Predicting SPY Move from Trump Twitter Analysis

By Alex Bell

Background

- NOTE! I am not a financial advisor. Invest at your own risk!
- Idea began at previous job
- Was able to learn skills here at GA to begin tackling the problem

Trump Tweets

Donald J. Trump @realDonaldTrump — Following

The Fake News Media is going CRAZY! They are totally unhinged and in many ways, after witnessing first hand the damage they do to so many innocent and decent people, I enjoy watching. In 7 years, when I am no longer in office, their ratings will dry up and they will be gone!

9:34 AM - 31 Jul 2018

23,083 Retweets 93,777 Likes

Donald J. Trump @realDonaldTrump — Following

Tariffs are the greatest! Either a country which has treated the United States unfairly on Trade negotiates a fair deal, or it gets hit with Tariffs. It's as simple as that - and everybody's talking! Remember, we are the "piggy bank" that's being robbed. All will be Great!

7:29 AM - 24 Jul 2018

21,250 Retweets 91,997 Likes

○ 30K   ⟲ 21K   ♡ 93K

Donald J. Trump @realDonaldTrump — Following

Some people HATE the fact that I got along well with President Putin of Russia. They would rather go to war than see this. It's called Trump Derangement Syndrome!

7:27 AM - 18 Jul 2018

33,852 Retweets 130,867 Likes

○ 55K   ⟲ 34K   ♡ 131K

Donald J. Trump @realDonaldTrump — Following

The people of Germany are turning against their leadership as migration is rocking the already tenuous Berlin coalition. Crime in Germany is way up. Big mistake made all over Europe in allowing millions of people in who have so strongly and violently changed their culture!

9:02 AM - 18 Jun 2018

34,488 Retweets 117,226 Likes

○ 37K   ⟲ 34K   ♡ 117K

Whats the Goal?

- Take a tweet from Donald Trump
- Analyze the text
- Find any correlation with the stock market
- Predict the stock market close as positive or negative (from previous day)

SPY

- Exchange Traded Fund (ETF)
- Follows S&P500 Index
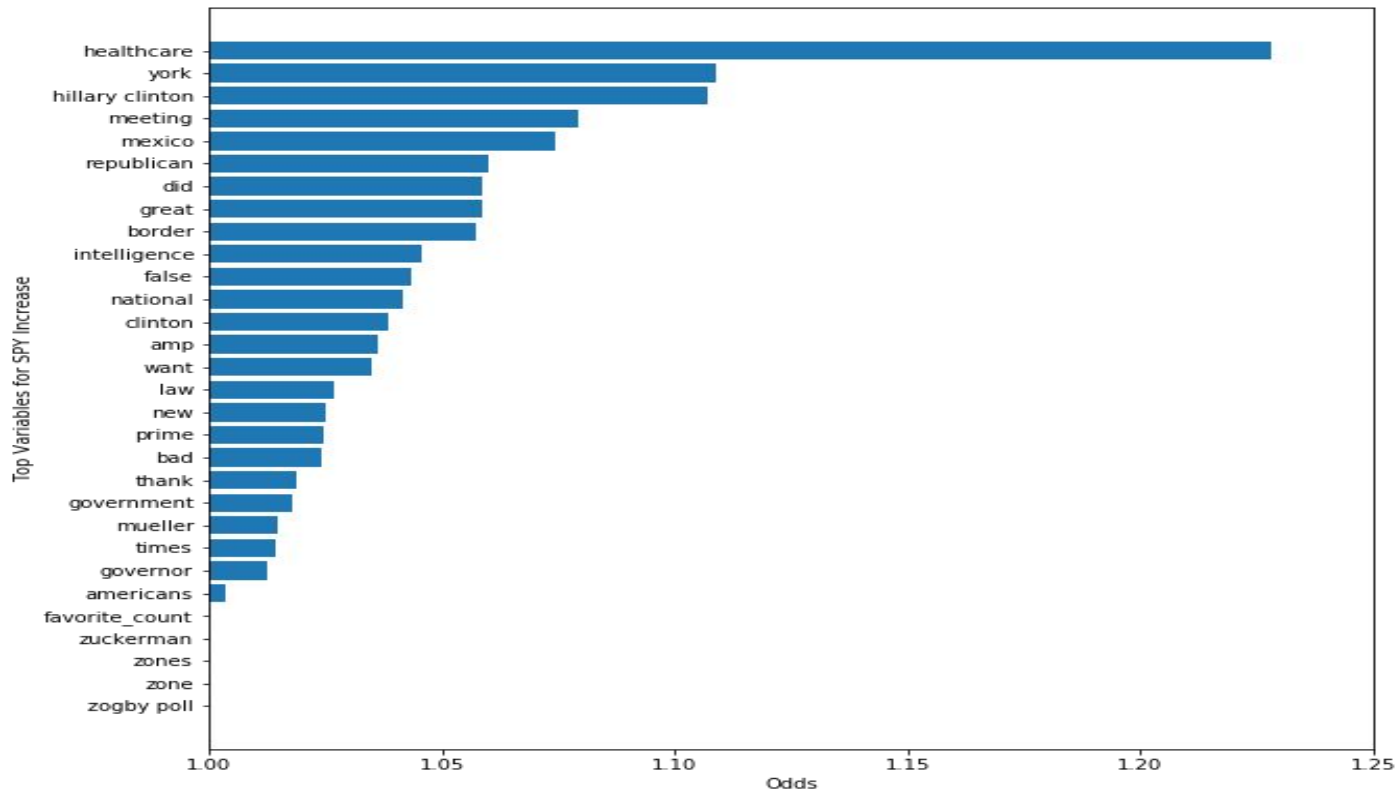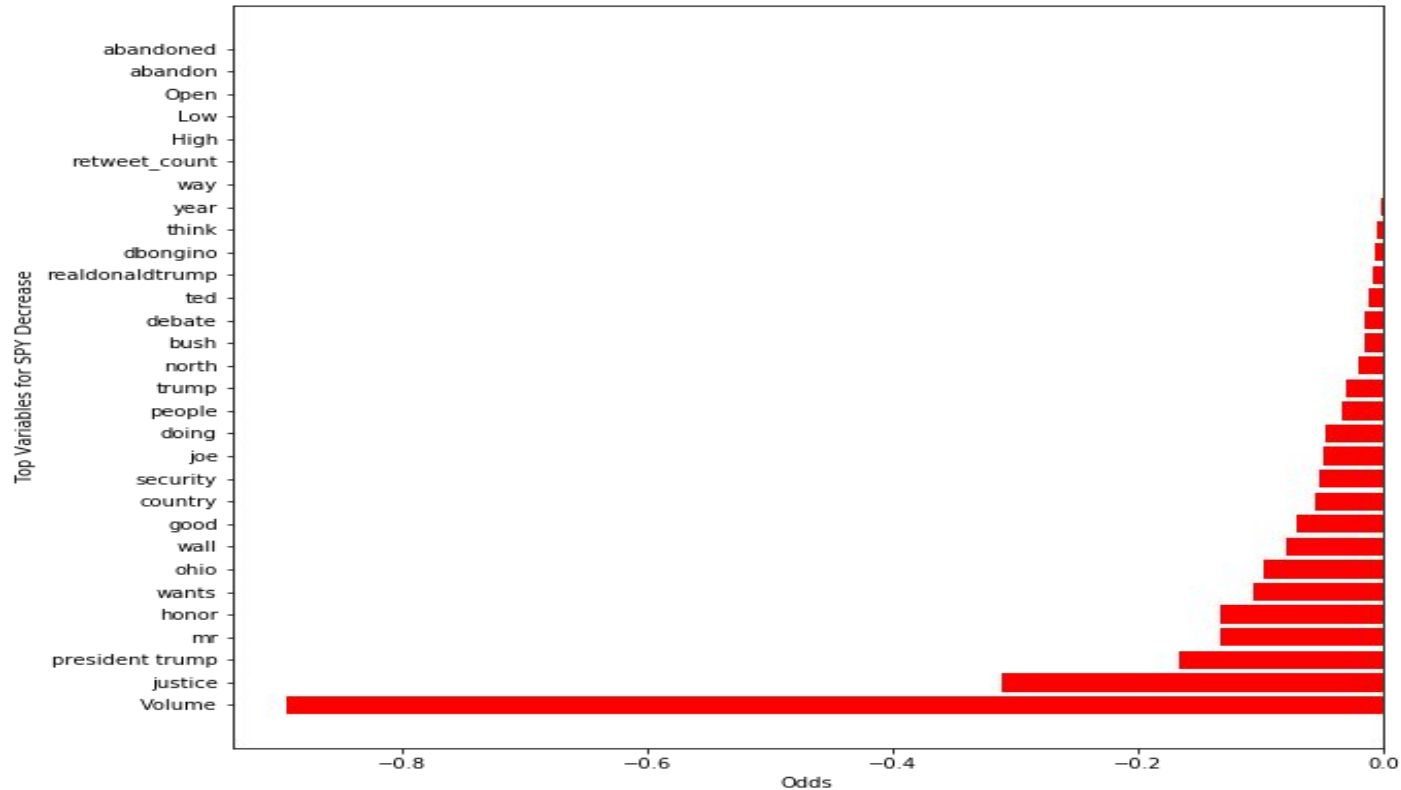- Great corollary for overall stock market

Data Collected

- Donald Trump Twitter Data
    - Source (ie. Twitter for iPhone)
    - **TEXT (Over 17,000 posts)**
    - Time created
    - Retweet and favorite count
- SPY Historical Market Data
    - Open
    - High
    - Low
    - Close
    - Volume

Positively Correlated Variables

Negatively Correlated Variables

Tweets with Compound Score

Text	Compound Score
"Congratulations to Patrick Reed on his great and courageous MASTERS win! When Patrick had his amazing win at Doral 5 years ago people saw his great talent and a bright future ahead. Now he is the Masters Champion!"	0.9899
"Here we go with Mitt Romney but so fast! Question will be is he a Flake? I hope not. Would much prefer that Mitt focus on Border Security and so many other things where he can be helpful. I won big and he didn't. He should be happy for all Republicans. Be a TEAM player & WIN!"	0.9843

Tweets with Compound Score cont'd

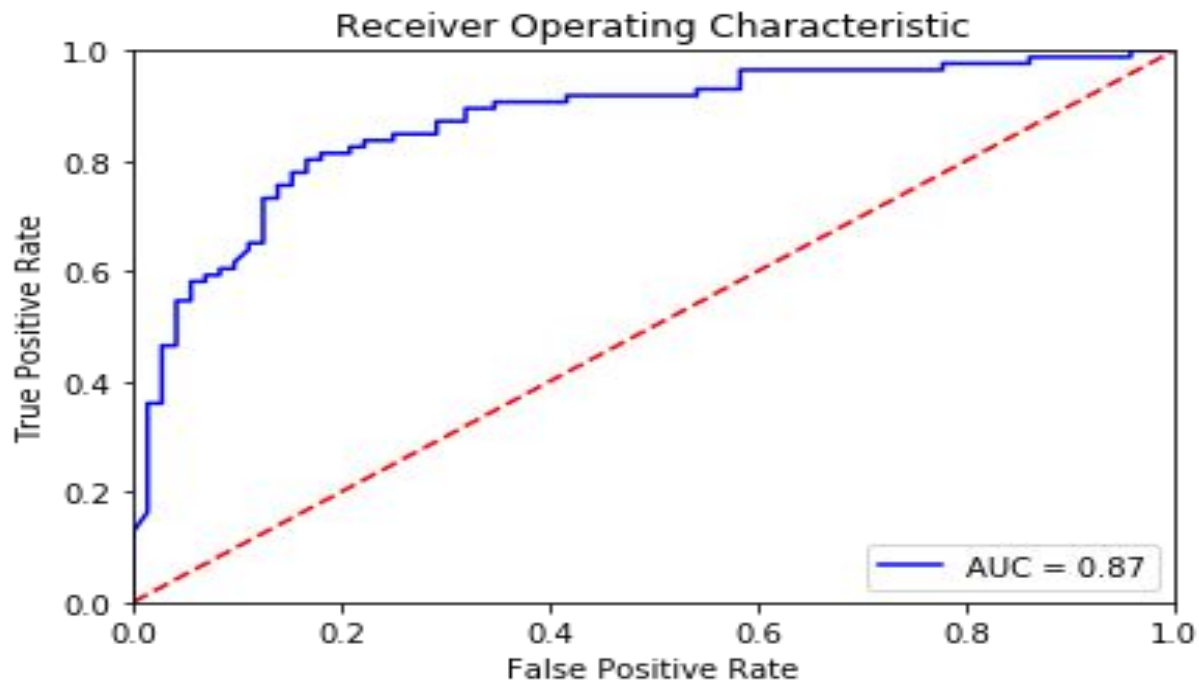| Text	Compound Score
"It is outrageous that Poisonous Synthetic Heroin Fentanyl comes pouring into the U.S. Postal System from China. We can and must END THIS NOW! The Senate should pass the STOP ACT – and firmly STOP this poison from killing our children and destroying our country. No more delay!"	-0.9864
"The Rigged Russian Witch Hunt goes on and on as the "originators and founders" of this scam continue to be fired and demoted for their corrupt and illegal activity. All credibility is gone from this terrible Hoax and much more will be lost as it proceeds. No Collusion!"	-0.9785

Modeling

- Made Target Variable of up day being any time closing price of SPY was greater than the day before.
- Metric used across all = Accuracy
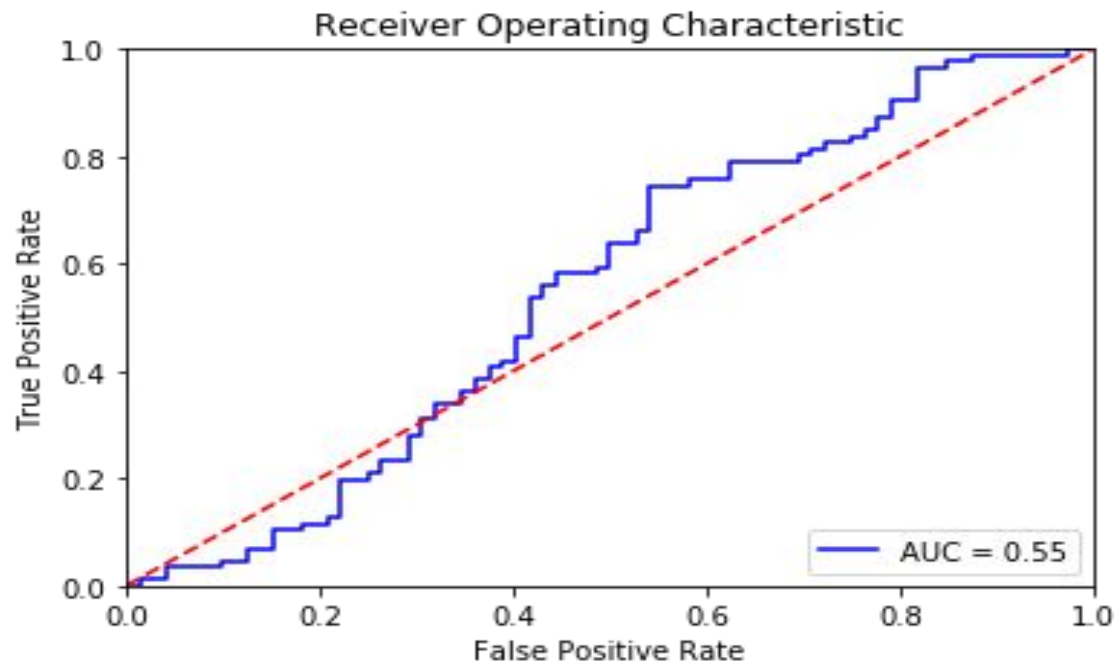- Baseline Score = 54% Accuracy

Modeling Cont'd

- Logistic Regression
  - Default = 50% Accuracy
  - Ridge = 50% Accuracy
  - Lasso = 58% Accuray
- Random Forest
  - Default = 66% Accuracy
- Gradient Boost Classifier
  - With Parameters = 81% Accuracy

More Scoring - Gradient Boost ROC AUC Graph

More Scoring Cont'd – Logistic Regression (Lasso)

Going Forward

- Best Model = Gradient Boost Classifier
- Nice to get a high accuracy score but…
- Implementation in real world wouldn't be the same
- Problems with data set up
  - Go by single Tweet instead of grouping together
  - Would give more importance to each tweet
  - Since grouped all tweets together by day, may have lost model efficiency
- Good stepping stone for further tweaks and future use!

Thank you!

Questions?