Doug Dellolio
NSF Summary Page

There are many programs in existence today that do different types of analysis on stocks. It is not hard to find one that will look at factors such as day volume, the amount of shares traded in a day, and predict based on past days what today will bring. One of the main drivers of a stock price is media and in particular social media. Media has a great impact on whether the price of a stock moves up or down. Companies listed on any exchange will publish press releases regularly to inform the public about events or products. After a press release is published, the price of the stock can move in any direction as a reaction by shareholders and traders. For example: Apple publishes a press release stating that it has created the first driver-less car. Almost immediately, the stock will increase in price because it would be considered a huge breakthrough in technology. On the other hand, Twitter publishes a press release stating that they are being investigated for securities fraud. At the open, shareholders will be selling out of the stock because they do not want to be invested in a company that is conducting fraud and can potentially be shut down.

One goal of this project is to analyze the words and phrases of these press releases into positive, negative and neutral categories. Categorizing these words and creating a significant word and phrase group will allow for the prediction of stock price movement for future press releases. This program will be implemented using a library called LingPipe, which is a tool kit for processing text using computational linguistics. LingPipe is currently embedded in a commercial product called ThreatTracker which is used by the U.S. Department of Defense because its algorithm is so powerful and trusted. In fact, over 200 government intelligence analysts were trained in the use of this program since it was adopted. Capabilities of this library include extracting significant multi-word phrases in one corpus and of relatively significant ("hot") terms in one corpus relative to another. What are the benefits to using this program as opposed to just reading on your own? On average a company will publish 10-20 press releases a year. For a human to read and analyze 10 press releases at the minimum takes up much time and energy. However, when using a program like this, the average run time is 1.3 seconds. This will include the analyzing of a years worth of volume data correlating to press release history, and a sentiment analysis on the words and phrases. Therefore, when a future press release is published, categorized terms and frequencies from past press releases will be available immediately and will serve as a guide to the trader to make proper investment decisions.

This will be the first finance application to allow a trader to view how the media affects the market's movement. As of now, there are no programs that will analyze and predict movement of stock price based on media alone. A goal of this project is to expand on the idea that media has a direct impact on the movement of stock price. Stock traders share a common objective, which is to invest money and grow the value. This program will also serve as an educational tool to companies because it illustrates how word choice may affect people's decisions. This can shape how a company might word a press release in the future. Lastly, it will educate traders to look for certain patterns of phrases in future press releases to make educated decisions about where to invest their money.