**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:
- $\text{AAPL}$ incredibly bullish this week! (Positive)
- $\text{AAPL}$ 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.

**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**


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.