Forecasting Stock Price via Social Media Analysis

Scott Coyne

Mentors Dr. Praveen Madiraju and Joseph Coelho

Objective

Volatile stock prices and over speculation have led to unstable investments and economic recessions. This project attempts to better understand and predict price movements from social media data.

Background

Most previous research is based on **sentiment analysis**, or labeling a post as positive or negative. Ex:

$AAPL incredibly bullish this week! (Positive)

$AAPL losing revenue. Sell now! (Negative)

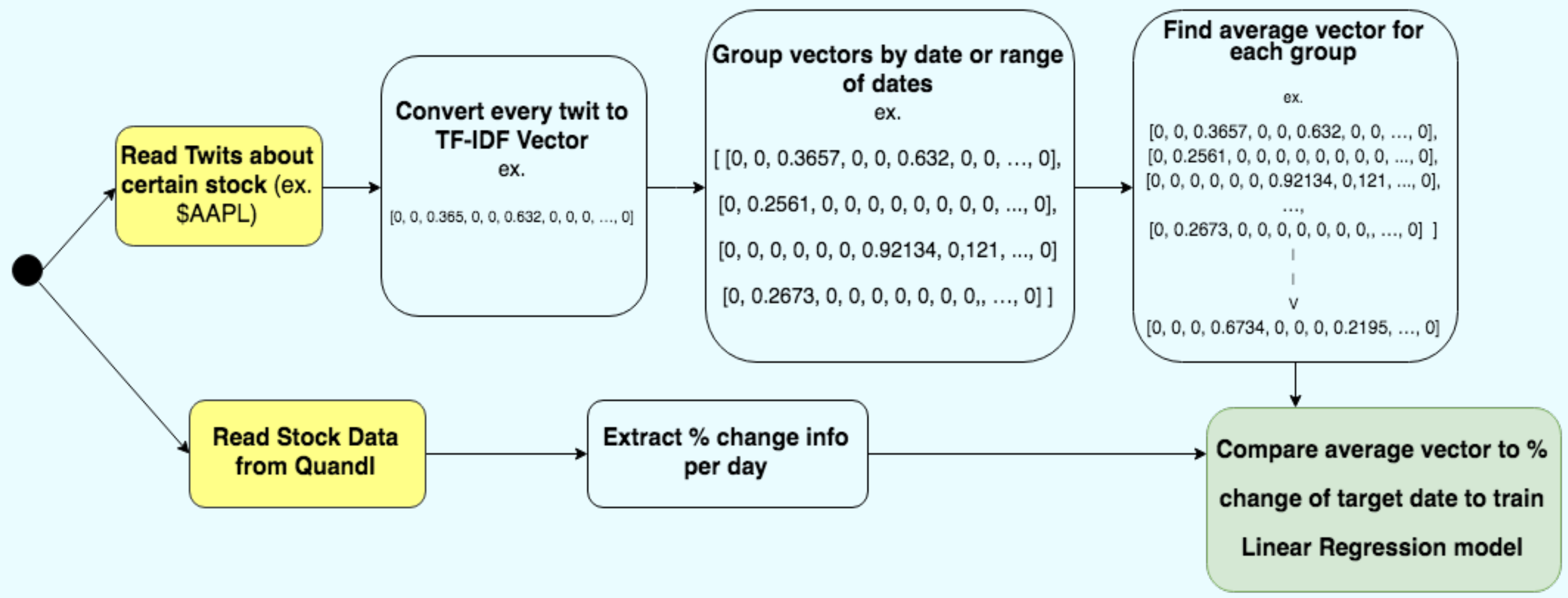
Sentiment is typically determined by a words **TF-IDF**, or term frequency inverse document frequency, to score words and relate them to sentiment.

Methods

Model 1 uses linear regression to relate TF-IDF's to price change.
Model 2 classifies posts by sentiment.
Model 3 uses model 2 to spot smart users and follow their predictions.

Model 1

Linear Regression

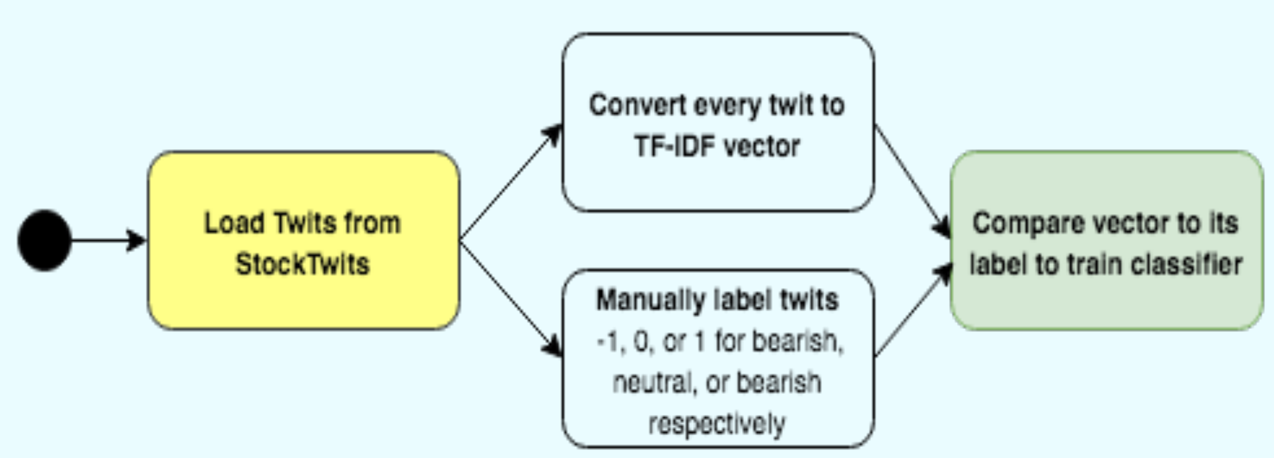


Above – Diagram visualizing linear regression model
Right – Individual stock results. Accuracies in green are better than base case.
Average accuracy came out to 52.5 percent.

Symbol	Accuracy	Base Case
JNJ	0.6333	0.6667
FISV	0.6000	0.6333
DISH	0.6000	0.5667
INTC	0.6000	0.5167
EBAY	0.5833	0.5667
NFLX	0.5333	0.5167
COST	0.5167	0.5500
SIRI	0.5167	0.4167
QCOM	0.5167	0.4667
WDC	0.5000	0.4333
ADI	0.5000	0.4167
MNST	0.4833	0.5000
CERN	0.4833	0.6000
VRTX	0.4833	0.5333
AAL	0.4667	0.5000
TSLA	0.4667	0.5167
UTX	0.4333	0.5333

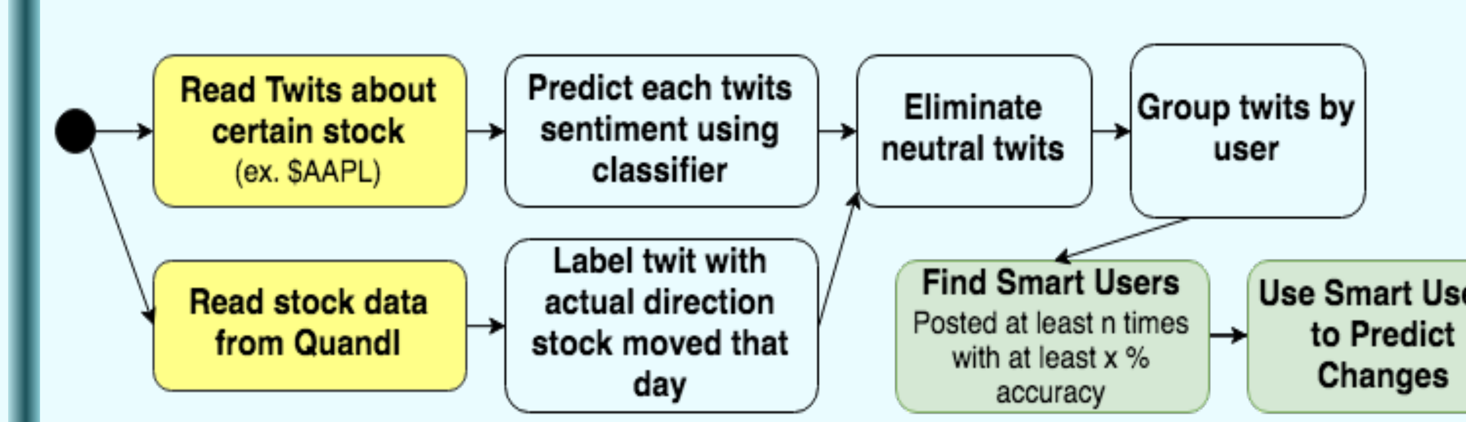
Model 2

Sentiment Classification



Model 3

Smart User Filtering



Symbol	Accuracy	Base Case
JNJ	0.786358	0.604348
JPM	0.703809	0.538598
AMZN	0.686308	0.550553
NFLX	0.618300	0.587832
BAC	0.613480	0.520525
INTC	0.612271	0.492095
AAPL	0.598293	0.652340
TSLA	0.591949	0.507714
GS	0.578090	0.573139

Conclusions

1. Aggregate TF-IDF **CANNOT** predict price change alone, despite previous sources suggesting otherwise.
2. TF-IDF **CAN** predict an individual stocks sentiment
3. Users who are correct in the past **CAN** predict the market with high success in the future

While Model 3 has its limitations, it yields an accuracy unprecedented by any prior work.

References and Acknowledgements

- Oh, C., & Sheng, O. (2011, December). Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement. In *ICIS*.
- Oliveira, N., Cortez, P., & Areal, N. (2013, September). On the predictability of stock market behavior using stocktwits sentiment and posting volume. In *Portuguese Conference on Artificial Intelligence* (pp. 355-365). Springer, Berlin, Heidelberg.

I would like to thank Marquette University as well as the Wehr Foundation for supporting my project. Without those organizations, this project would not have been possible.