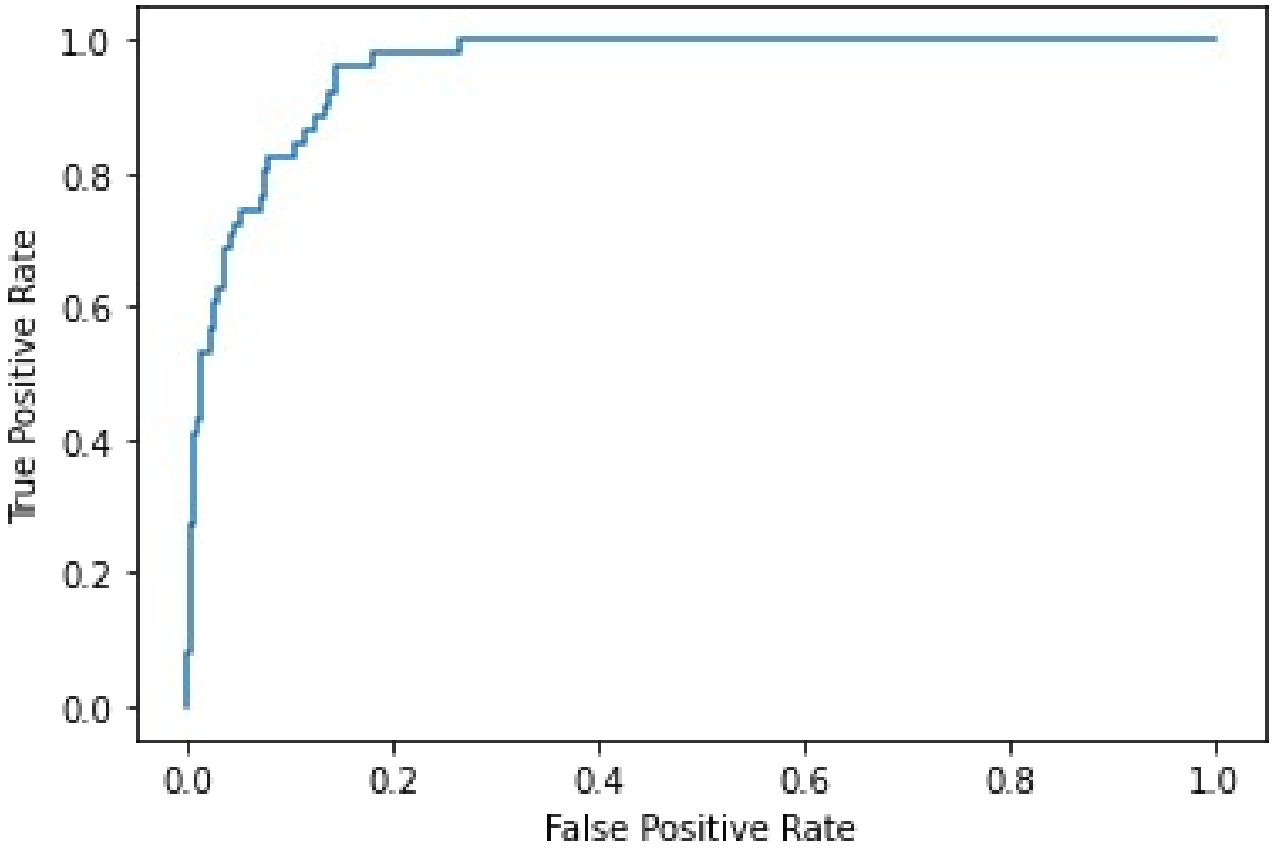
	precision	recall	f1-score	support
0	0.97	1.00	0.98	1313
1	0.71	0.29	0.42	51
accuracy			0.97	1364
macro avg	0.84	0.64	0.70	1364
weighted avg	0.96	0.97	0.96	1364



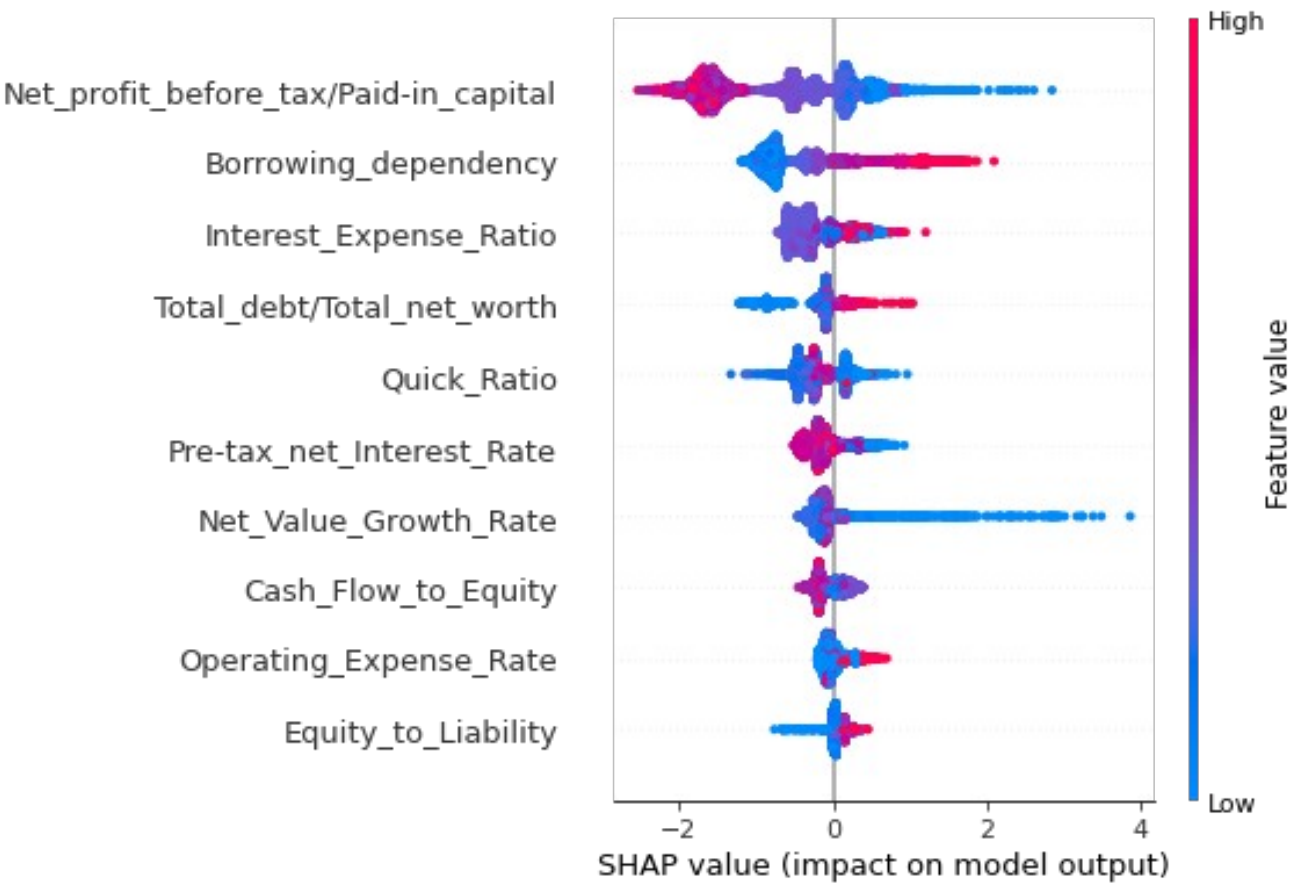
Probability cutoff 0.5 results in precision 0.71 and recall 0.29.

Performance

ROC AUC score:
0.9581



Feature Importance - SHAP



Summary and Conclusions

- XGBoost modeling with 95 features gave ROC AUC score of 0.9588.
-
- Shuffle feature selection didn't give any insight because all the features reduced the score by only 0.009 to 0.023.

- Ordering features by single feature prediction and then iteratively selecting k best features gave ROC AUC score of 0.9581 and reduced nubmer of features to 10.

The end

Thank you for your attention!