



Company Bankruptcy Prediction

Predicting bankruptcy risk through
Machine Learning Classification

Cindy Su - 10/29/2021

Introduction

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

Data & Methodology



Data & Methodology



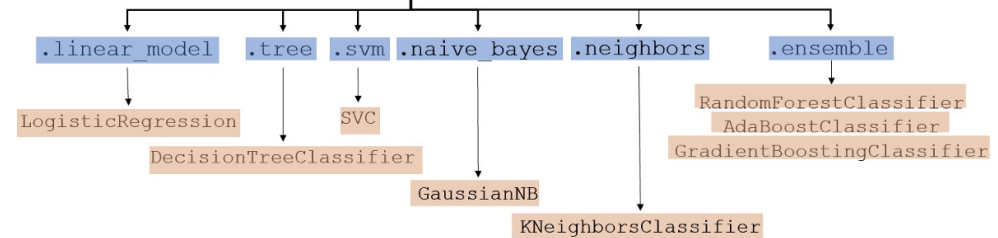
- Taiwan Economic Journal
- 800-900 listed companies
- From 1999 to 2009
- rows: 6819
- features: 96



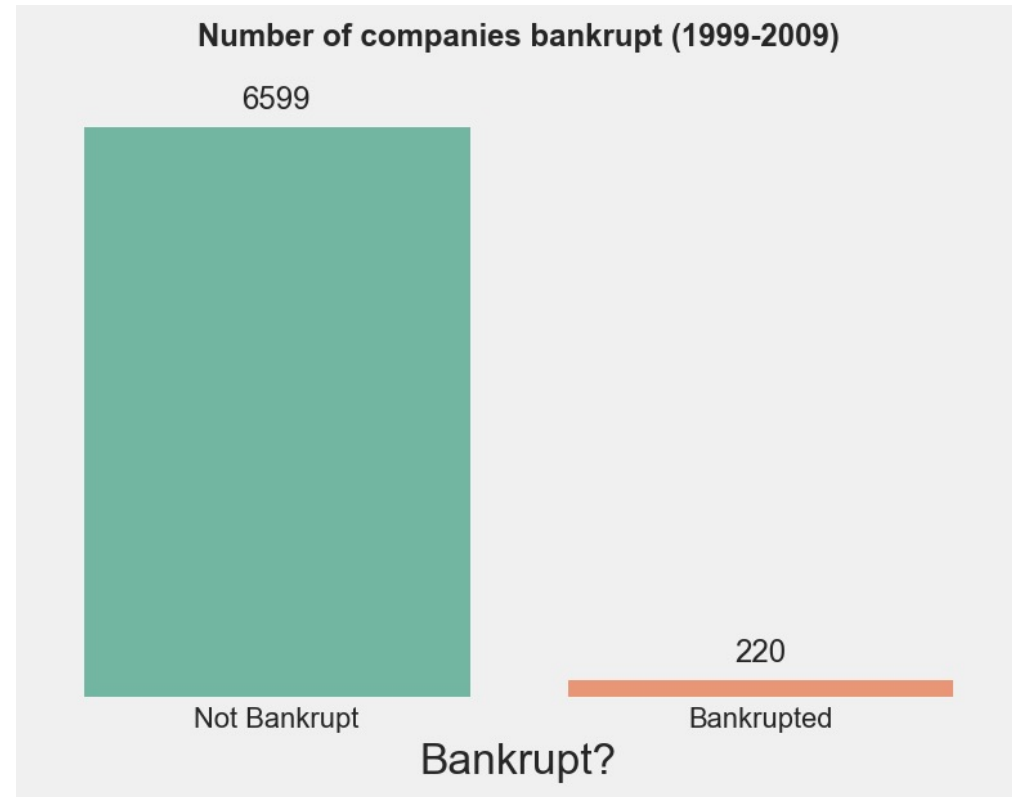
- S&P500



- Feature selection & Dimensionality reduction

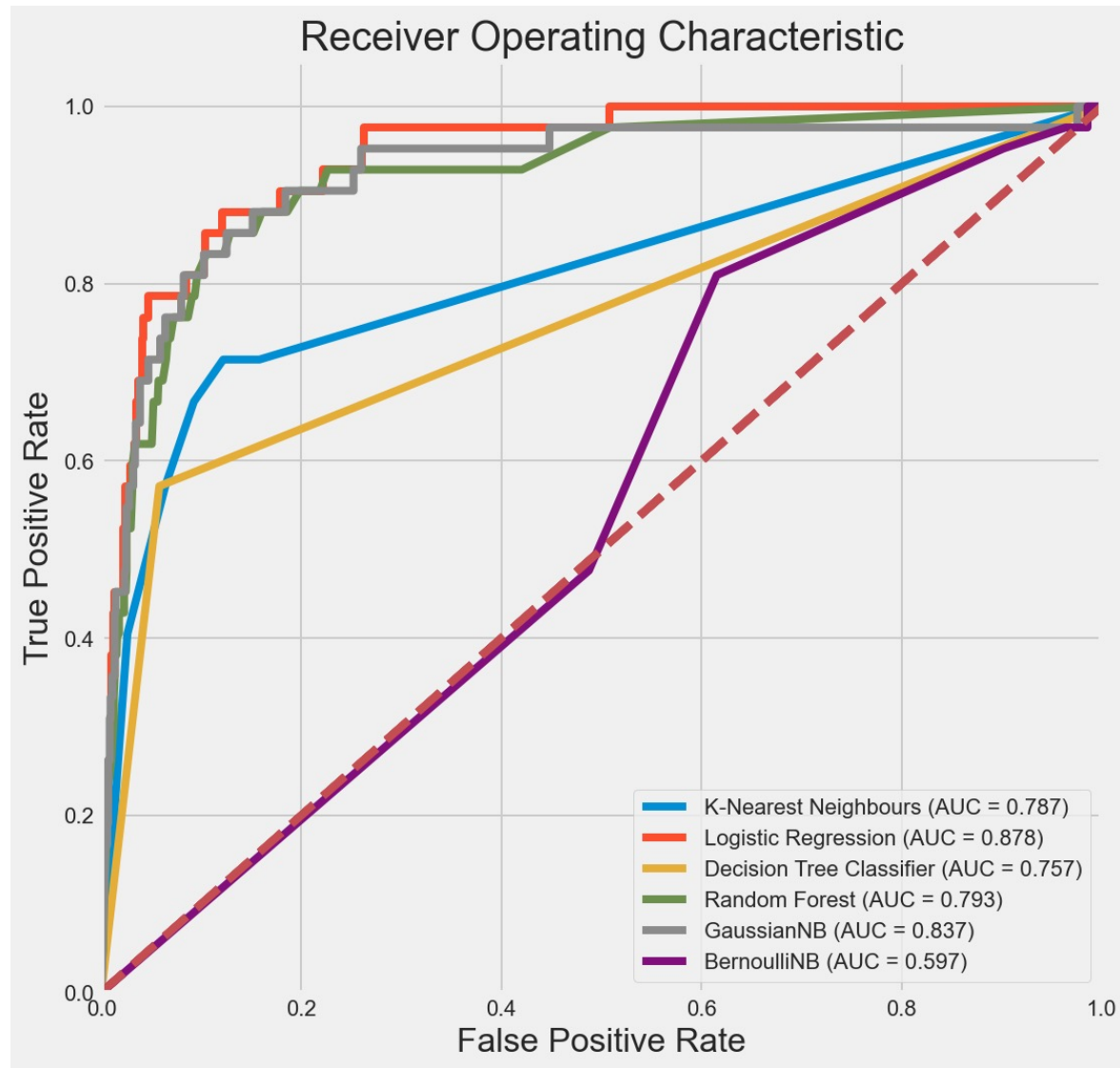


Exploratory Data Analysis



- True Class: Bankrupted
- False Class: No bankrupt
- Goal: Minimize False Negative ➔ Maximize Recall

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



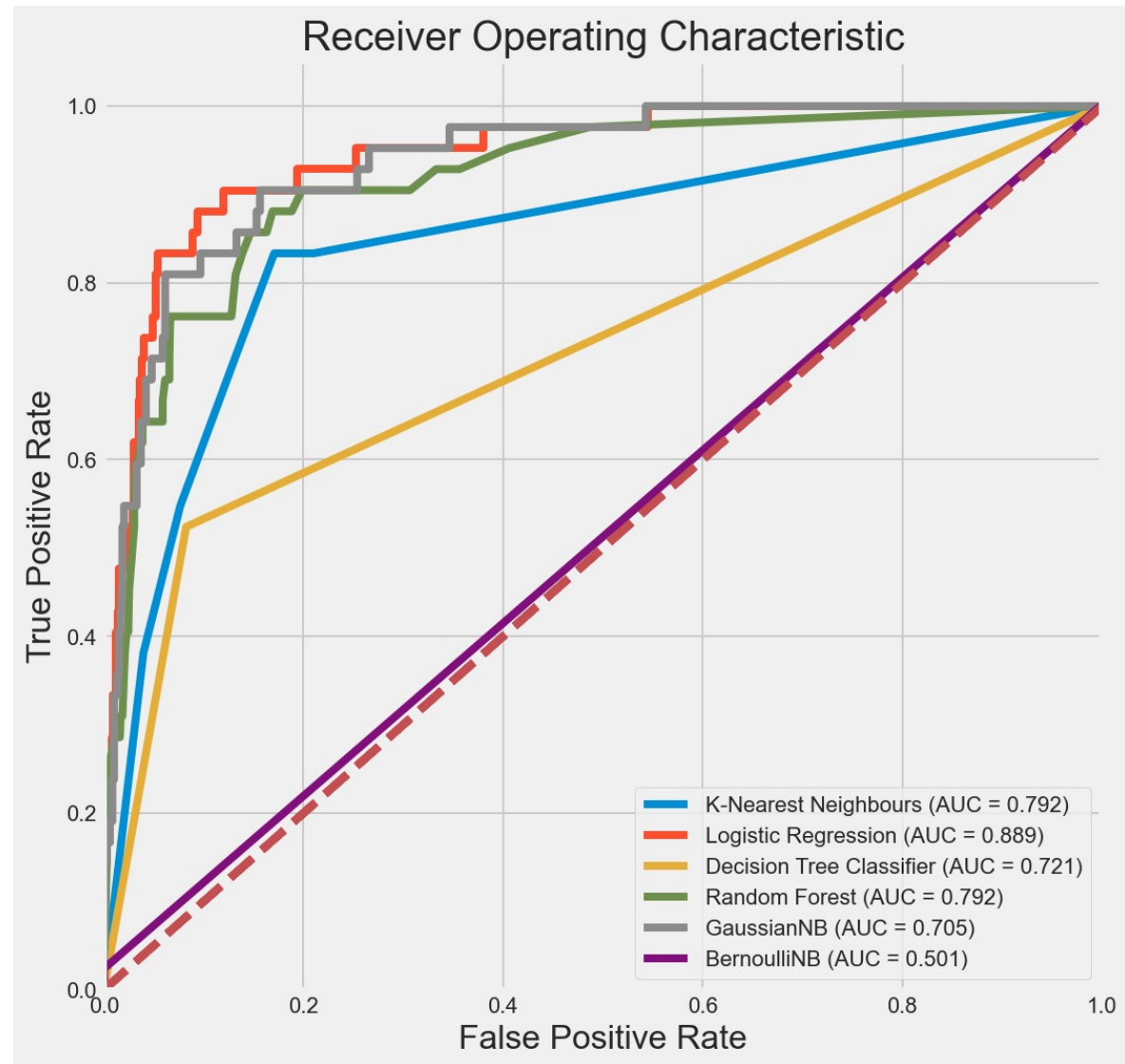
Naïve Result

Top 2 Models

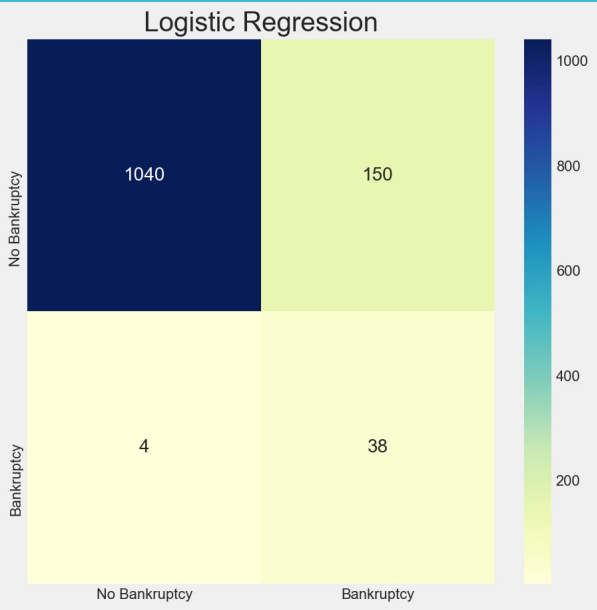
- Logistic Regression
- Gaussian Naive Bayes




Tuned Result

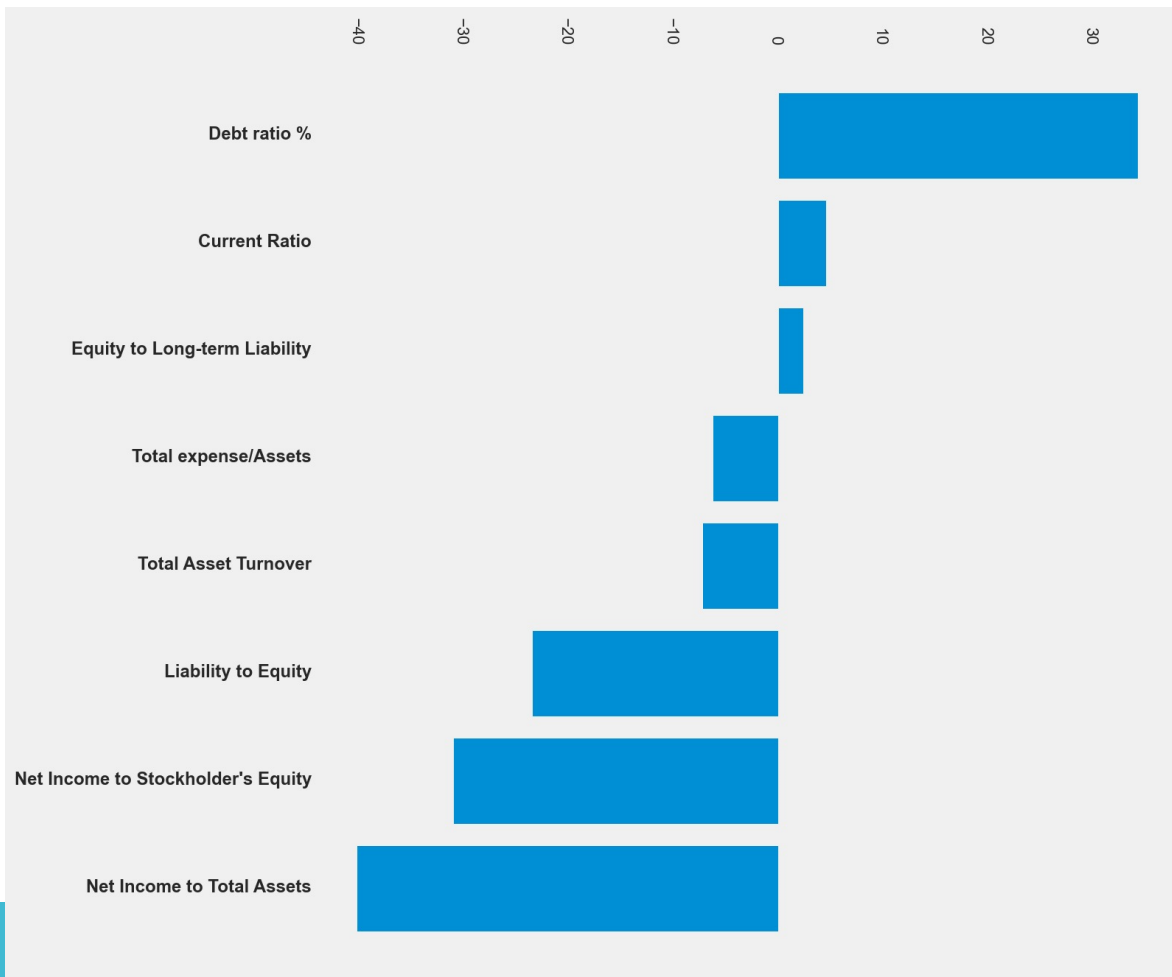
- Removing multicollinearity



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



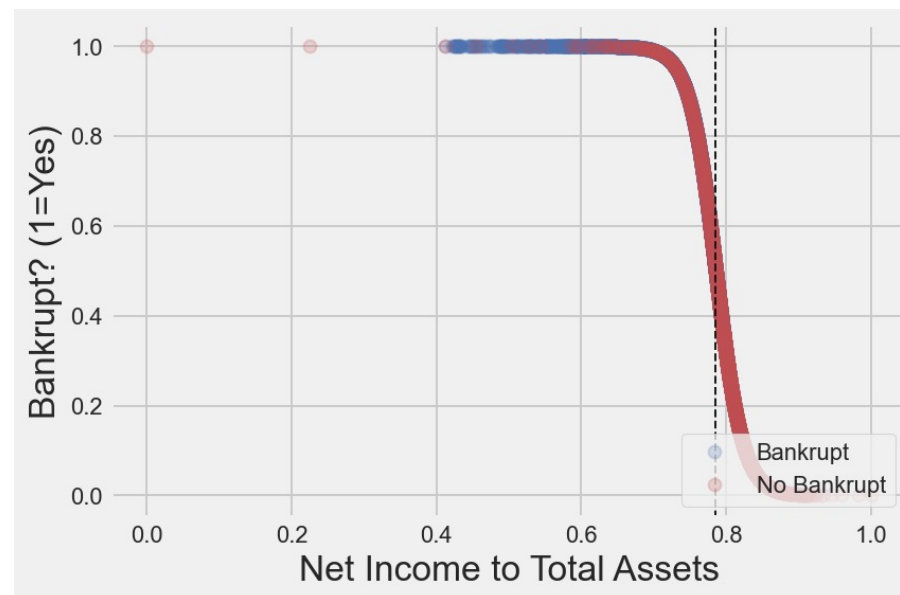
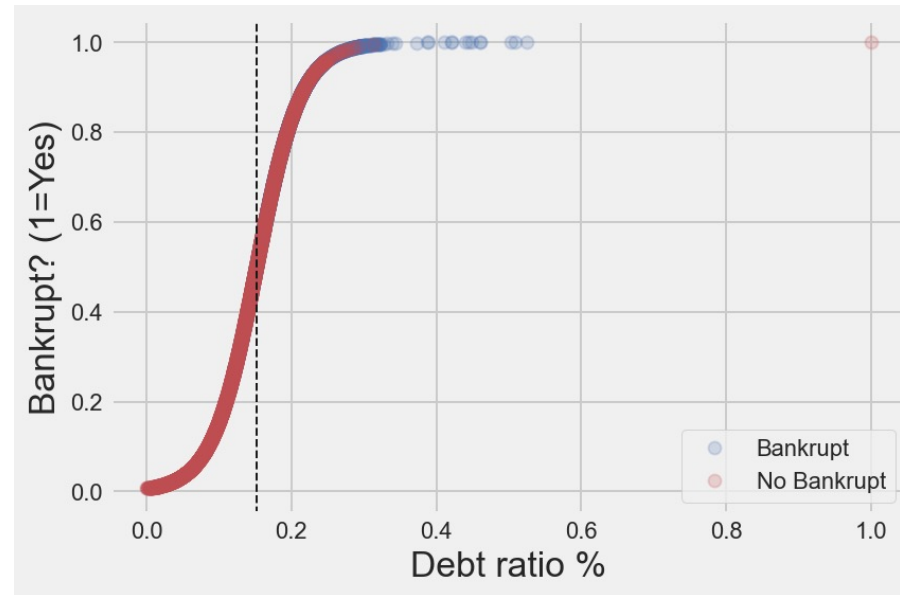
Feature	Feature Coefficient
Debt ratio %	34.34437
Current Ratio	4.62239
Equity to Long-term Liability	2.40398
Total expense/Assets	-6.22252
Total Asset Turnover	-7.189
Liability to Equity	-23.468
Net Income to Stockholder's Equity	-30.9835
Net Income to Total Assets	-40.1931

Logistic Regression

Reduced to 12 features

Tuned Result

- Removing multicollinearity

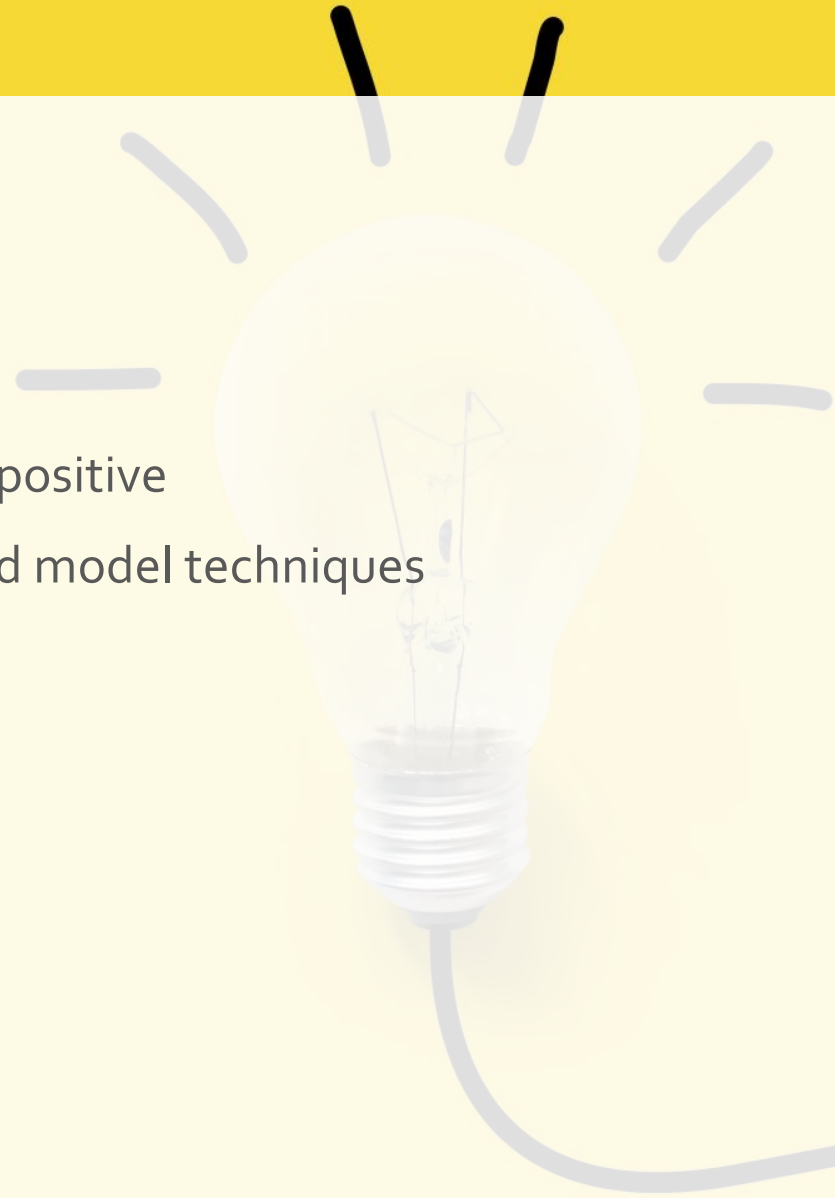


S&P500 Top identified companies

- Diamondback Energy, Inc.
- SBA Communications Corp (New)
- AutoZone, Inc
- Autodesk Inc

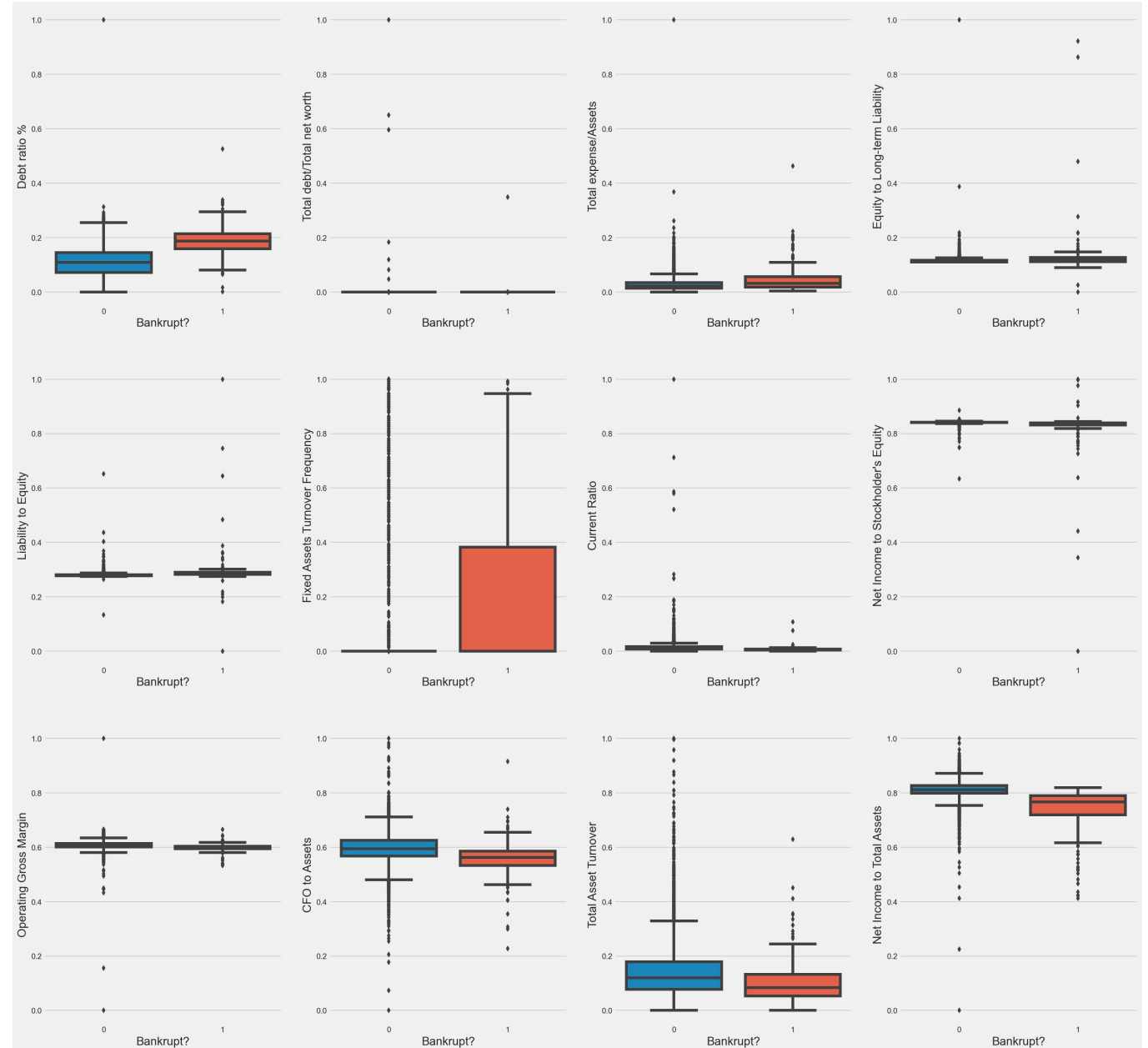
Future works

- Improving High False positive
- Try more sophisticated model techniques
 - XGBoost



Thank you

Feature Classification Performance



Pairplot of 12 features

