

A Look into the Past: Predicting Bankruptcies in Taiwan

By: Sergio Rodriguez



Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Data

- Data used in this project was 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.
- Column Names:
 - Cash Flow to Equity
 - Cash Flow to Sales
 - Total Asset Turnover
 - Net Income to Total Asset
- Kaggle link below:

-<https://www.kaggle.com/fedesoriano/company-bankruptcy-prediction>|

Business Problem

Taiwan Rating's Agency is revamping their current bankruptcy models to ensure the financial of the companies listed on the Taiwan Stock Exchange. With the amount of volatility and bankruptcies currently being experienced due to Covid-19, they would like to assess the solvency of the companies listed on the TSE. They have tasked us with creating a predictive model that takes into account the Taiwan bankruptcies from '99 - '09 to use on the currently listed companies on the TSE.

Summary

- Our XGBoosted Tree was our best performing model predicting 48/58 bankrupt companies
- Our top 3 most important features were:
 - Net income to total assets
 - Total debt/ total net worth
 - Interest-bearing debt interest rate

Methods: Pre-processing and Transformation

- Pre-processing:
 - This entailed cleaning our data in order for us to be able to model without any issues.
 - Checked for missing values, duplicates or troublesome datatypes
- Transformation:
 - To deal with multicollinearity, we dropped 21 features that were highly correlated to each other
- Class Imbalance:
 - Most importantly, our data only had 3% of all data with companies that had gone bankrupt which is a concern when training models
 - There is a method called Smote which allowed us to alleviate this issue by oversampling our minority class i.e. make synthetic copies of companies that were bankrupt

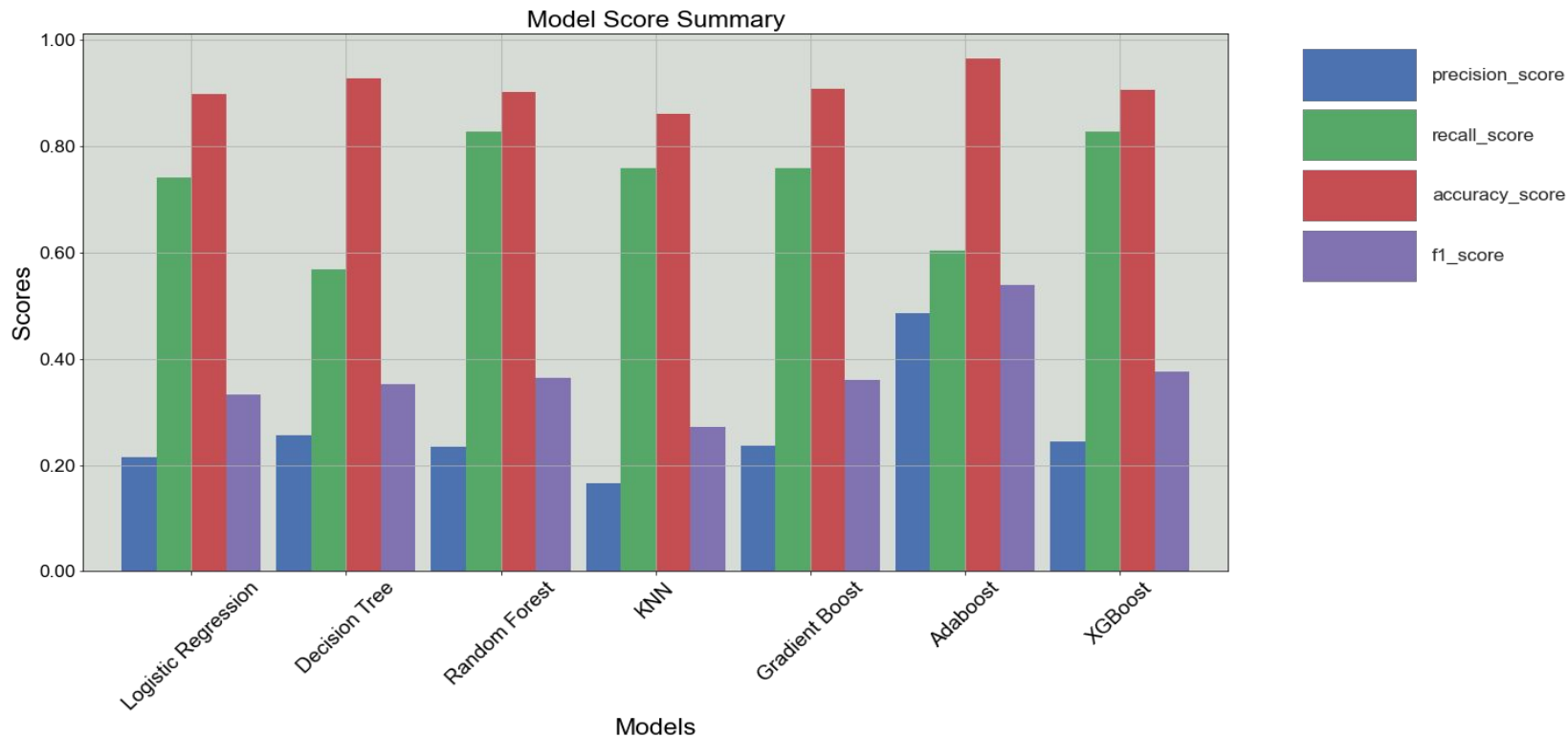
Results: Summary Table

- Note high recall, but low precision score

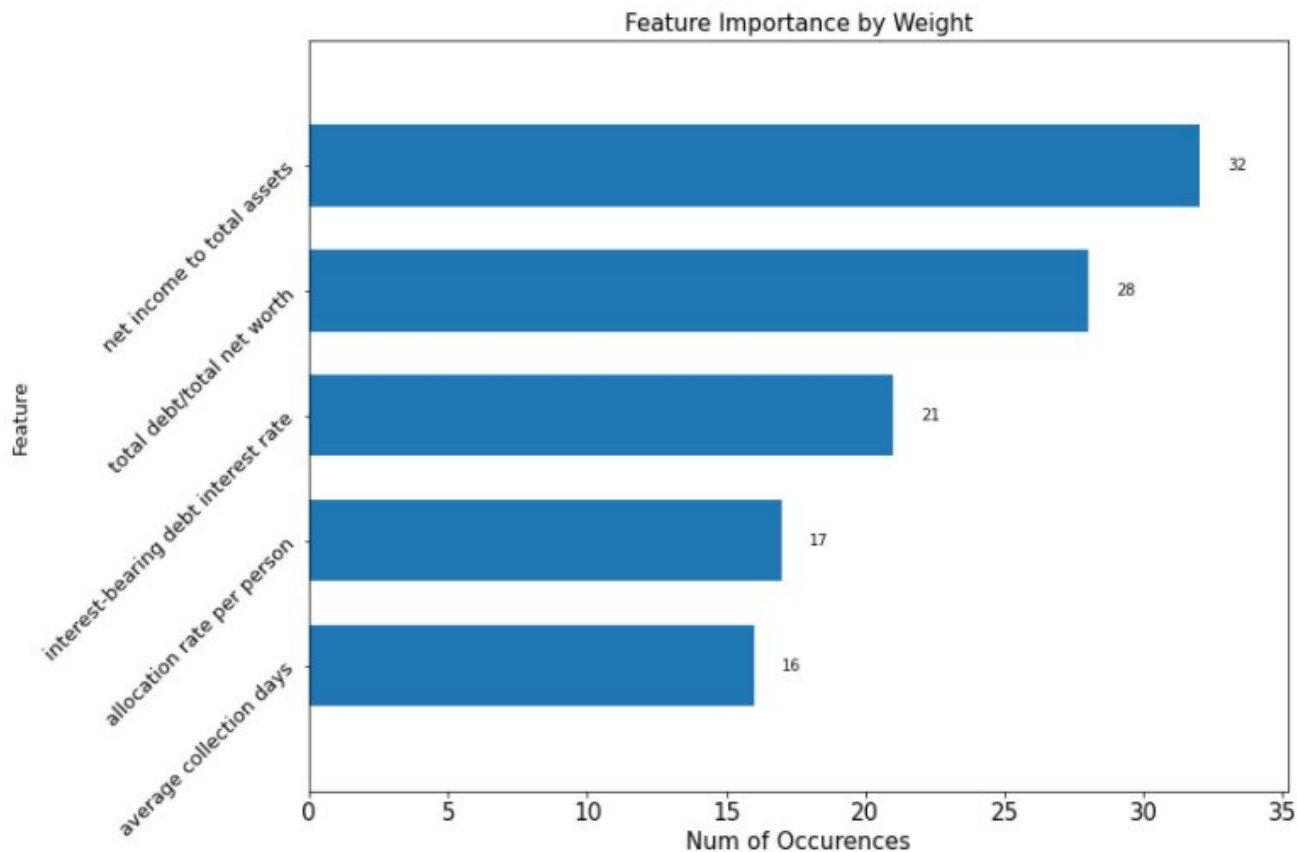
	precision_score	recall_score	accuracy_score	f1_score
Model				
XGBoost	0.245000	0.844828	0.906158	0.379845
Random Forest	0.228155	0.810345	0.900293	0.356061
KNN	0.171533	0.810345	0.860411	0.283133
Gradient Boost	0.239362	0.775862	0.908504	0.365854
Logistic Regression	0.213930	0.741379	0.898534	0.332046
Decision Tree	0.253521	0.620690	0.924927	0.360000
Adaboost	0.439024	0.620690	0.960117	0.514286

Results

Our best performing model was XGBoost which was able to predict 49/58 companies

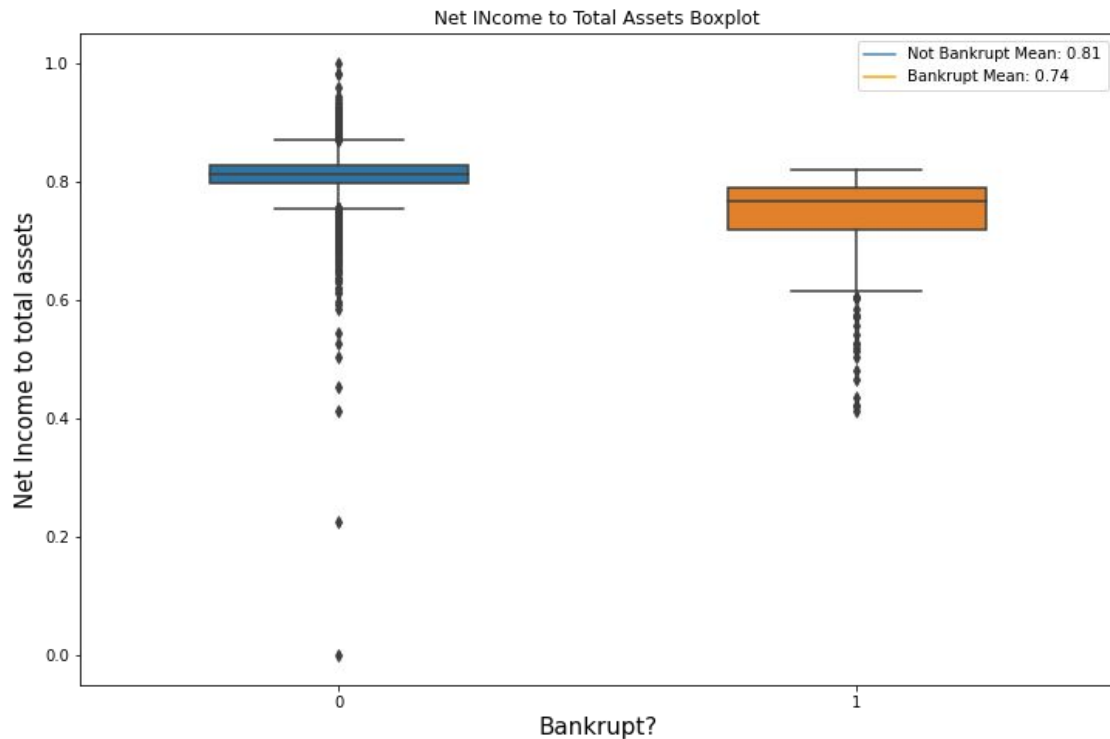


Results: Feature Importance



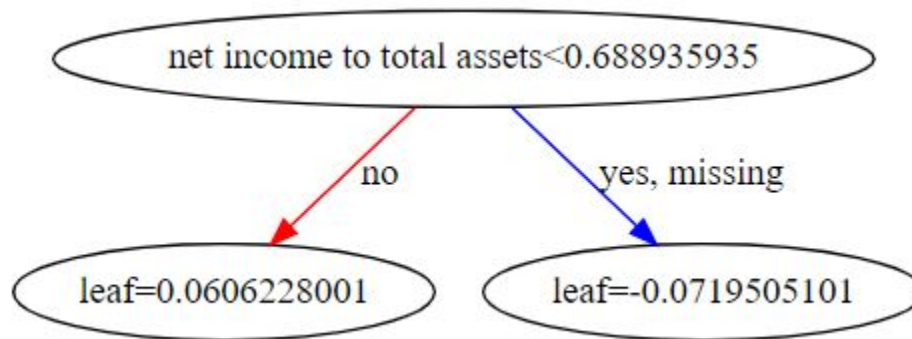
Results: Feature Importance

- Here we see that our mean Net income to total assets is higher for companies that did not go bankrupt



Results: XG Boost Tree

- Our tree's threshold for deciding if company was bankrupt or not was .69



Conclusions

- Of all of the models ran, the XG Boosted Tree was our best model at predicting bankrupt companies with a recall score of .84 predicting 49/58 bankrupt companies
- Across all of our models we were able to increase our recall score but at the expense of precision i.e. predicted companies to be bankrupt that were not
- Our most important features were:
 - Net Income to total assets
 - Total Debt/ Total Net Worth
 - Interest Rate on Interest Bearing Debt

Future Work

- Predict Credit Ratings
- Include more financial numbers than just ratios
- Assess stock price or tradeable instrument affects

Thank You!

Email: srodriguez2742@gmail.com

GitHub: [@srodriguez2742](https://github.com/srodriguez2742)