# Final Capstone - Prediciton of Stock Movement Based on Earnings Call Analysis

By Jacob Kovach -- Proprietary and Confidential

# Introduction

- Many methods have been utilized for prediction of stock movements. Most focus on more prolific sources of information such as previous trends, news articles, public sentiment, etc.
- For this project, we attempted to determine if movement could be determined strictly from the proverbial horse's mouth: that is, from the word of the company executives and outside analysts that meet quarterly to discuss the performance of a public company
- The completed product is a model capable of taking a raw transcript of a company's earnings call and providing a prediction of one of three movement classes over the week following the call

# Method

- Transcripts obtained from The Motley Fool (www.fool.com) using web scraping module to determine link titles from sitemap data. More than 13,000 transcripts were saved locally for recall.
- Python scripting conducted to retrieve datetime and raw file content then clean test and engineer additional features, specifically ones needed to create the target variable.
- Primary feature extraction conducted using TF-IDF vectorization. Considered single words, bi-grams, tri-grams, and quadra-grams (4 word sequences). Reduced feature space to 400 features through SVD.
- Modeling conducted using random forest, gradient boosting, logistic regression, and linear SVC.

# Results

- All classifiers predicts above 60% accuracy on the test set (75/25 training test split)
- Random forest and logistic regression performed best, with accuracy of > 65% for test set data
- All models struggled to accurately differentiate between class 0 and class 1 transcripts
- Most inaccuracies were a result of low recall scores between class 0 and class 1

# SWOT Analysis

# SWOT - Strengths and Weaknesses

Strengths:

- Comparatively to other prediction methods, model predicts movement with a high degree of accuracy
- Relatively simple implementation
- Computationally efficient due to smaller corpus size

Weaknesses:

- Model underpredicts the zero class (price decline)
- Limited application since we only evaluate text of earnings calls
- Less accurate as more data is introduced

# SWOT (con't) - Opportunities and Threats

Opportunities:

- Begin trading with algorithm strategy, using new earnings calls and gauging results on new data
- Divide the outcome variable into two binaries:
  - Increase vs. decrease
  - Significant vs. Insignificant
- Expand corpus to include other text sources

Threats:

- Potentially, other external factors play a bigger role in deciding stock movement.
- Wide time gap between earnings calls prevents long term trading strategy
-