Bankruptcies Sprint 3

Mark Stansky
Brainstation NYC

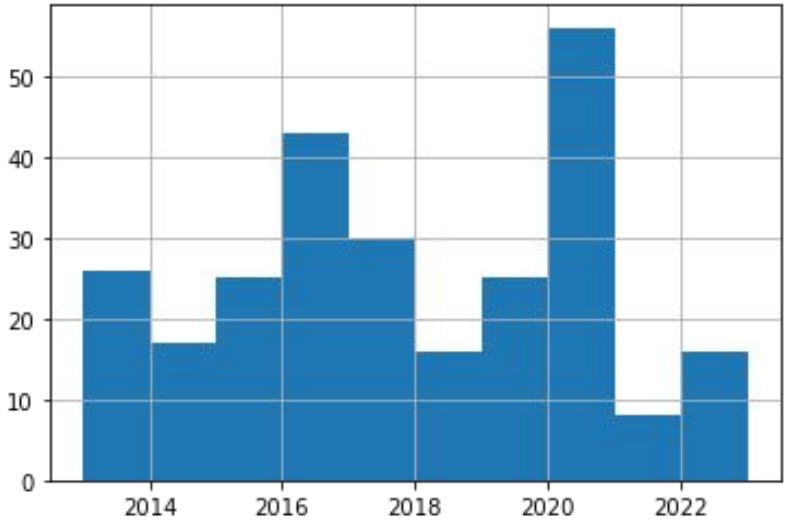
Non-Technical Overview

- Non-Technical Overview: Predicting Future Bankruptcies in the S&P 500

- Proposed Solution:
  - Use ML techniques to better understand and categorize past bankruptcies
  - Use Logistic Regression to identify potential bankruptcies among public companies
- Estimate of potential impact:
  - Support both policymakers and investors to take appropriate steps to reduce losses
  - Early warning for operators

Bankruptcies

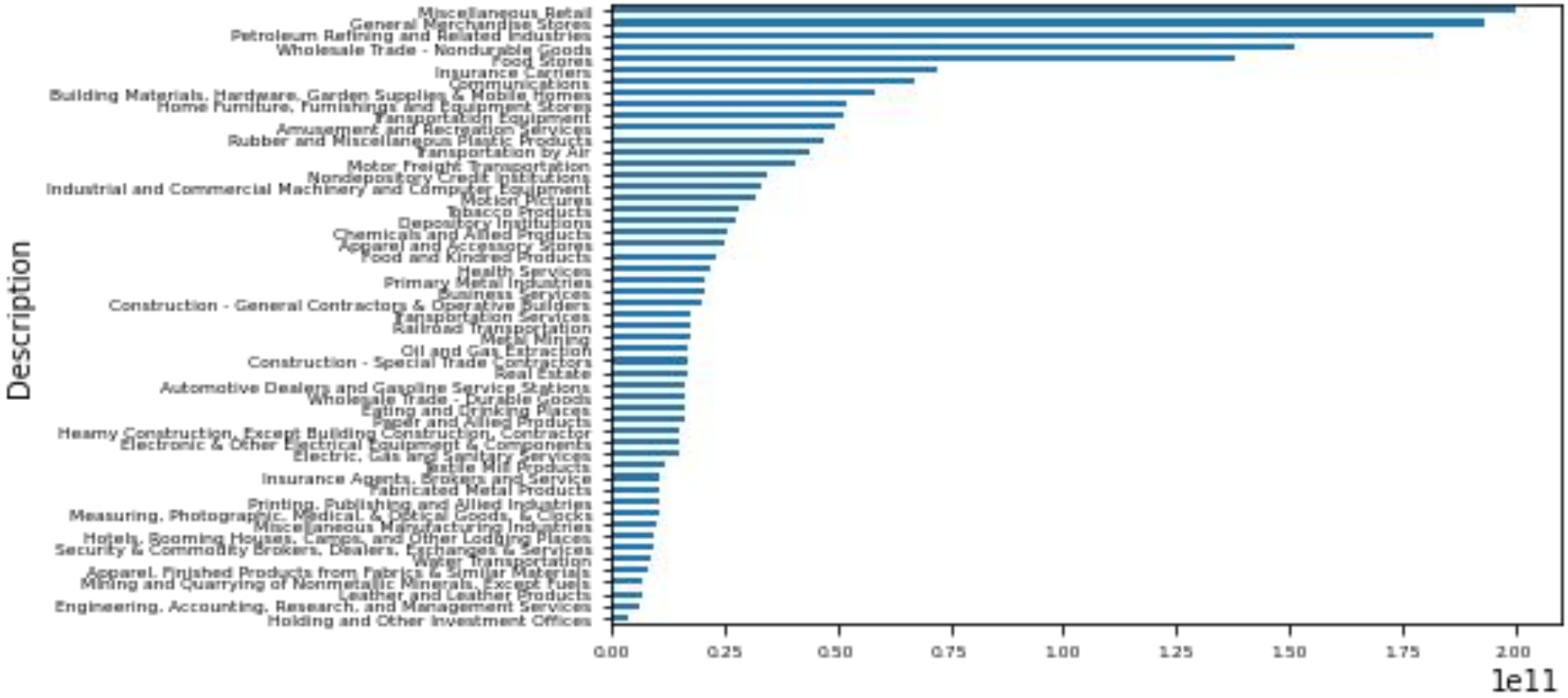
- 86 data points ranging from 2013-2023



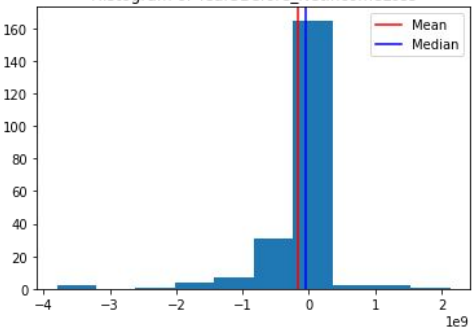
S&P 500

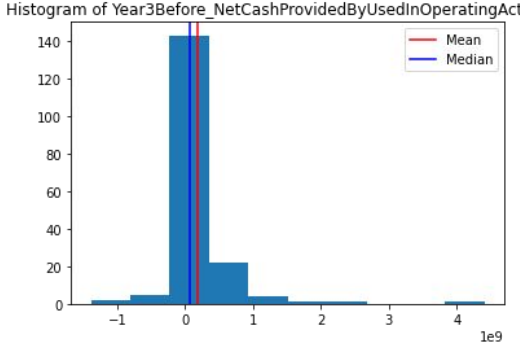
- 448 (complete) data points

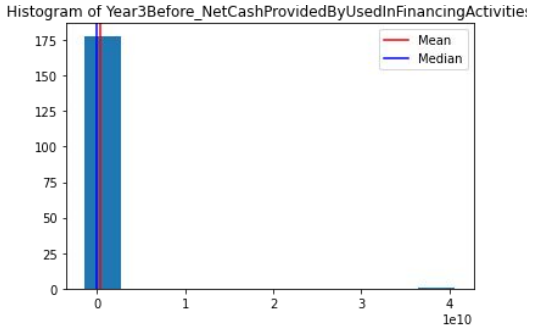
S&P 500 Data - Average Revenue by Industry



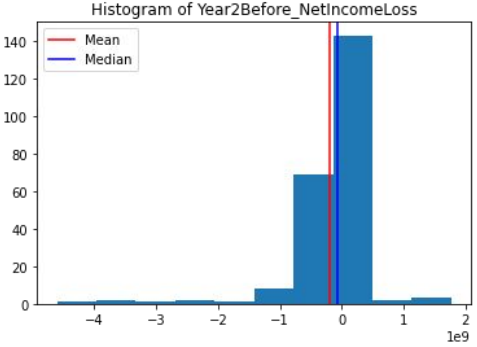
3

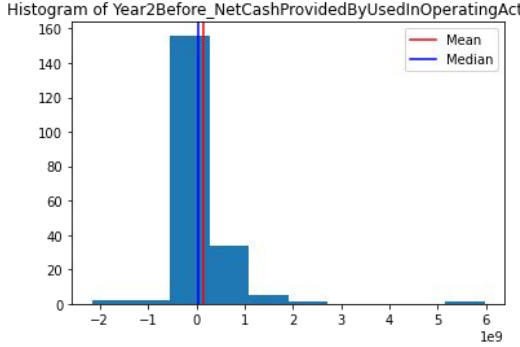

Histogram of Year3Before_NetIncomeLoss
Mean
Median
1e9

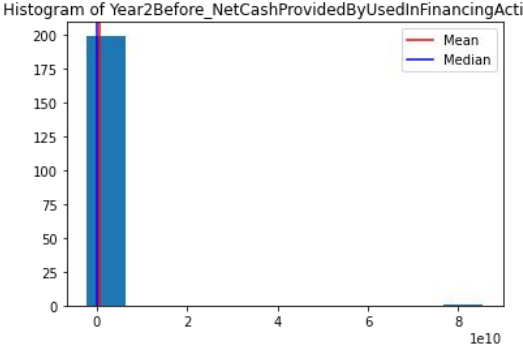

Histogram of Year3Before_NetCashProvidedByUsedInOperatingActivities
Mean
Median
1e9


Histogram of Year3Before_NetCashProvidedByUsedInFinancingActivitie
Mean
Median
1e10

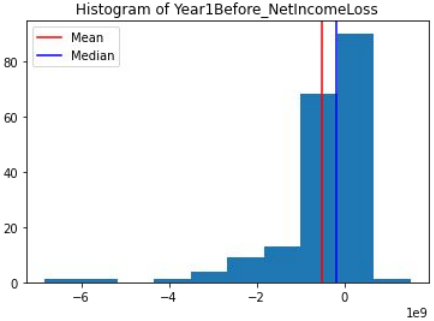
2

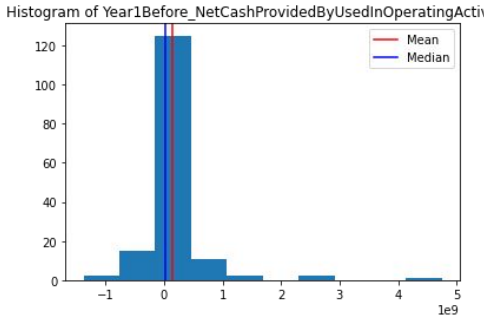

Histogram of Year2Before_NetIncomeLoss
Mean
Median
1e9

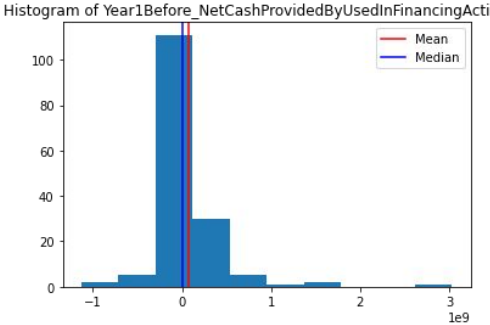

Histogram of Year2Before_NetCashProvidedByUsedInOperatingActivities
Mean
Median
1e9


Histogram of Year2Before_NetCashProvidedByUsedInFinancingActivities
Mean
Median
1e10

1

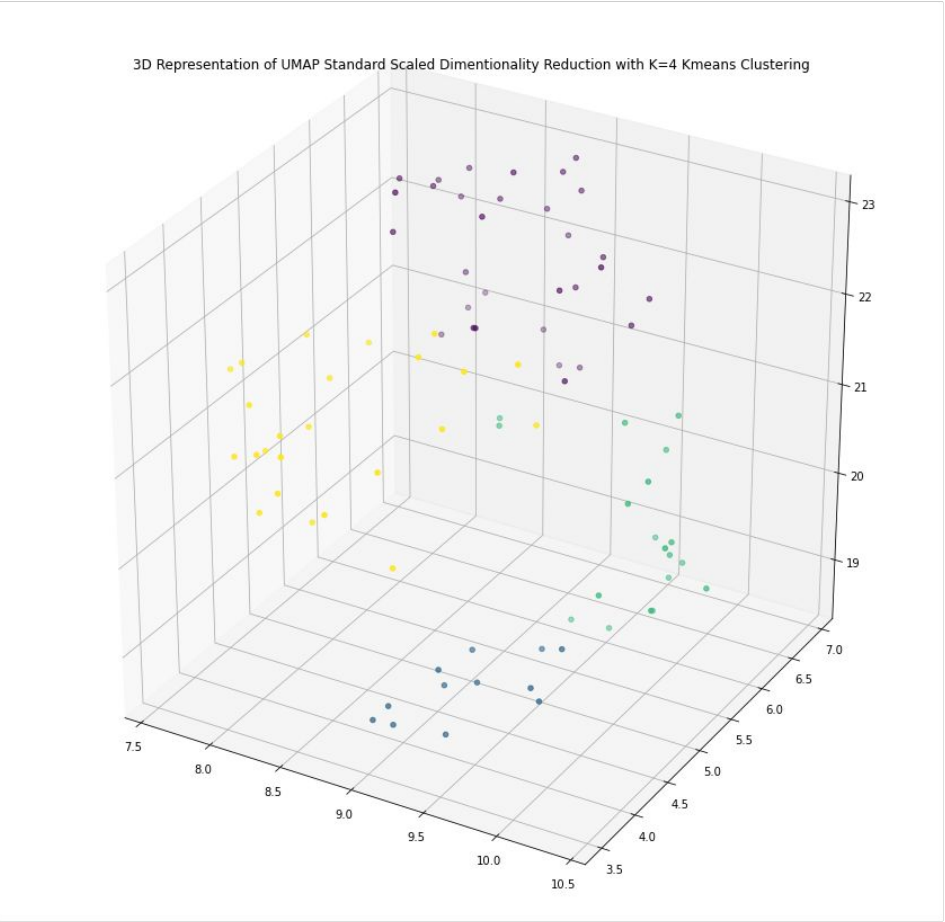

Histogram of Year1Before_NetIncomeLoss
Mean
Median
1e9


Histogram of Year1Before_NetCashProvidedByUsedInOperatingActivities
Mean
Median
1e9


Histogram of Year1Before_NetCashProvidedByUsedInFinancingActivities
Mean
Median
1e9

Clustering Analysis

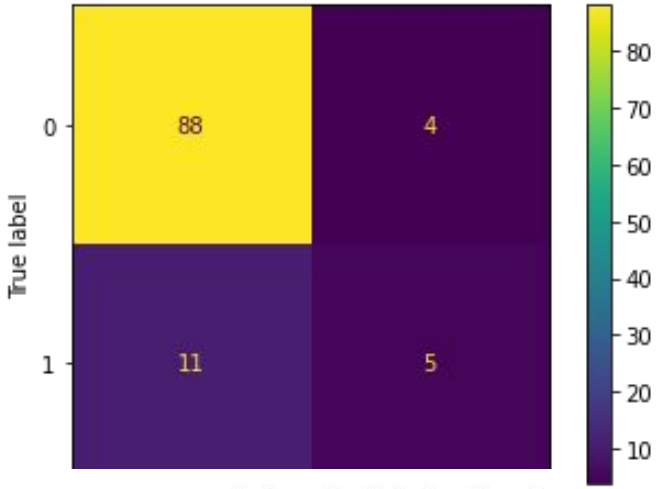
- Group 0:
  - Small companies in size
- Group 1:
  - Large Companies
  - Defined by highest median revenue, assets, etc.
  - Pos. Op CF
- Group 2:
  - Larger Historical Losses, but better in recent term
- Group 3:
  - Small Revenues, Large Losses
  - Pos. Op CF
  - Young companies?



Combined Data Logistic Regression

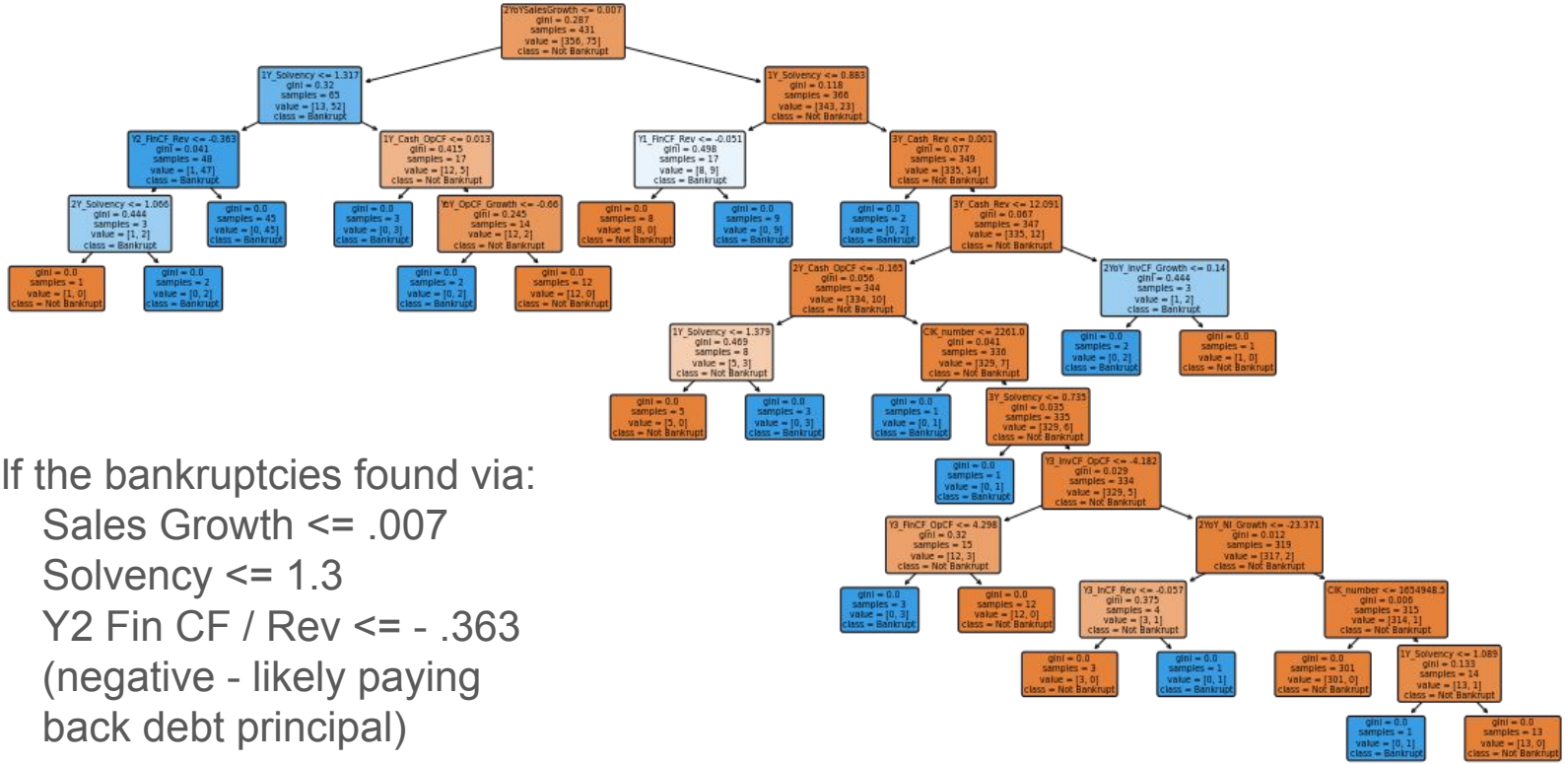
- 11/16 classified as False Negatives
- F-1 of 0.81
- Accuracy not a good metric
  - Missed 11 / 16 metrics
- Recall: .06 - not a good classifier
  - Bankruptcy is hard to spot!
  - Model not very sensitive

- Note: my confusion matrix got worse with class_weight parameter added



	Feature	Coefficient	odds_ratio
2	2YoYSalesGrowth	0.002609	1.002613
14	Y2_InvCF_OpCF	0.001041	1.001041
15	Y3_InvCF_OpCF	0.000946	1.000946
13	Y1_InvCF_OpCF	0.000498	1.000498
11	Y2_FinCF_Rev	0.000483	1.000483
7	Y1_InCF_Rev	0.000438	1.000438
23	YoY_FinCF_Growth	0.000354	1.000354
8	Y2_InCF_Rev	0.000299	1.000299
24	2YoY_FinCF_Growth	0.000245	1.000245
22	2YoY_InvCF_Growth	0.000183	1.000183

Decision Tree Analysis



Half the bankruptcies found via:

- Sales Growth <= .007
- Solvency <= 1.3
- Y2 Fin CF / Rev <= - .363 (negative - likely paying back debt principal)

Key Insights Discovered

- Clustering: Grouped companies largely by size
  - Interesting nuance between those with Positive Operating CF vs not
- Logistic Regression Predictions:
  - Surprised by the lack of predictiveness
  - Ratio of Operating CF to Investing CF as 2nd-5th most predictive interesting

Product Demo

- Likely going to manifest itself as a ‘research report’ in Tableau
- Run a few more modeling techniques - which works best?
- Would love to create a dropdown in Streamlit for SP 500 to evaluate company on the decision tree weights and otherwise quantify the risk of bankruptcy