Predict Stock Price From LinkedIn Profiles
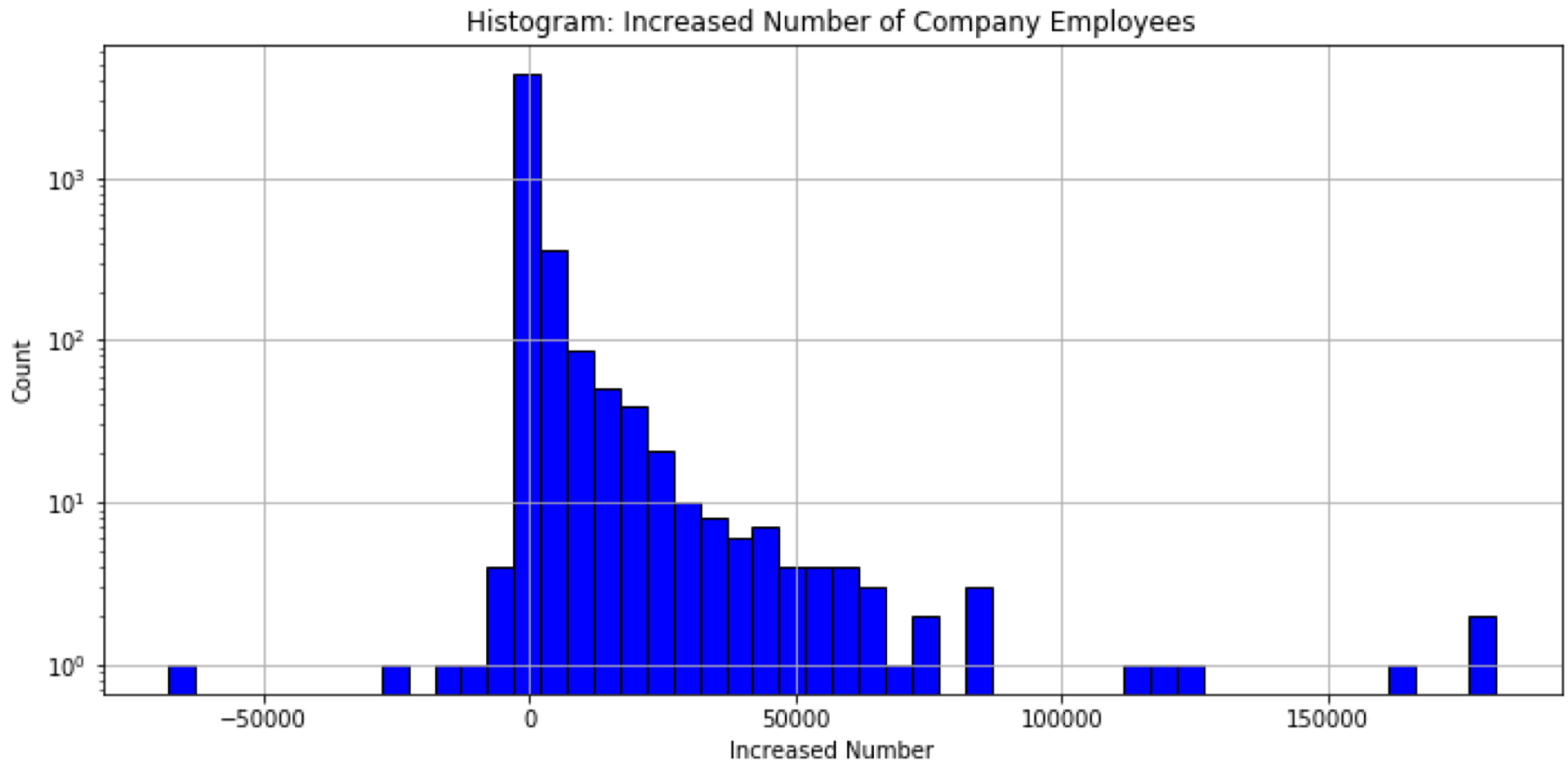
Yichao Zhang


VS

Companies Expansion
# 2015 - 2018

Histogram: Increased Number of Company Employees

Top 20 Companies

company_name	expand_num	expand_ratio
Walmart	181442.0	0.790146
Amazon	176769.0	2.087716
IBM	163813.0	0.395551
Accenture	122157.0	0.473460
McDonald's	119906.0	1.160497
Marriott International	115879.0	1.861959
Cognizant	85812.0	0.543995
Ford Motor Company	84251.0	1.024702
Apple	82126.0	0.790251
Google	72997.0	1.078529
JPMorgan Chase & Co.	71919.0	0.379084
HSBC	69610.0	0.501513
Vodafone	65685.0	0.839307
Citi	63873.0	0.441759
Bank of America	63626.0	0.321339
Sprint	61329.0	1.599484
Starbucks	60479.0	1.015907
GE	58862.0	0.332875
AT&T	58101.0	0.337875
Wells Fargo	56078.0	0.303592

Stock vs Number of Employees: Amazon

Number of Employees & Stock Price

Correlation Coefficients vs **Leading Days:** Amazon
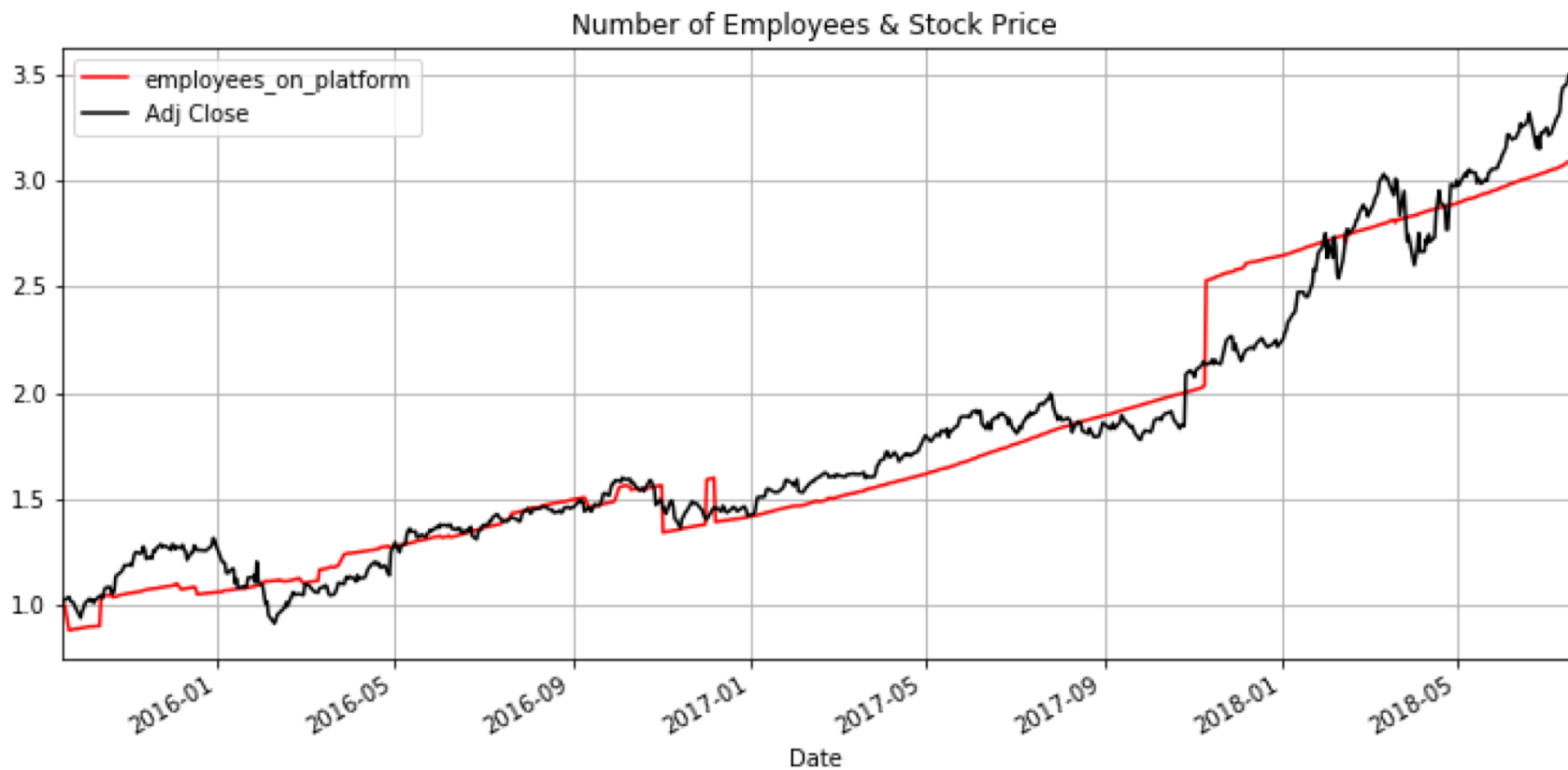
Correlation Coefficients vs Leading Days
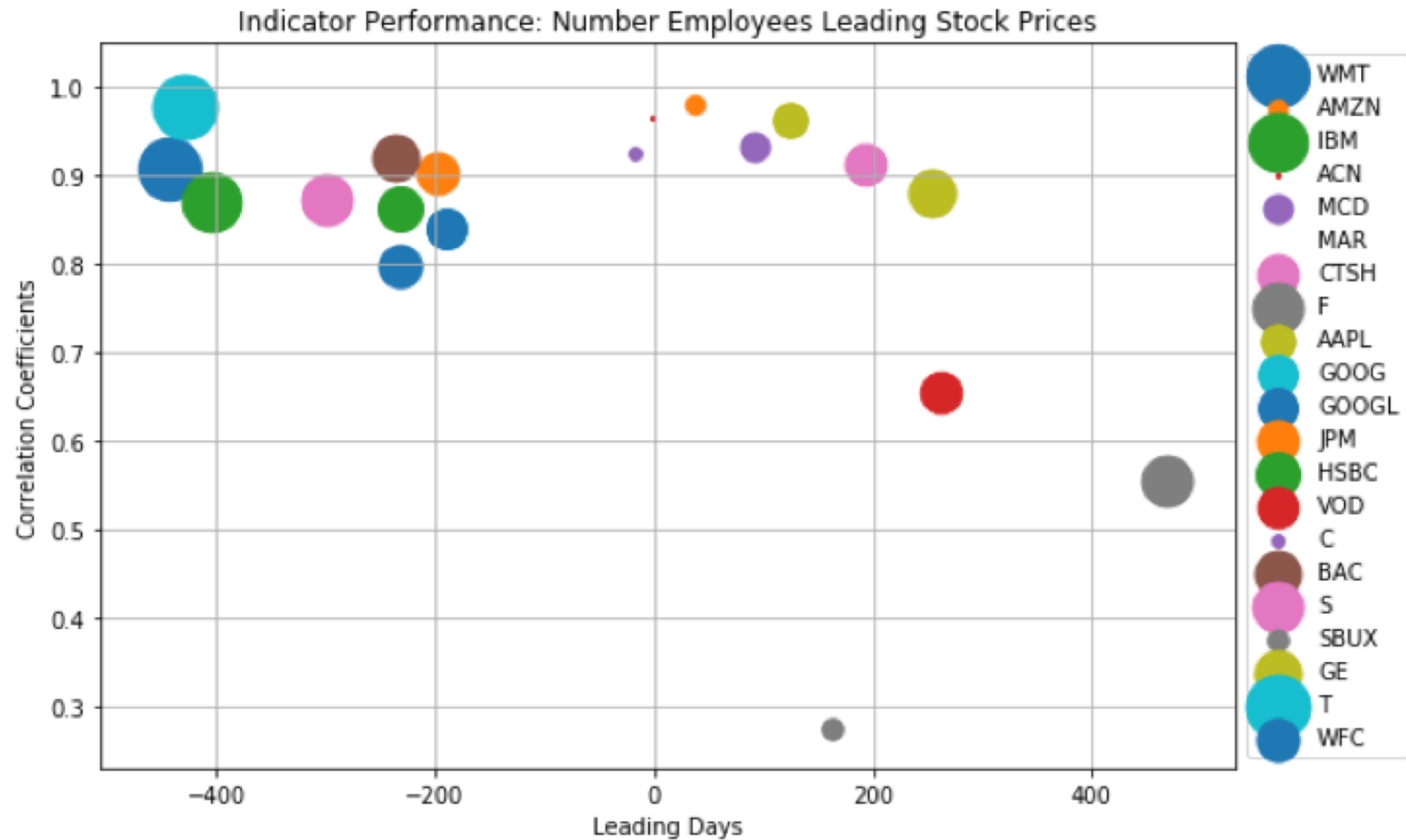Amazon
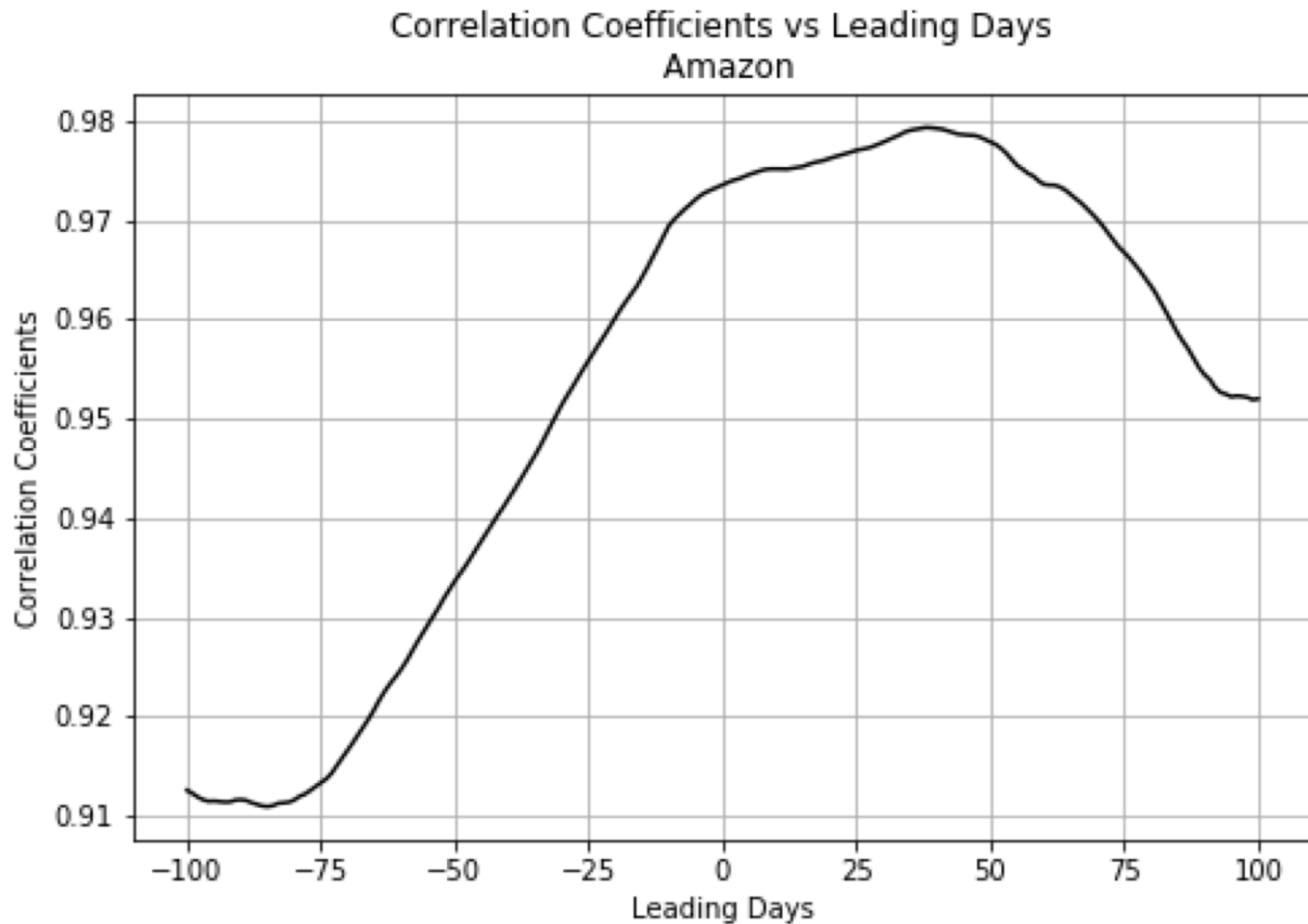
- Leading: Stock Prediction
  - » From LinkedIn Profile

- Lagging: Job Seeking
  - » From Stock Price
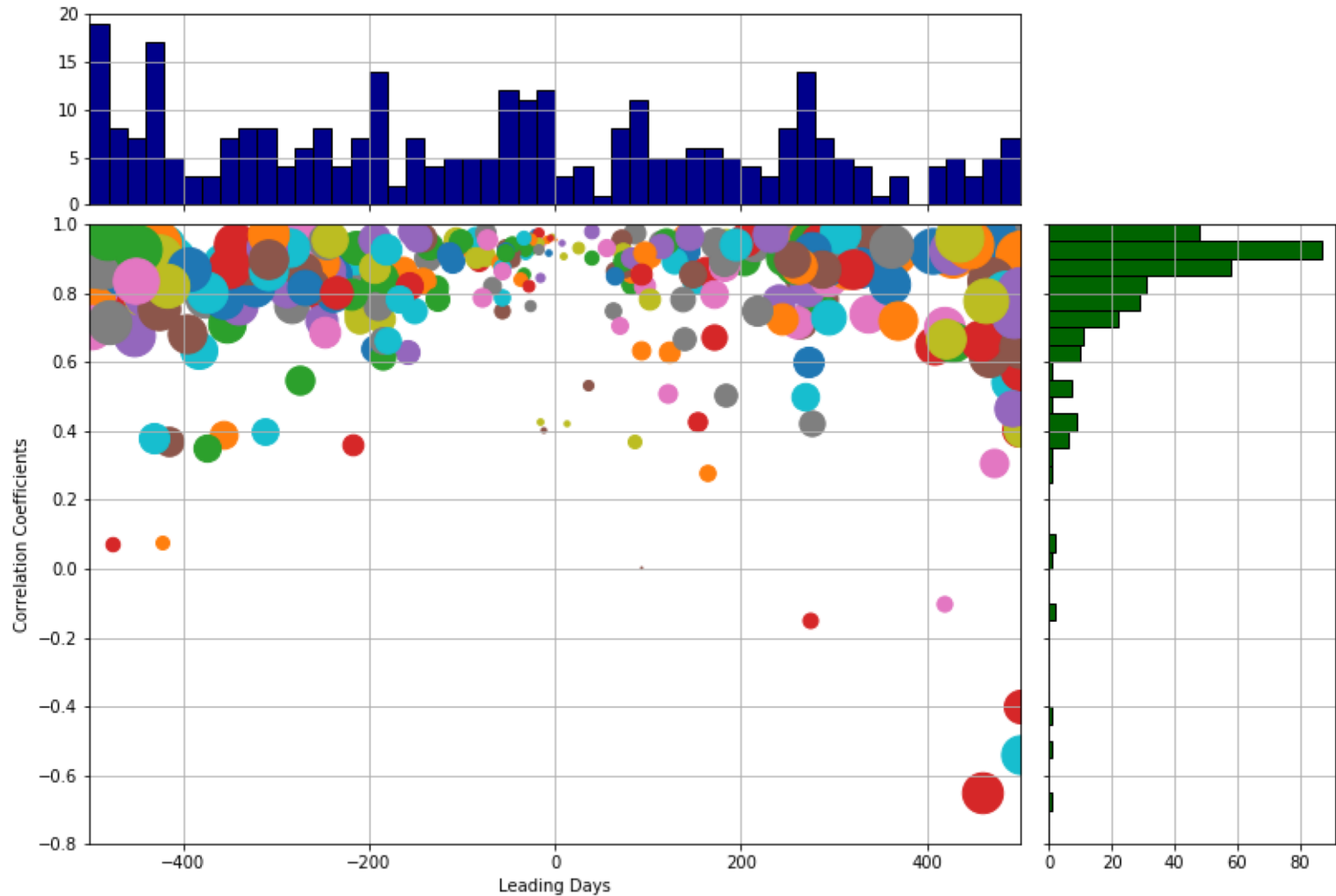
Indicator Performance of Top 20:
Number Employees Leading/Lagging Stock Prices

Indicator Performance: Number Employees Leading Stock Prices

Indicator Performance
Search 1900+ companies, whose data cover 2016-2017

Further Steps and Approach:

- Split the date range into in sample period (for training) and out of sample period (for testing)

- **Predictive modeling** (Naive Bayes, Random Forest, LSTM, etc.). Based on the expansion of a company is leading or lagging the stock price, we build 2 different type of models:

  - **Leading model**: select companies with high correlation coefficients leading indicators, and build a machine learning model to predict the stock price from the number of employees

  - **Lagging model**: select companies with high correlation coefficients lagging indicators, and build a machine learning model to predict the number of employees from the stock price

- Explore the followers_count, and add it as an additional indicator into the models

- **Backtesting** in sample and out of sample, fine tune the model

- The **performance metric**:

  - For regression model (e.g. predict the stock price values), we use Root Mean Square Error (RMSE)

  - For classification model (e.g. predict of the stock price goes up or down), we use Area Under the Receiver Operating Characteristics (AUROC)