

# Industry Classification

Jen Hilibrand



# The Industry Classification Problem

- Do fundamentals alone have predictive power?
- Can certain types of misclassification inform us on industry patterns?
- Is there a systematic way to generate closest peers?

# Workflow

yahoo!  
finance



# Data

- Russell 3000 Constituents

## Features

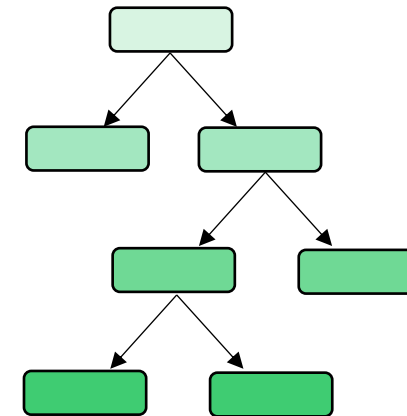
- Market Cap
- Dividend Rate
- Price to Book
- EV to Revenue
- EV to EBITDA
- Trailing EPS
- Profit Margin
- Implied Vol

## Target Categories

- Basic Materials
- Communication Services
- Consumer Cyclical
- Consumer Defensive
- Energy
- Financial Services
- Healthcare
- Industrials
- Real Estate
- Technology
- Utilities



**XGBClassifier** (`n_estimators = 100`,  
`learning_rate=0.4`, `objective="reg:squarederror"`,  
`max_depth=3`, `subsample=1`, `gamma = 0`,  
`colsample_bytree=0.4`)



## Challenges

- 11 categories
- 3k data points
- Sporadic nulls

- Grid search for hyperparameters



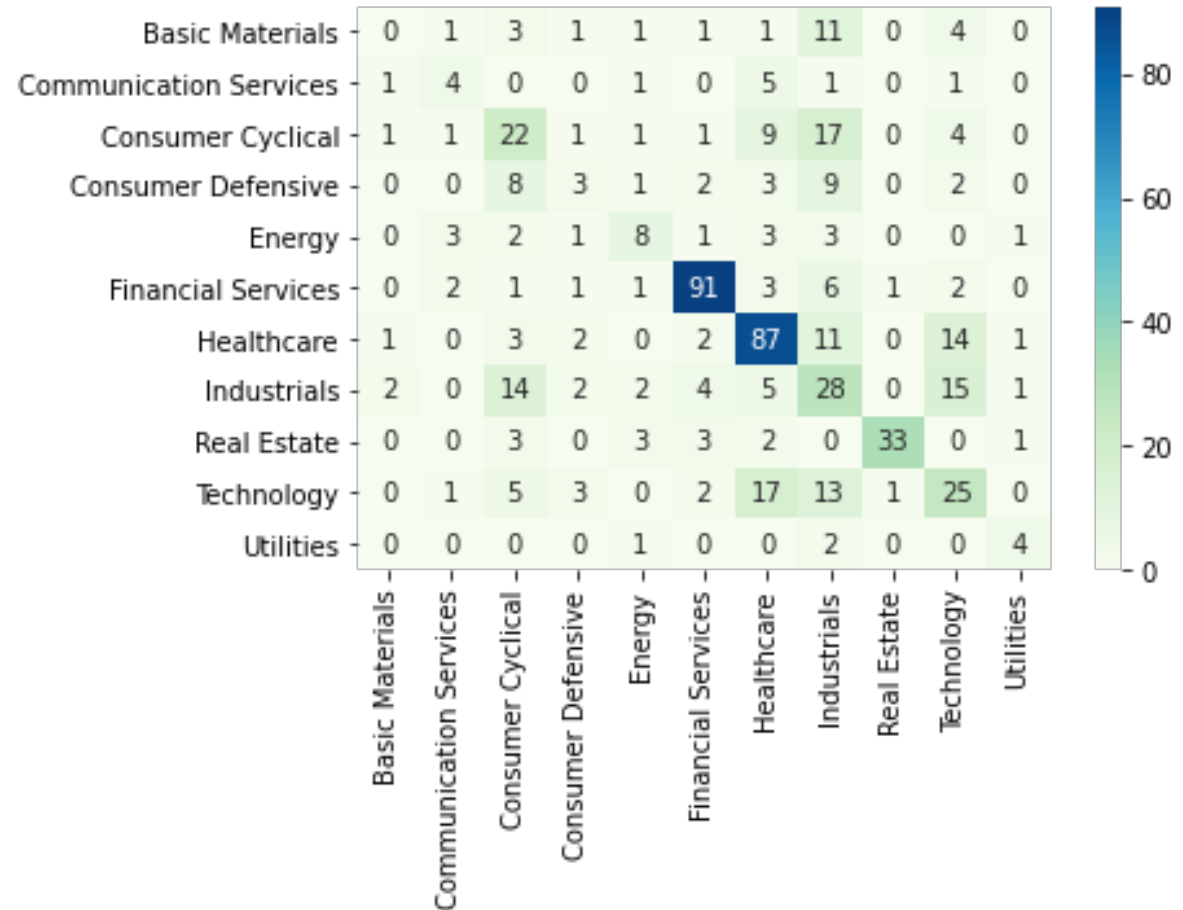
# How the model performs

## Misclassifications:

- Consumer Cyclical and Industrial misclassifications
- Technology and Healthcare misclassifications

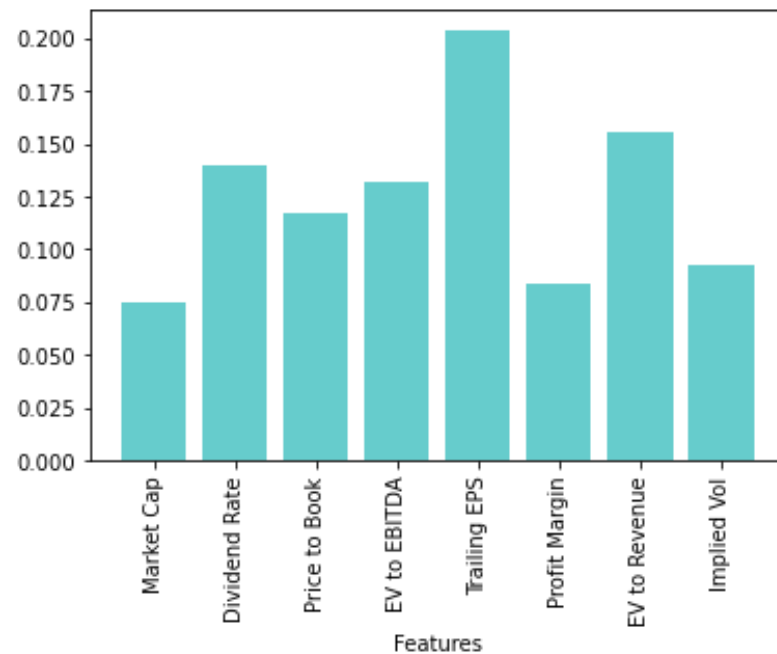
## Low Recall:

- Basic Materials
  - 0/23 first rank
  - 2/23 second rank
- Consumer Defensive
  - 3/28 first rank
  - 6/28 second rank



## How the model performs

- Accuracy for first prediction: .54
- "Top 2" Accuracy: .69



	Recall	Precision
Basic Materials	0.00	0.00
Communication Services	0.31	0.33
Consumer Cyclical	0.39	0.36
Consumer Defensive	0.11	0.21
Energy	0.36	0.42
Financial Services	0.84	0.85
Healthcare	0.72	0.64
Industrials	0.38	0.28
Real Estate	0.73	0.94
Technology	0.37	0.37
Utilities	0.57	0.50

	Dividend Rate	Price to Book	EV to Revenue	EV to EBITDA	Trailing EPS	Profit Margin	Implied Vol	Market Cap (\$Bn)
true_sector								
Basic Materials	0.98	3.93	2.79	16.44	2.18	0.03	0.42	9.24
Communication Services	0.30	4.09	19.55	-157.32	2.57	-0.07	0.52	62.47
Consumer Cyclical	0.54	11.86	10.64	18.48	6.06	0.01	0.46	16.94
Consumer Defensive	1.00	8.85	2.82	15.10	3.46	0.01	0.33	23.89
Energy	0.50	3.69	8.61	-47.91	-3.25	-0.20	0.58	11.38
Financial Services	0.95	2.34	3.50	14.90	5.16	0.25	0.40	11.55
Healthcare	0.18	8.74	630.85	-0.45	-0.72	-0.22	0.74	10.13
Industrials	0.78	6.99	2.40	89.88	3.18	0.01	0.42	10.48
Real Estate	1.47	3.81	14.03	15.47	1.03	0.13	0.37	23.10
Technology	0.36	13.10	16.19	-38.98	1.28	0.00	0.44	32.42
Utilities	1.70	2.44	20.68	13.44	2.52	0.10	0.28	15.55



	Dividend Rate	Price to Book	EV to Revenue	EV to EBITDA	Trailing EPS	Profit Margin	Implied Vol	Market Cap (\$Bn)
true_sector								
Basic Materials	0.98	3.93	2.79	16.44	2.18	0.03	0.42	9.24
Communication Services	0.30	4.09	19.55	-157.32	2.57	-0.07	0.52	62.47
Consumer Cyclical	0.54	11.86	10.64	18.48	6.06	0.01	0.46	16.94
Consumer Defensive	1.00	8.85	2.82	15.10	3.46	0.01	0.33	23.89
Energy	0.50	3.69	8.61	-47.91	-3.25	-0.20	0.58	11.38
Financial Services	0.95	2.34	3.50	14.90	5.16	0.25	0.40	11.55
Healthcare	0.18	8.74	630.85	-0.45	-0.72	-0.22	0.74	10.13
Industrials	0.78	6.99	2.40	89.88	3.18	0.01	0.42	10.48
Real Estate	1.47	3.81	14.03	15.47	1.03	0.13	0.37	23.10
Technology	0.36	13.10	16.19	-38.98	1.28	0.00	0.44	32.42
Utilities	1.70	2.44	20.68	13.44	2.52	0.10	0.28	15.55



	Dividend Rate	Price to Book	EV to Revenue	EV to EBITDA	Trailing EPS	Profit Margin	Implied Vol	Market Cap (\$Bn)
true_sector								
Basic Materials	0.98	3.93	2.79	16.44	2.18	0.03	0.42	9.24
Communication Services	0.30	4.09	19.55	-157.32	2.57	-0.07	0.52	62.47
Consumer Cyclical	0.54	11.86	10.64	18.48	6.06	0.01	0.46	16.94
Consumer Defensive	1.00	8.85	2.82	15.10	3.46	0.01	0.33	23.89
Energy	0.50	3.69	8.61	-47.91	-3.25	-0.20	0.58	11.38
Financial Services	0.95	2.34	3.50	14.90	5.16	0.25	0.40	11.55
Healthcare	0.18	8.74	630.85	-0.45	-0.72	-0.22	0.74	10.13
Industrials	0.78	6.99	2.40	89.88	3.18	0.01	0.42	10.48
Real Estate	1.47	3.81	14.03	15.47	1.03	0.13	0.37	23.10
Technology	0.36	13.10	16.19	-38.98	1.28	0.00	0.44	32.42
Utilities	1.70	2.44	20.68	13.44	2.52	0.10	0.28	15.55

# Generating Most Similar Peers

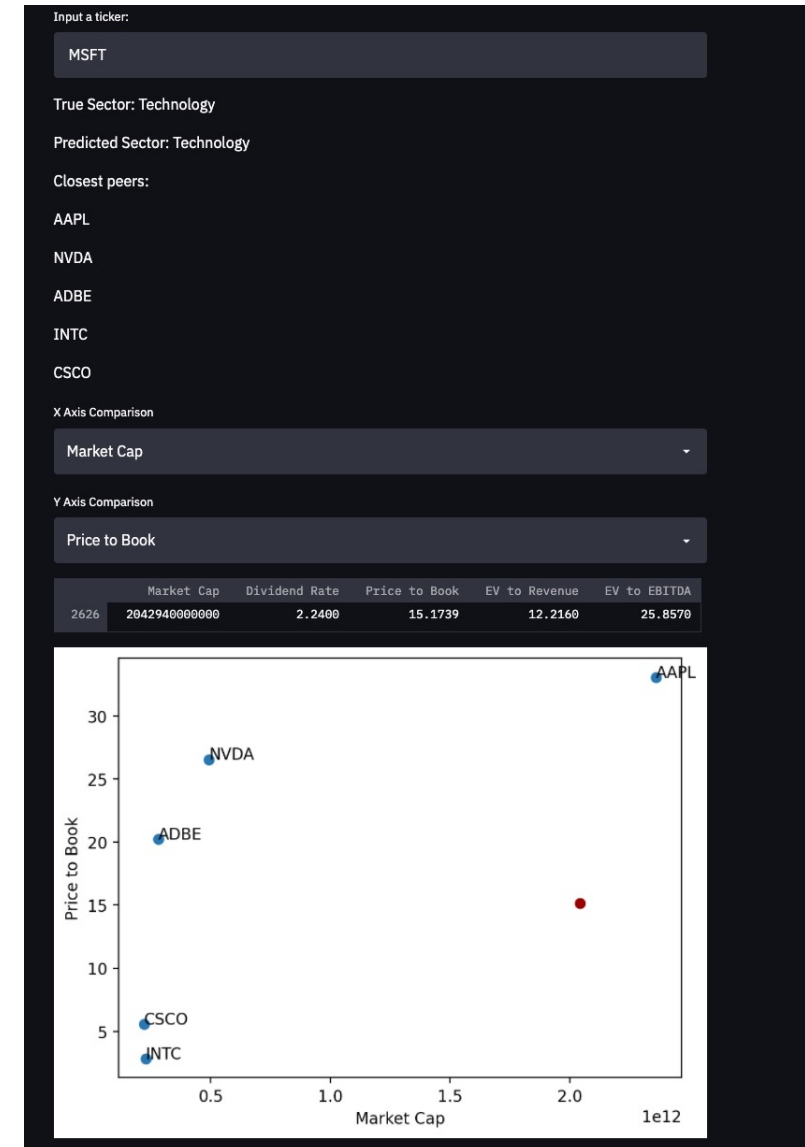
Scale data



Filter on sector classification



Euclidean distance across features



## Future work

- More feature engineering
- Further anomaly detection
- Predictive power for other financial targets?