



Company Bankruptcy Prediction

Predicting bankruptcy risk through
Machine Learning Classification

Cindy Su - 10/29/2021

Introduction

- **Goal:** Predicting bankruptcy risk and possible financial distress of public companies through financial ratios.
- **Client:** Bank Creditors or Investors
- **Product:** Machine Learning Classification model for bankruptcy risk prediction

Data & Methodology



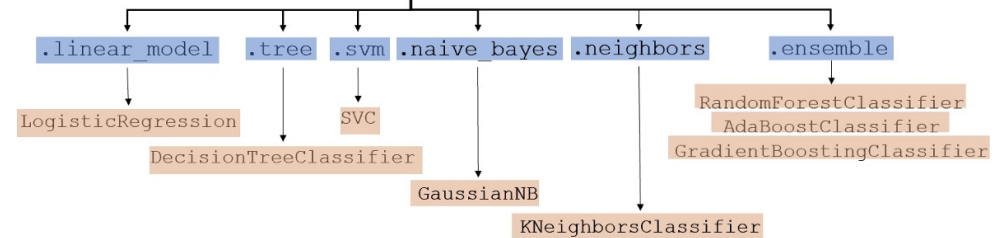
Data & Methodology



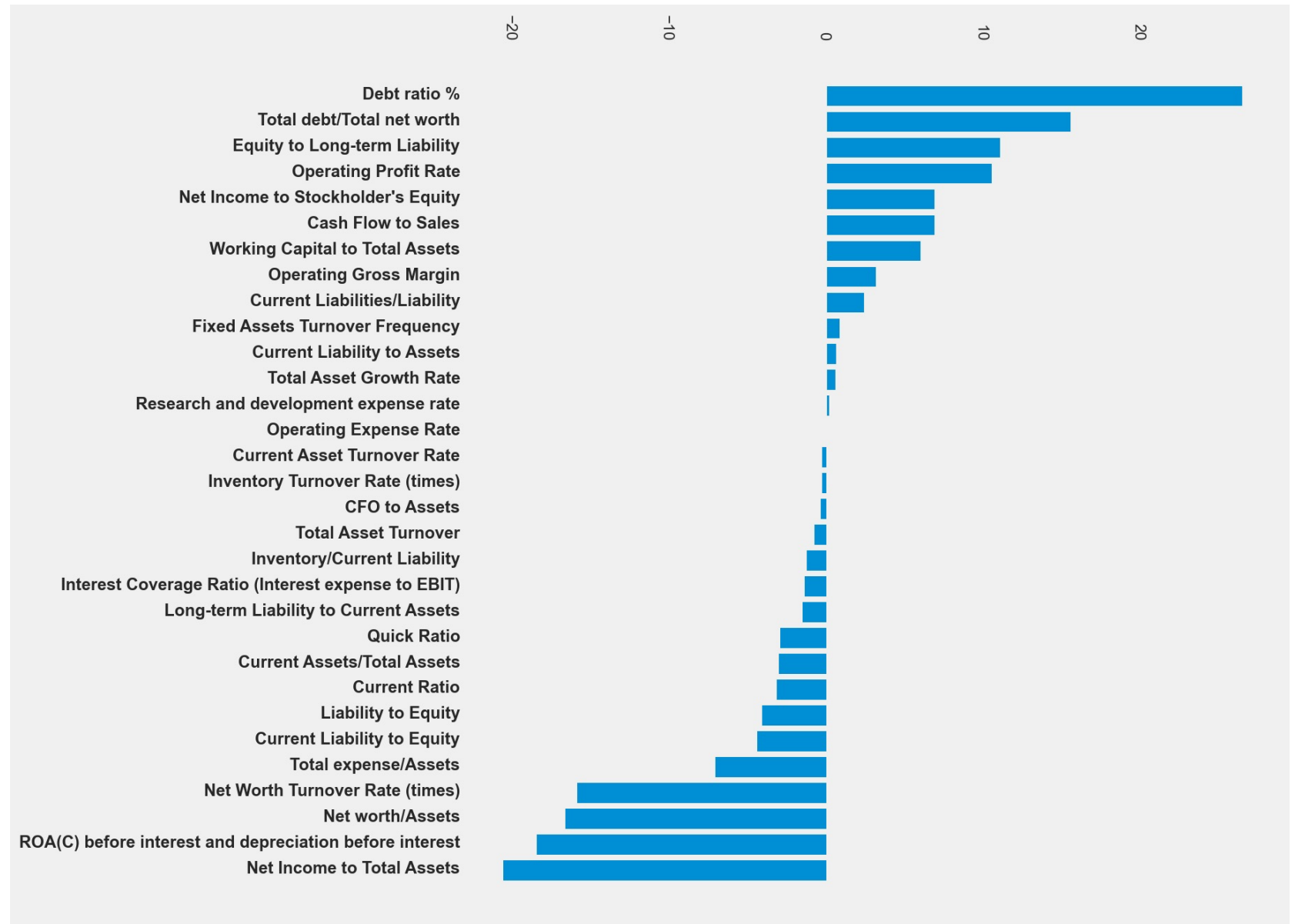
- Taiwan Economic Journal
- 800-900 listed companies
- From 1999 to 2009
- rows: 6819
- features: 96



- Feature selection & Dimensionality reduction

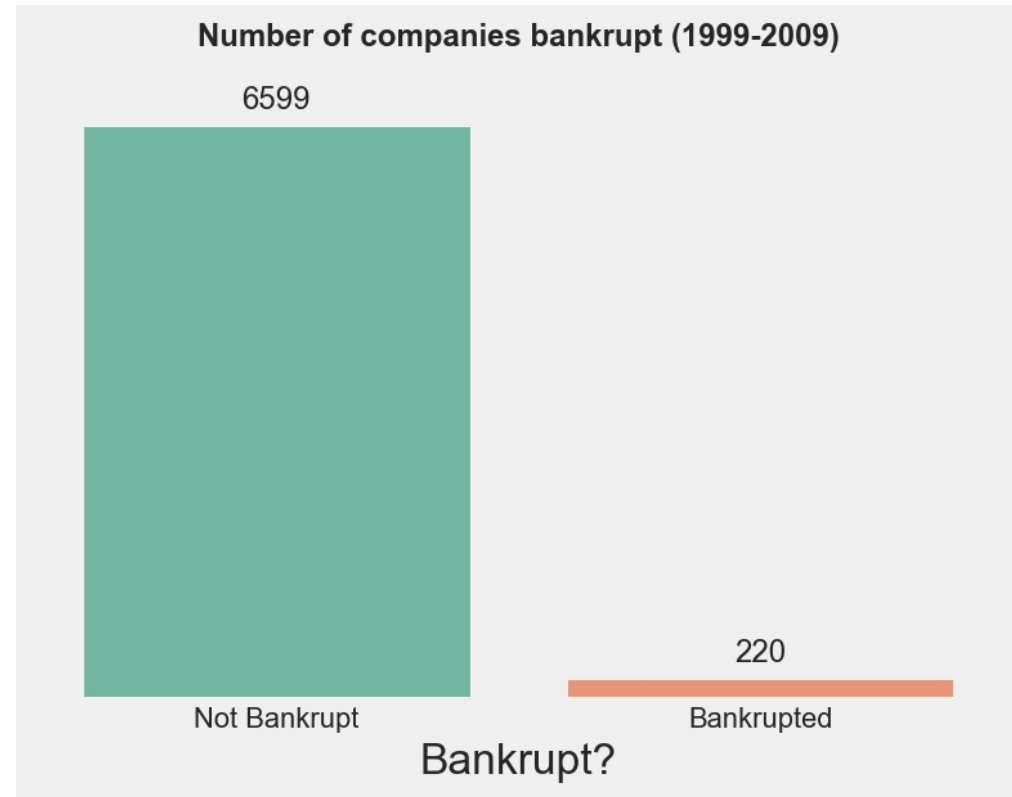


Exploratory Data Analysis



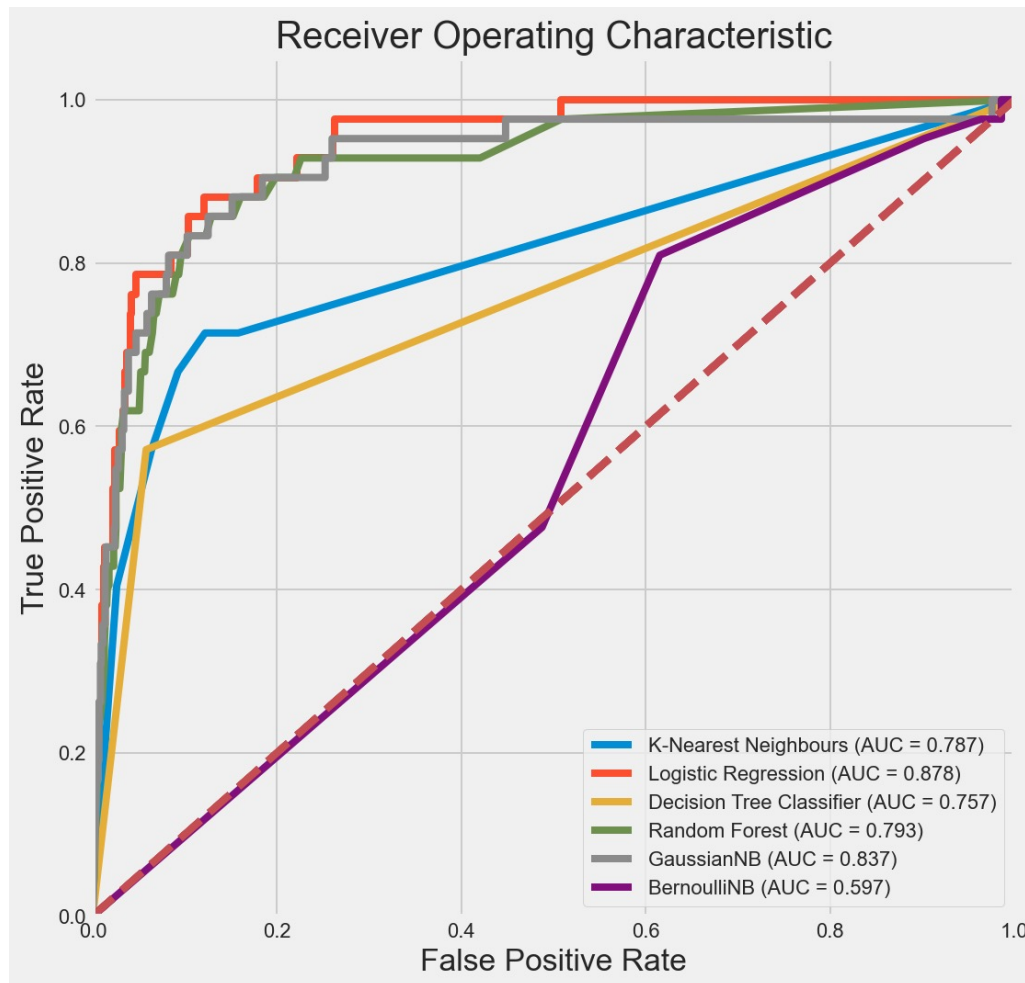
- Narrow down from 96 features to 31 features to corporate with S&P500 data

Exploratory Data Analysis



- True Class: Bankrupted
- False Class: No bankruptcy
- Goal: Minimize False Negative → Maximize Recall

$$Recall = \frac{TP}{TP + FN}$$



Naïve Result

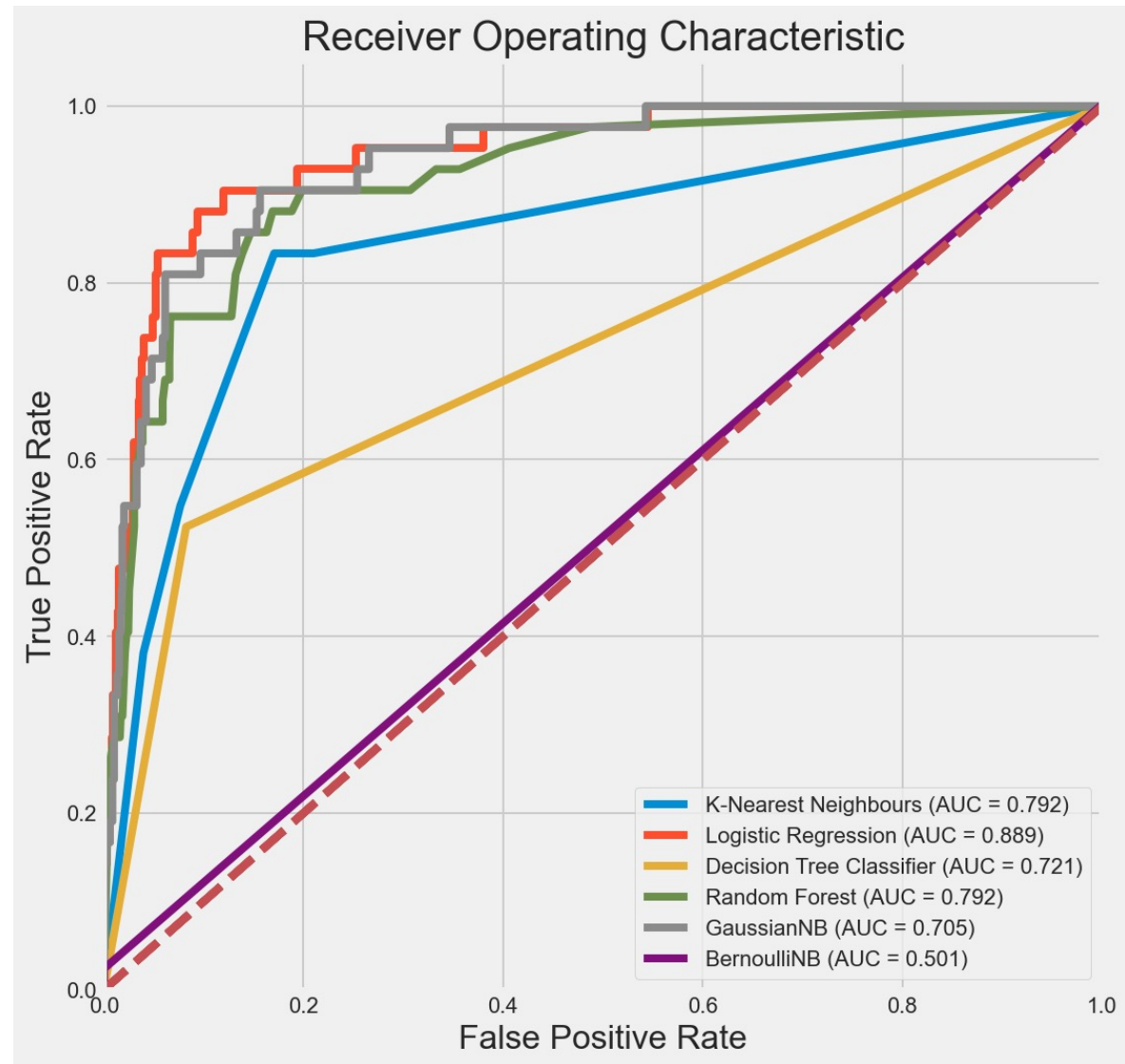
Top 2 Models

- Logistic Regression
- Gaussian Naive Bayes

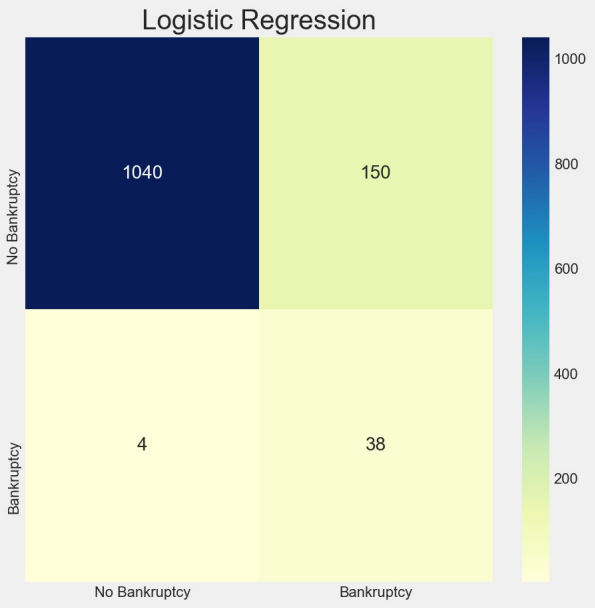
Algorithm	Train Score	Model Score	Recall	Precision	F1 score	F-Beta	ROC-AUC score
Logistic Regression	88.86000	87.5%	0.88000	0.20000	0.32000	0.32000	0.88000
GaussianNB	82.36000	72.89%	0.95000	0.11000	0.19000	0.19000	0.84000
K-Nearest Neighbours	96.47000	89.94%	0.67000	0.20000	0.31000	0.31000	0.79000
Random Forest	100.00000	95.05%	0.55000	0.35000	0.43000	0.43000	0.76000
Decision Tree Classifier	100.00000	92.45%	0.45000	0.21000	0.29000	0.29000	0.70000
BernoulliNB	68.02000	39.94%	0.81000	0.04000	0.08000	0.08000	0.60000




Tuned Result

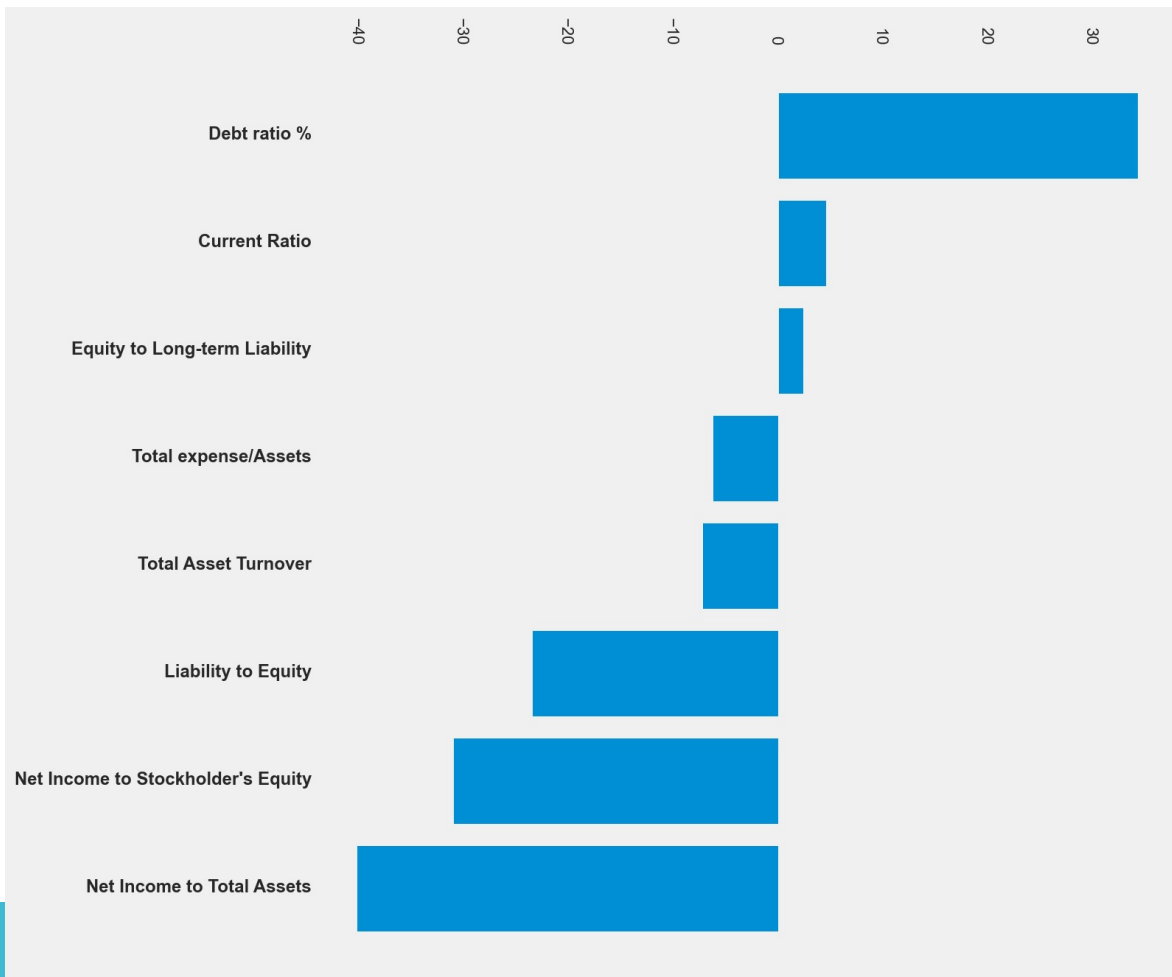
- Removing multicollinearity



Result



Model	Baseline	Tuned
Recall	88%	90%  0.02
Precision	20%	20%
F1	0.32	0.33  0.01
ROC_AUC	0.88	0.89  0.01



Feature	Feature Coefficient
Debt ratio %	34.34437
Current Ratio	4.62239
Equity to Long-term Liability	2.40398
Total expense/Assets	-6.22252
Total Asset Turnover	-7.189
Liability to Equity	-23.468
Net Income to Stockholder's Equity	-30.9835
Net Income to Total Assets	-40.1931

Logistic Regression

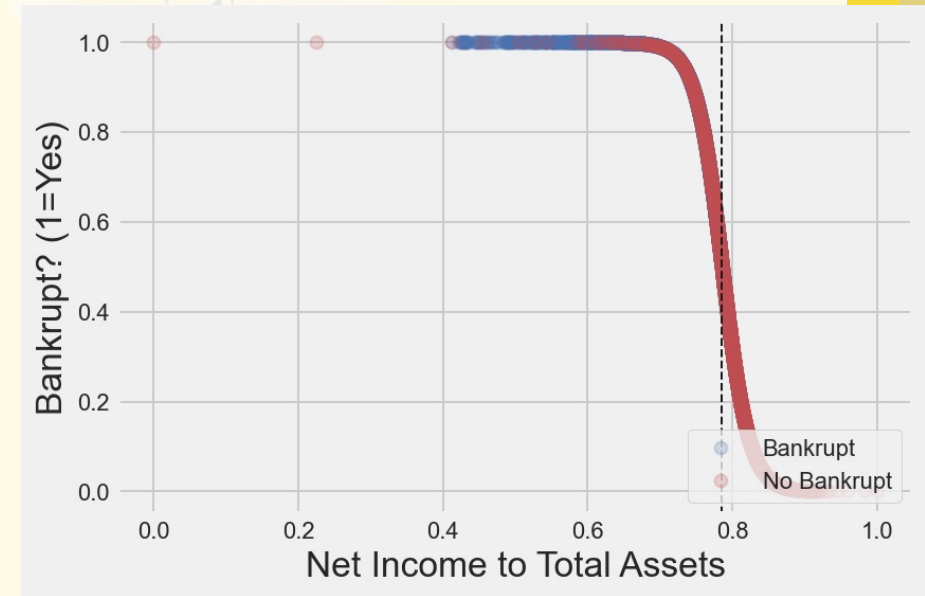
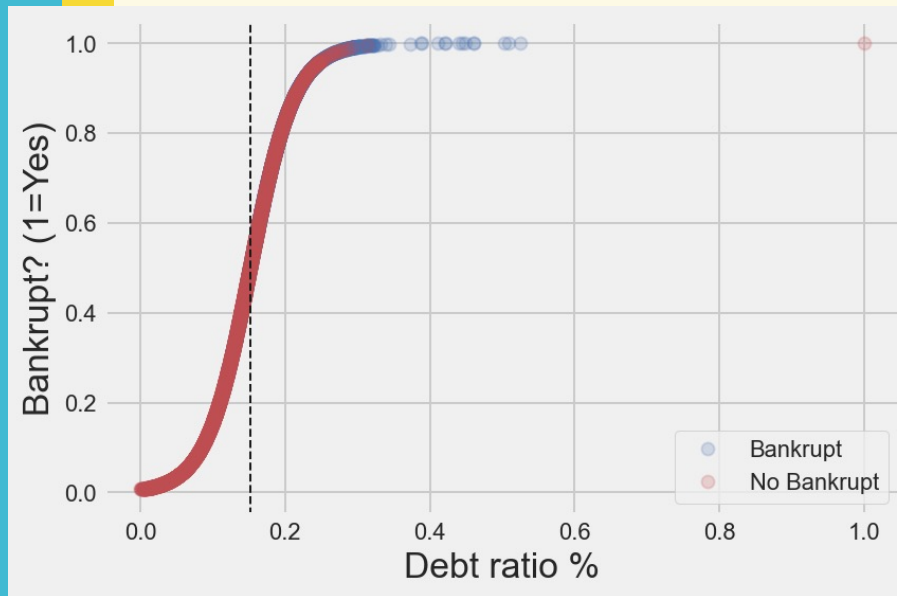
Reduced to 12 features

S&P500 Top identified companies

- TransDigm Group Inc
- MSCI Inc
- American Airlines Group Inc
- Kimberly-Clark Corp.
- Lamb Weston Holdings Inc

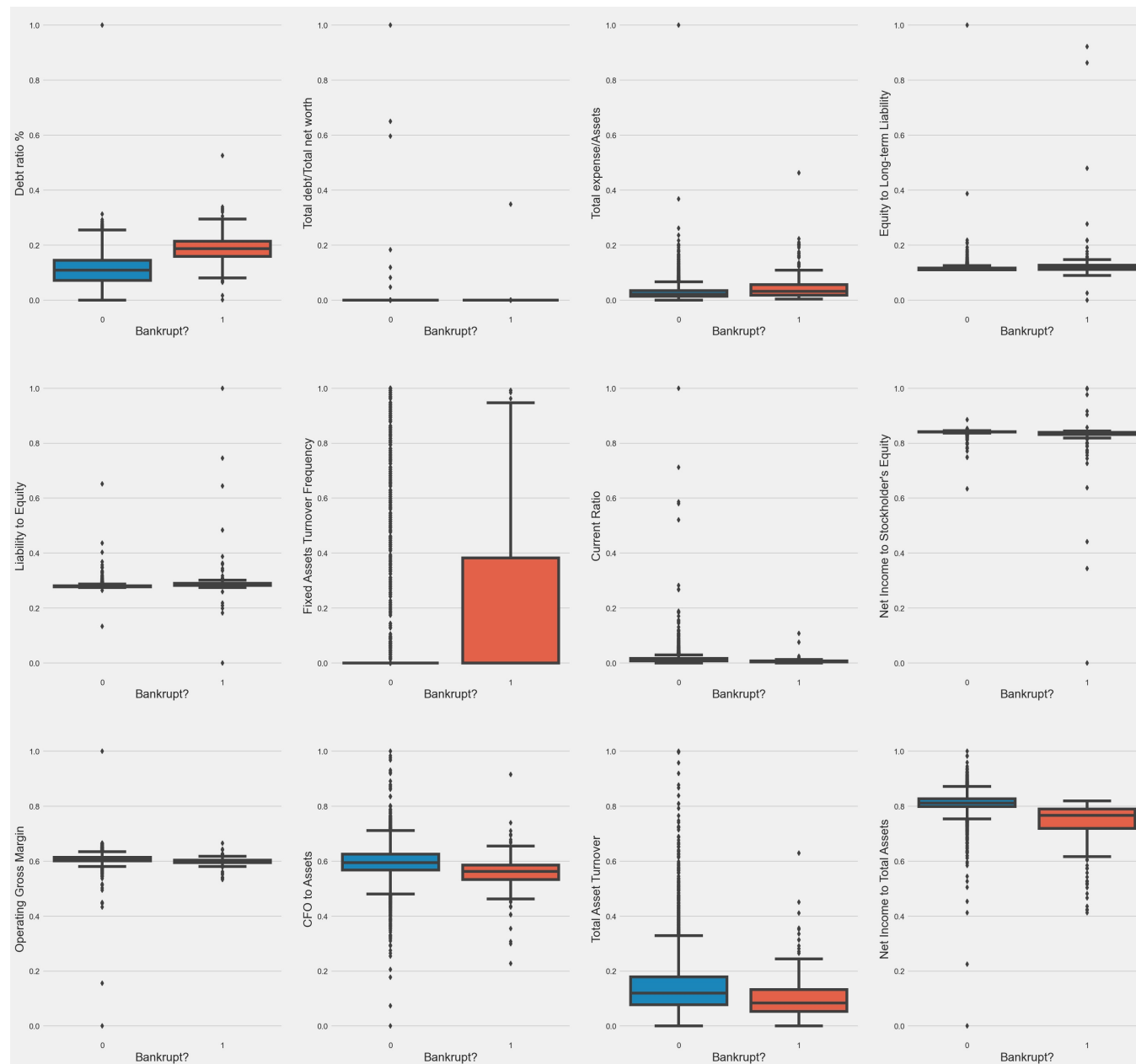
Future works

- Improving High False positive
- Improving the incorporate of SP500 data
- Try more sophisticated model techniques
 - XGBoost



Thank you

Feature Classification Performance



Pairplot of 12 features

