**Executive Summary**

The impact of news media information on stock prices is one common topic of interest in sentiment analysis. For this project, we aim to show a high positive correlation between positive news sentiment and positive stock price changes in the S&P500. Sentiment analysis will be performed through the use of 3 different libraries, through which the results will be published and noted for future research.

**Research Topic & Hypothesis**

Stock market prediction features a long standing debate over whether a stock’s historical prices can determine its future prices. A theory called the Random-walk analysis has sought to claim that a stock’s future price is independent of its history. This means that stock prices should instead be determined by the information of tomorrow. This has led to research trying to uncover what kinds of information are meaningful in determining future prices. For example, Malkiel and Fama posit that market information is embedded in market price if it is an efficient market.[[1]](#footnote-1)

For this project, we are looking to perform sentiment analysis on financial news sources to test whether negative or positive sentiments about future stock market prices fuel market growth, in this case the effects on the S&P500, stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States. By quantifying language, we seek to examine if it provides novel information about the firm's earnings and returns. We will compare and contrast the different sentiment word lists to determine dictionary models the sentiments better, and which brings more prediction performance improvements than the other dictionaries.

**Possible extensions to the research topic**

Combining technical indicators from stock prices and news sentiments from textual news is an interesting extension to the research problem, as shown by Xiaodong Li et al. [[2]](#footnote-2)

**Datasets used**

***Sentiment Word Lists that can be considered:***

1. Loughran and McDonald Sentiment Word Lists for longer form articles:

<https://sraf.nd.edu/textual-analysis/resources/#LM%20Sentiment%20Word%20Lists>

2. VADER (Valence Aware Dictionary for sentiment Reasoning) for short form postings - microblogs and tweets:

<https://github.com/cjhutto/vaderSentiment>

3. Harvard iv-4 dictionary:

<http://www.wjh.harvard.edu/~inquirer/homecat.htm>

**Relevant python packages:**

<https://pypi.org/project/pysentiment2/>

Stocks news data that features over some 4 Million articles scraped covering some 6000+ stocks: <https://www.kaggle.com/miguelaenlle/massive-stock-news-analysis-db-for-nlpbacktests>,

<https://www.kaggle.com/aaron7sun/stocknews>

**Literature/Articles of reference**

Predicting the Effects of News Sentiments on the Stock Market

<https://ieeexplore.ieee.org/document/8621884>

News impact on stock price return via sentiment analysis

<https://doi.org/10.1016/j.knosys.2014.04.022>

Incorporating stock prices and news sentiments for stock market prediction

<https://doi.org/10.1016/j.ipm.2020.102212>

**Specific timeline**

|  |  |
| --- | --- |
| Date | To be completed |
| 20 January 2021 | Submission of Project Brief |
| 30 January 2021 | DSS Project Onboarding |
| 8-14 February 2021 | First meeting |
| 22-28 February 2021 | Mid-Semester project discussion |
| 15-21 March 2021 | Final meeting and project submission |

1. Malkiel, Burton G., and Eugene F. Fama. "Efficient capital markets: A review of theory and empirical work." *The journal of Finance* 25, no. 2 (1970): 383-417. [↑](#footnote-ref-1)
2. Li, Xiaodong, Pangjing Wu, and Wenpeng Wang. "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong." *Information Processing & Management* (2020): 102212. [↑](#footnote-ref-2)