

Bankruptcy Prediction

Problem Statement

To predict bankruptcy and the main features involved, which could help a company detect early signs and can help reduce economic losses.

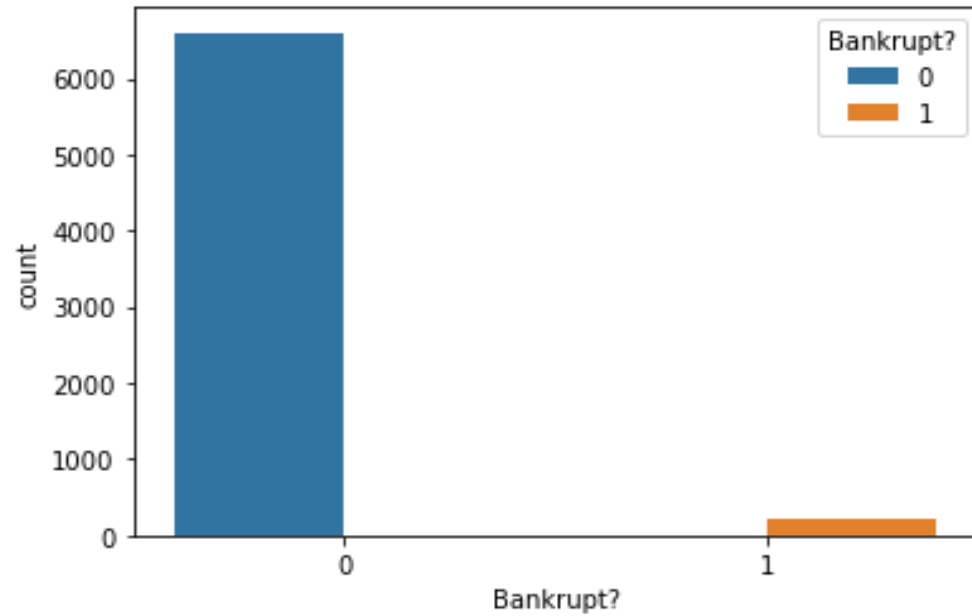
Bankruptcy Prediction

- Classification model
- ~ 7000 rows and 95 columns
- Yes/ No class label

Data Wrangling

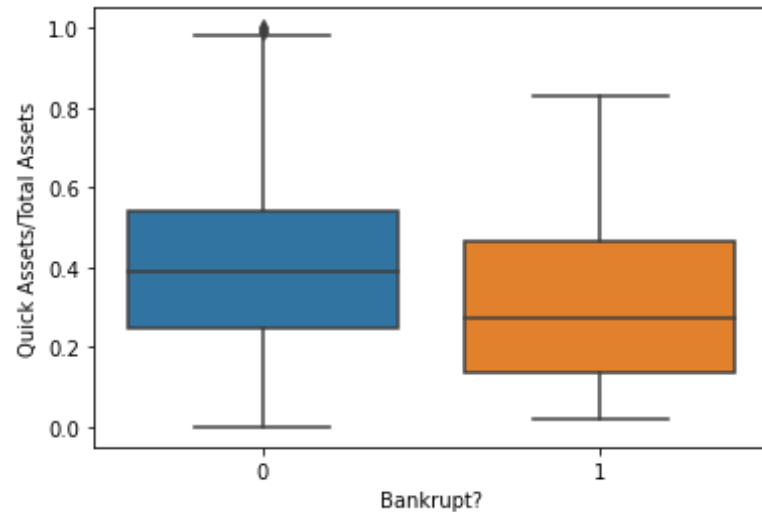
- No missing values
- Explored Outliers

Exploratory Data Analysis



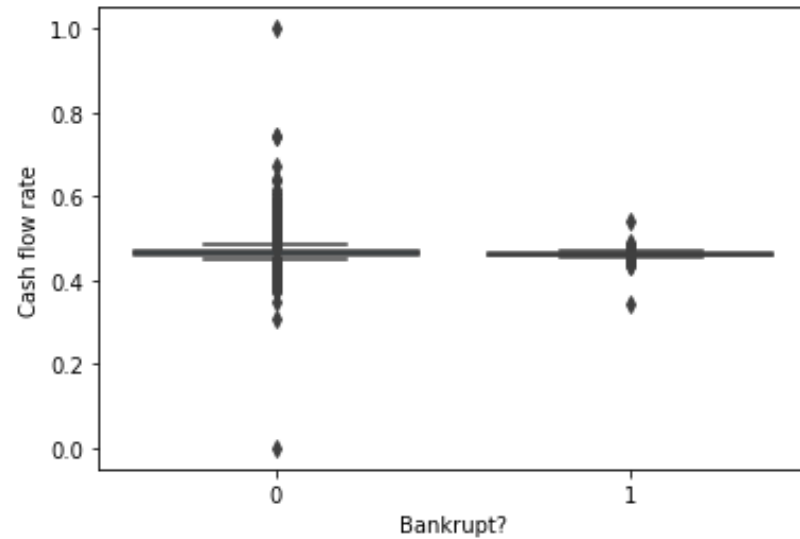
- 6599 No class label
- 220 Yes class label
- ~ 3% Yes class

Correlation



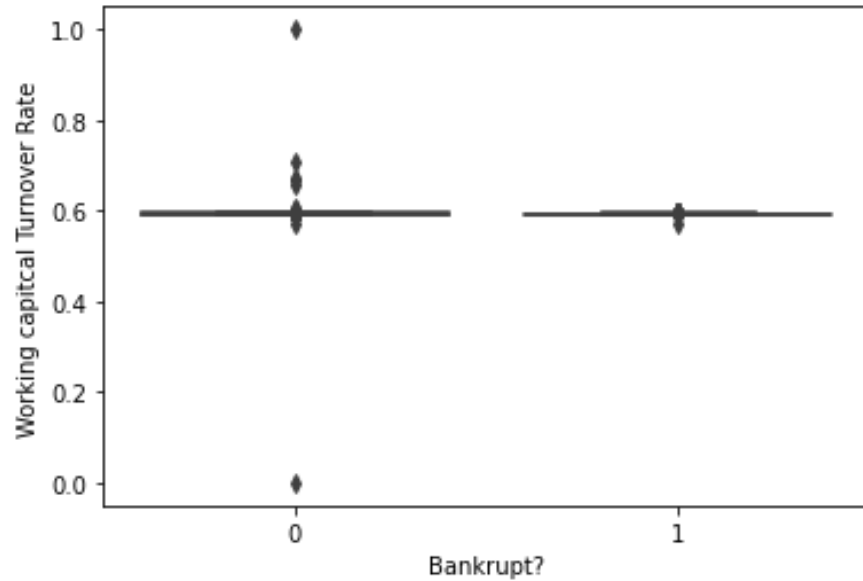
- Compared features with the class labels
- The average of Total Assets of the companies that went bankrupt were lower

Correlation



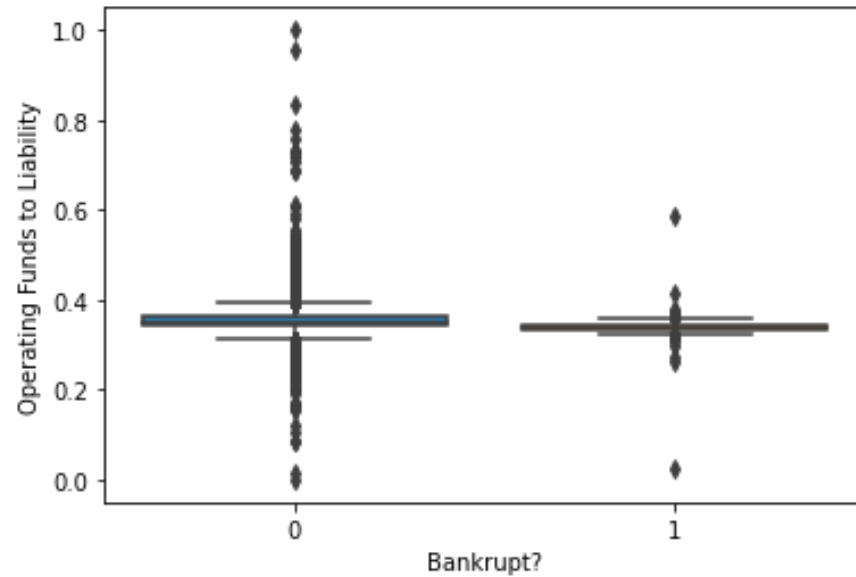
Cash flow rate was
ranged higher for the
No class label

Correlation



Working capital turnover rate ranged higher for the No class label

Correlation



Operating Funds to Liability ranged higher for the No class label

Preprocessing and Training Data

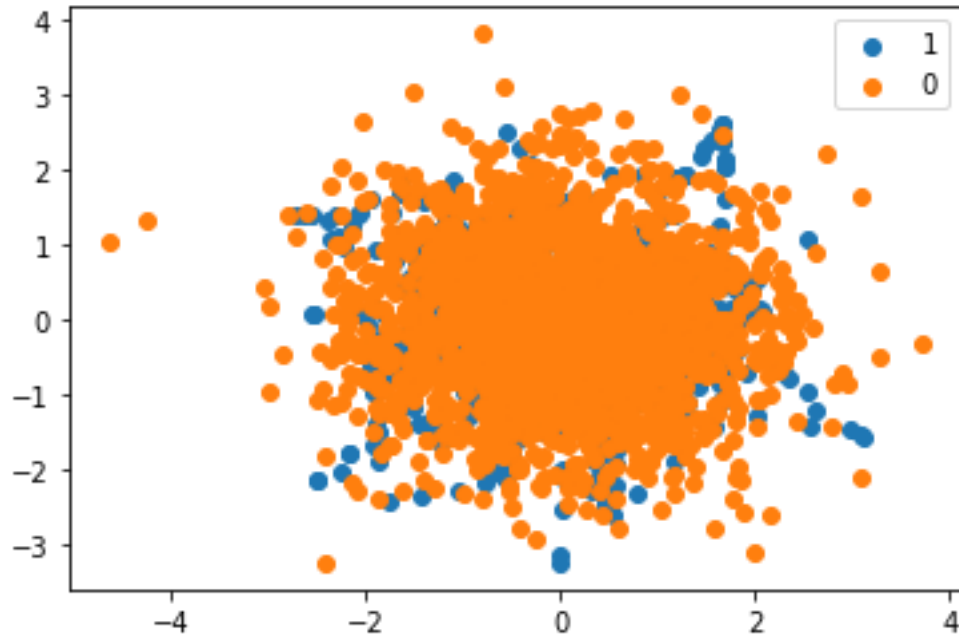
- 80-20 split
- Initial models
- Logistic Regression
- Decision trees

Preprocessing and Training Data

	precision	recall	f1-score	support
0	0.97	0.99	0.98	1313
1	0.61	0.22	0.32	51
accuracy			0.97	1364
macro avg	0.79	0.61	0.65	1364
Wt. avg	0.96	0.97	0.96	1364

- F1-score
- Between .3-.4 for No class label for all models

Preprocessing and Training Data



- Class imbalance
- SMOTE (Synthetic Minority Oversampling Technique)

Modelling – Logistic Regression

	precision	recall	f1-score	support
0	0.98	0.97	0.97	387
1	0.94	0.96	0.95	207
accuracy			0.96	594
macro avg	0.96	0.96	0.96	594
weighted avg	0.96	0.96	0.96	594

Modelling – Decision Tree

	precision	recall	f1-score	support
0	0.90	0.93	0.91	387
1	0.96	0.94	0.95	207
accuracy			0.93	594
macro avg	0.93	0.93	0.93	594
weighted avg	0.93	0.93	0.93	594

Decision Tree – Important Features

- Liability to Equity
- Working Capital Turnover Rate
- Revenue per Share
- ROA before share % and after tax
- Operating Profit Rate
- Quick Assets/ Total Assets
- Net worth
- Average collection days
- Cash flow to sales
- Debt Ratio %

Modelling – Gradient Boosting

	precision	recall	f1-score	support
0	0.98	0.97	0.98	387
1	0.95	0.96	0.95	207
accuracy			0.97	594
macro avg	0.96	0.97	0.96	594
weighted avg	0.97	0.97	0.97	594

Modelling – XGBoost

	precision	recall	f1-score	support
0	0.99	0.98	0.99	387
1	0.96	0.99	0.97	207
accuracy			0.98	594
macro avg	0.98	0.98	0.98	594
weighted avg	0.98	0.98	0.98	594

Summary – Important steps

- Looked at the data/ outliers.
- Compared features with class labels and with each other.
- Applied initial models and got low F1-score for No class label.
- Oversampled No class label to solve class imbalance using SMOTE.
- Applied models again and saw a significant increase in accuracy.
- Important features through decision trees.
- Best prediction with XGBoost.