Senior Project – Electrical Engineering – 2020

Twitter as a Predictor of Stock Market Volatility

Maxwell Chase Pisano Advisor – Prof. Webb

INTRODUCTION:

- Sentiment is the feelings behind words.

$AAPL: Stock sunk 2.8% biggest slump since oct how most valuable tech company fall 2.8% #Losing Streak #SELL%#BEAR

Positive Words	Negative Words	Neutral Words	Total Words	Positive #	Negative #	Total #
3	7	6	16	0	2	3

Points Negative	Points Positive	Overall Raw Score
-9.5	3	-6.5

Take raw score and divide by total words so (RAW SCORE) / (TOTAL WORD COUNT) = (Sentiment weight)

So the tweet above = -6.5 / 16 = -0.40625 = -40.625%

- Volatility is the amount of uncertainty or risk in the size of changes in a security's value. Volatility is calculated using Implied Volatility which takes the present and future sentiment of the security or Index
- Volume is the amount of shares traded or the amount of tweets that happen in a certian time frame
- Since changes in price and volume are directly correlated to new information entering the market one would expect that when there is an increase in information into the market.
- Markets are neither efficent or inefficent which is demonstrated by traders being able to take advantage of market situations to yield returns.
- If the stock or index has high liquidity then buying or selling units which is volume will not effect the price and that is where volatility comes in

SYSTEM OVERVIEW:

- The system is compose of three basic parts
- Financial Data Feed: Market Data is collected from Bloomberg Terminal Subscription, Factset, and Thomson Reuturs EIKON
- Twitter feed: Tweets are scraped directly from twitters api and fed into the data frame which is hosted in Google colab which is a web based VM for Machine learning and Data Analytics
- News Data: Top News from 30 Top News providers is imported and used to filter tweets for relevance to current events
- Predictive Engine: This is where the processing of twitter data and market data takes place and using Python and R Data analytics takes place and the results are output to the graphing engine
- Graphing: Using Tableau which connects to the Data Frame graphs the predicted data

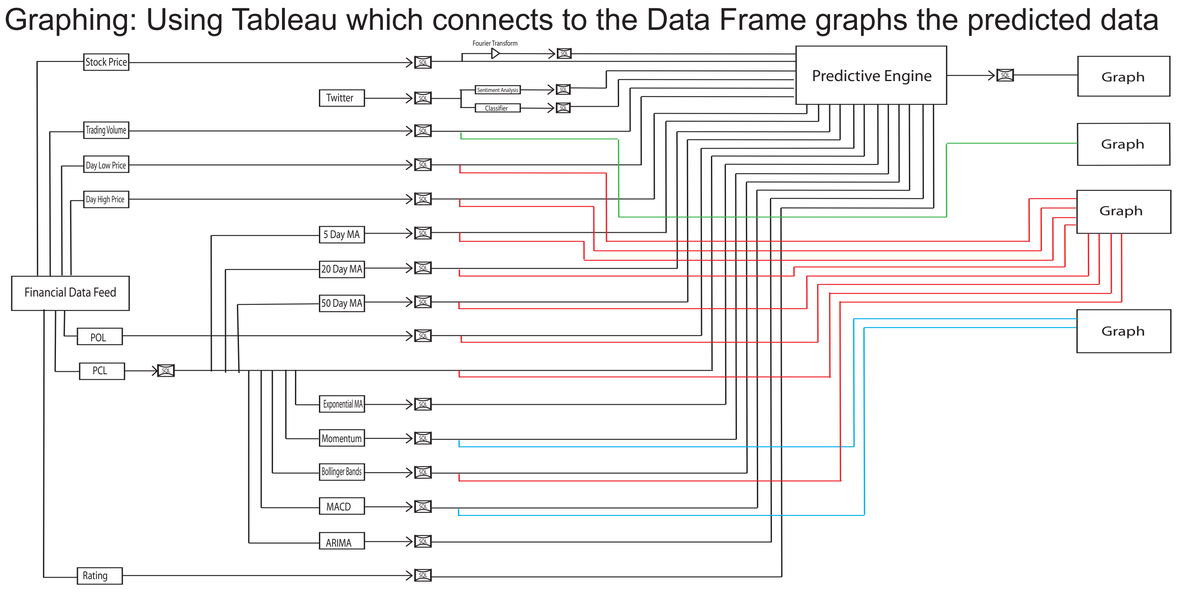

Figure 1: Block Diagram of System

TESTING:

- Six months of historical tweet data and market data are loaded into the Data Frame and using machine learning the system determines when there is a correlation between the tweet volume and sentiment and the volume and price changes of the index and the calculated implied volatility
- Using the 6 months of historical data that was hand sorted and pre tagged for correlation to train the machine learning algorithim with then compare the implied volatiity it calculated for the Index to the Historical Volatility for that time period.
- If there is a signicant correlation between the historical data and training data the machine learning algorithm will then start predicting volatility in real time
- Tweets are constantly being currated and the list of current events is constantly being updated to ensure only tweets with any significance are included

PRELIMINARY RESULTS:

- Inital Results concluded that there is a corellation between stock price and twitter sen timent and volume of tweets and volume of shares traded
- There is an increase in volotility in the trading price as the volume of shares increase
- This spike happens roughly 2 minutes after the tweets are posted and the market stabilizes amoust 10 minutes after the tweet was posted.
- When there is an increase in Twitter volume and the sentiment is positive there is an increase in price and volume.
- When there is an increase in Twitter volume and the sentiment is negative we will see a decrease in price and increase in volume. If twitter volume decreases we will also see a decrease in volume until its below the daily moving average

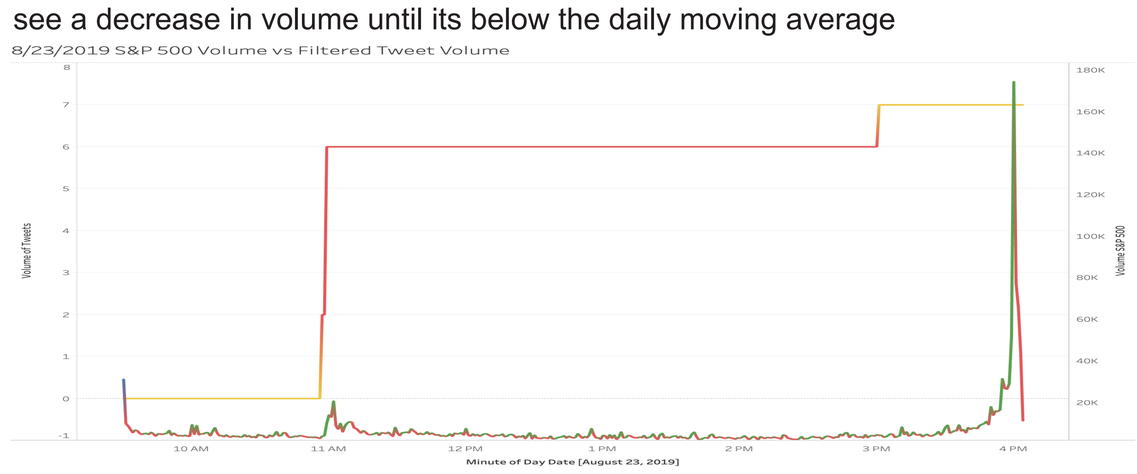

Figure 2: Graph of S&P 500 Volume vs Twitter Volume

Figure 3: Graph of S&P 500 Price vs Twitter Sentiment

FUTURE WORK:

- The next step is to run the data through the machine learning algorithm to deterimine the validity of the training data and to train the model.