

---

## STOCK MARKET PREDICTIONS USING NLP SENTIMENT ANALYSIS



# Project Proposal

Prepared for: Metis Staff Instructors(Chris Bruehl, Leon Johnson)

Prepared by: Rahul D. Raju, Metis Student

September 9, 2021

---

---

# EXECUTIVE SUMMARY

## Opportunity

The stock market is a reflection of the complex interactions amongst humanity, nature and technology. It is a forward looking mechanism that distills and discounts information in order to make probability based investment and trading decisions. The list of dependent variables that are inputted into these decisions are nearly infinite. Historically, the way financial investment and trading companies have dealt with this was to create numerous sub-specialties and assign analyst, traders and bankers to each of these. Due to the digital revolution, however, the size, accessibility and delivery speed of information has grown to the point that even specialists cannot practically assess data in a timely manner. In response to this, trading firms for example, have been designing and implementing trading algorithms that quickly incorporate a vast amount of information in order to make trading decisions far faster than any human. The quantitative inputs into these algorithms are relatively easy to process, analyze and extrapolate from. The challenge, however, has been how to effectively translate and incorporate qualitative(textual) information into actionable trading schemes. The textual information can be in the form of financial statements, earnings releases and general company specific or macroeconomic news articles. Any tool that could quickly “decode” this information would provide enormous advantages to competitive market participants who place a significant premium on even a nano-second of trading time.

## Impact/Solution

Natural Language Processing is a machine learning tool that allows for the distillation and analysis of vast amounts of textual information. Specifically “sentiment analysis” can make use of NLP in order to determine if textual data is positive, negative or neutral in relation to a pre-

---

---

specified target. This is accomplished by machine learning algorithms that translate this information into objective, quantitative data that can be incorporated into trading algorithms. In this case the data that will be analyzed are news headlines which can have an immediate impact on stock market trading decisions. The target will be daily movements in the Dow Jones Industrial Average. The effect of news can also be measured on prices the same day or one to three days after its release. Movements will be labeled as downward, upward or unchanged.

## Project Data

The data set used in this analysis was found on Kaggle.com. The data ranges in time from 2008 to 2016. It includes the daily top 25 news headlines from the Reddit WorldNews Channel as ranked in importance by Reddit users. In terms of the target data, a “1” will represent the daily closing price of the Dow Jones Index if it has risen or remained unchanged since the previous day’s close. A “0” will represent a decrease in value from the previous day’s close.

Source: Sun, J. (2016, August). *Daily News for Stock Market Prediction, Version 1*. Retrieved [September 8th, 2021] from <https://www.kaggle.com/aaron7sun/stocknews>.

## Project Tools

Exploratory Data Analysis: Pandas, Numpy

Analysis: SKlearn, NLTK, TextBlob, Regex, VaderSentiment

Visualizations: Matplotlib, Seaborn, Tableau

---

---

## Minimum Viable Product

A Minimum Viable Product would include;

- Topic Modeling on the corpus of headlines
- Accuracy of predictions on the training set