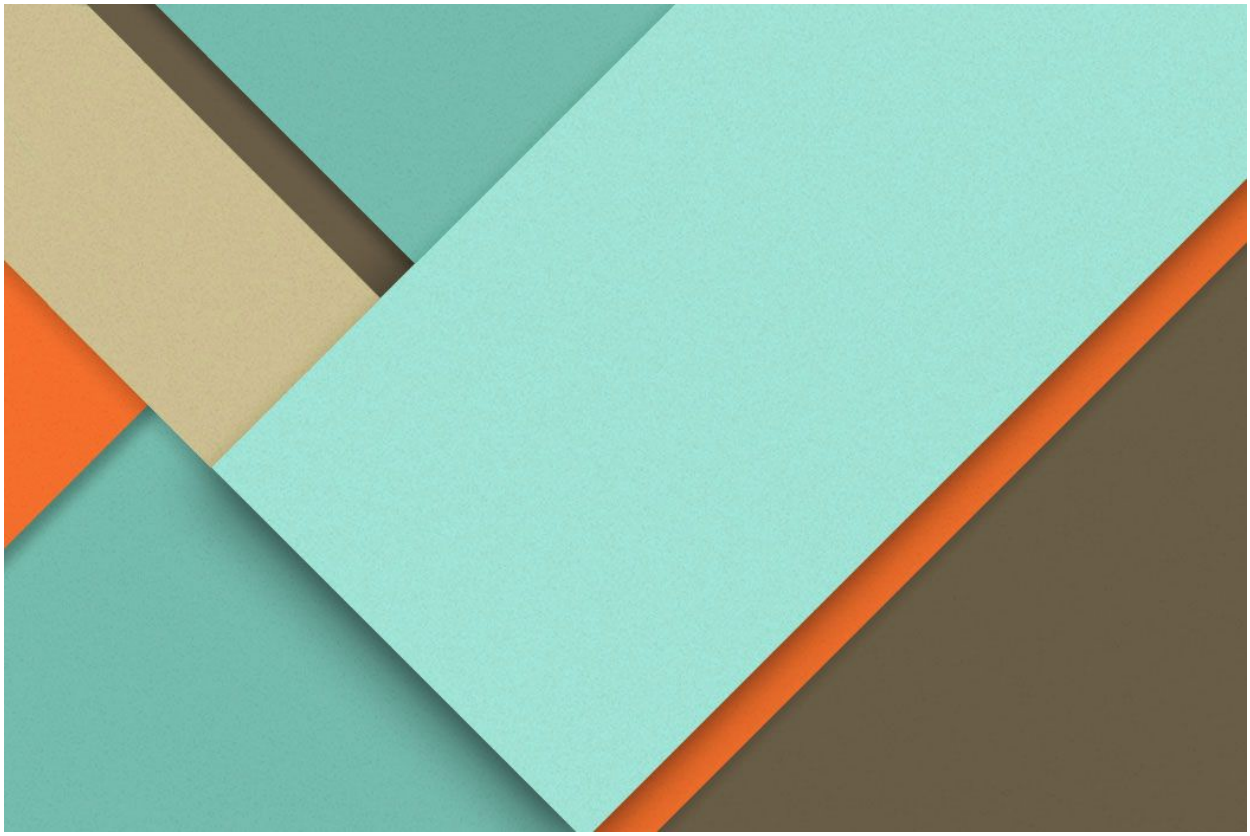
Final Project: Proposal



Stock Market Prediction

February 18th, 2019

Team Name: Stock Brokers

Tommy Betz, Gaurav Prachchhak, Veekesh Dhununjoy, Mihir Gajjar

Simon Fraser University

1. List 3 questions that you intend to answer (1 point)

Question 1: How do different factors like news and social media affect the stock prices?
Question 2: What would be the prices/trend of the stocks of different companies in future?
Question 3: How can a user decide where to invest?
Question 4: Which news topics affect the prices/trend of the stocks for different companies?

2. List all the datasets you intend to use (1 point)

1. Daily Historical Stock Prices Dataset: https://www.kaggle.com/ehallmar/daily-historical-stock-prices-1970-2018/version/1
2. Real-Time Stock Market Information API: https://www.alphavantage.co/
3. Social Media APIs.
4. Web Scraping from Yahoo finance for missing years (if needed).

3. Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)

Data Collection: Where/how to get data

We are using the stock data from Kaggle from 1970-2018. In addition, we will be using news, social media and stock market APIs to collect the data. We will also perform web scraping if needed for the missing data from Yahoo finance.

Data Exploration: Do you need to conduct EDA in order to understand the data?

Yes, we are planning to carry EDA on the fields in the datasets to better understand the data and in order to find out the missing values.

Data Cleaning: Do you need to clean data? How to clean them?

We will have to clean the text data from the APIs for news and social media data using various data cleaning techniques such as string tokenizing, removing stop words, removing duplicates, filtering outliers, etc.

Data Integration: Do you need to integrate data from multiple sources?

Yes, we are integrating data from datasets, APIs and web scraping (if needed).

Data Analysis: What analysis do you intend to do? (e.g., SQL, Statistics, Deep Learning) How to evaluate your analysis results? (e.g., evaluation metrics, confidence intervals, benchmark)

We intend to use NLP techniques to understand the text data which we get from news and

social media APIs. With the help of the above analysis and historical data for stocks, we are planning to predict the stock prices/trend in the future using various Machine Learning models. We explored certain confidence metrics and finalized https://www.kaggle.com/c/two-sigma-financial-news#evaluation which was used in a Kaggle competition.

Data Product What product do you want to build? (e.g., visualizations, an interactive web app, a jupyter notebook)

We are planning to make an interactive web application which will take as an input how much the user wants to invest, type of investment and gives as an output a report which contains the expected gains from the investment, projected growth, the stocks in which one should invest, the most relevant news articles or social media data which supports the suggested stock, etc.

4. Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)

It can help the user decide in which stock to invest based on how much and for how long one is willing to invest with the help of current news, social media data and historical financial data.