



Dali Souayah, CFA

Investment Professional | Data Scientist

Greater Boston Area



CFA Institute





# Machine Trading

Capstone Project

Dali Souayah

5-29-2019



## **Data Science Process**

1. Problem Definition
2. Data Collection and Processing
3. Model Building
4. Model Selection
5. Model Evaluation
6. Future Work

## Real World Problem

Apr 30, 2019: AAPL 199.90



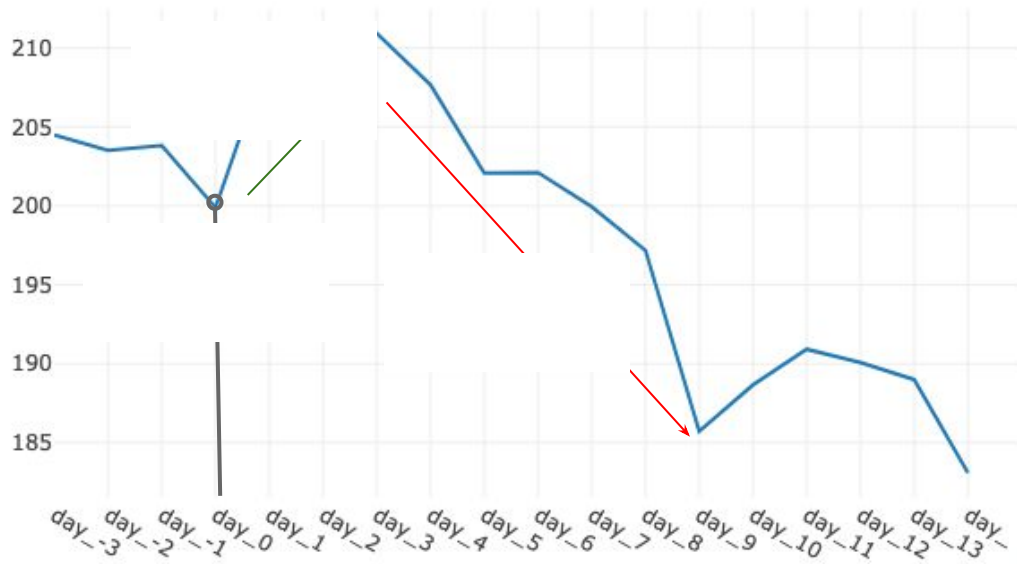
Press Release Financial Results Q2





## Real World Problem

Apr 30, 2019: AAPL 199.90



Press Release Financial Results Q2



## Data Science tools

- Natural Language Processing
- Analyze the text of the 8-k SEC filings
- Focus on qualifiers that echo, negativity, positivity or uncertainty
- Train a model to learn the trend between what is said and the movement of the stock price



## Data Collection

- SEC Edgar website
- Dictionary of words
- Yahoo Finance website

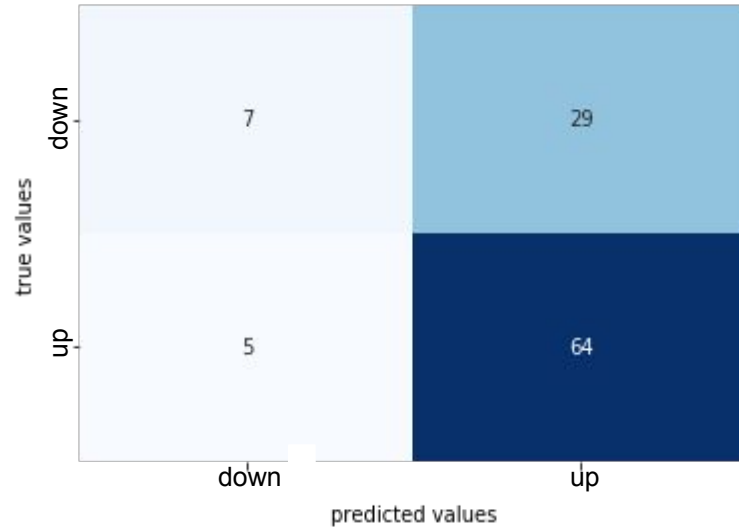
## Model selection

- Naive Bayes and Support Vector Machine consistently beat the baseline prediction
  - Baseline prediction: 58.9%
  - NB accuracy score: 68.5%
  - SVM accuracy score: 66.6%



## Model Evaluation

- Predicted almost all instances of the price going up
- Not predicting enough of prices going further down



## Simulation

The model out-performed passive investment by 2%





## **Future Improvements**

- Multi-classification and Regression models
- Introduce transaction fees
- Cover more stocks and different industries
- Use other text sources (e.g. social media)



# Thank you

For your attention

