RO3

FYP Proposal

<u>Bull & Bear</u>

A System for Comparative Analysis of Different Stock Prediction Methodologies for Beginner Investors

by

Ishan Jain, Mohibul Hassan, Nimisha Goyal, Tanmay Goel

RO3

Advised by

Prof. David Rossiter

Submitted in partial fulfillment

of the requirements for COMP 4981

in the

Department of Computer Science

The Hong Kong University of Science and Technology

2018-2019

Date of submission: April 17, 2019

Table of Contents

Abstract

Investing in stocks, if done correctly, is a very effective way to earn passive income for an average user. However, for an average individual from a non-finance background, it can be confusing, overwhelming and even dangerous if given the wrong advice.

Bull & Bear has been designed for those individuals. We want to make stock investing an appealing and educational activity for beginners by giving them confidence to make investment decisions based on strong analysis and comparison of the major stock prediction techniques within Technical Analysis indicators such as MACD, RSI, EMA and Bollinger Bands as well as upcoming new technologies like Social Media Sentiment Analysis.

Our algorithms are validated by back testing against actual data, and we have been able to compare the effectiveness of some major stock prediction methods for the 30 companies listed on the Dow Jones Index. An interesting find from our analysis was that the most common parameters for the stock prediction techniques may not always be the best ones. For e.g. common EMA parameters are 20, 50, 100 but we got the best results with 128.

85% of the testers of Bull & Bear were more confident about investing in the stock market after using our platform.

1 Introduction

1.1 Overview

Investing in the stock market can be a great method to increase or decrease one's personal wealth in the long term. But after the 2008-2009 financial crisis, many individuals became skeptical about investing their hard-earned money in the stock market. Even though the S&P 500 is up 300 percent since 2009, real estate is still voted as the best long-term investment [1].
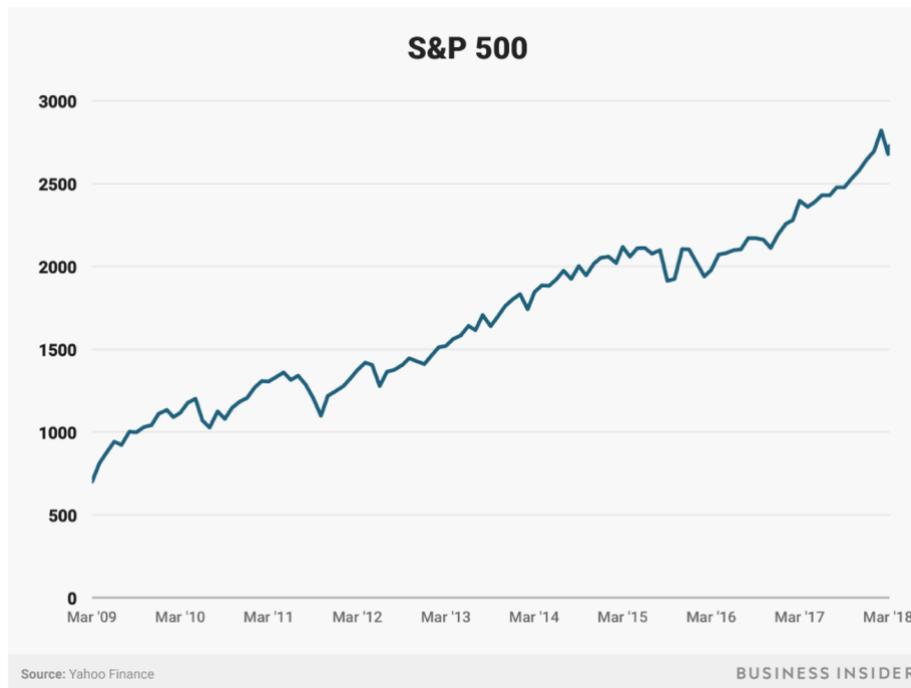
S&P growth over the past 9 years

Moreover, even if people do wish to trade, the amount of research and financial knowledge required sets a potentially steep learning curve, and it can overwhelm new entrants into the stock market. Numerous sources provide stock predictions and suggestions on which stocks to buy, but many of them are misleading. Also, news is often a good indicator of how a stock is doing [2] but analyzing the news can also be difficult for new-comers. The lack of education on this subject, especially amongst the younger generation is still very prevalent [3].

There already exist several algorithms, formulas and techniques to predict the price of stocks. There are also many websites which give predictions based on their own algorithms. An active and intelligent trader must be able to understand which techniques work best and also stay updated of any upcoming news concerning different

stocks. Having this knowledge can enable the trader to make smart investments and receive financial gains from the stock market.

In today's fast-paced environment, most people are always on the move and use smartphones and mobile applications for a lot of tasks. Stocks can rise or fall at any given moment, so it is important for traders to be aware of any new updates, as these could be signals to buy or sell stocks.

However, our research shows that many of the mobile applications related to the stock market are mainly focused on letting users follow their portfolios or buy or sell on the stock market [4]. There is currently no known platform which provides the user with a variety of different stock prediction techniques and gives them information regarding the effectiveness of each of these techniques, especially on demand via a smartphone application.

Our project focused on solving this problem by empowering beginning-level stock traders to make smart trading decisions by allowing them to compare the effectiveness of different stock prediction techniques for particular stocks that they're interested in. Moreover, they dynamically received relevant insights from recent news related to the stocks.

1.2 Objectives

We aimed to create an interactive and easy-to-use application that provides beginner level traders with relevant information that they need so as to make sound and effective trading decisions. Moreover, we intend to share this information with as many users as possible to promote the habit of healthy trading for personal wealth generation.

We did this by
1. Providing predictions made by different indicators within Technical Analysis
2. Performing Sentiment Analysis on financial news and tweets related to the stocks
3. Comparing these techniques with each other and informing the user about the effectiveness of each prediction technique over a period of time
4. Encompassing all the above features in an aesthetically pleasing and intuitive user interface

1.3 Technical Challenges

There were a few challenges we have encountered during the implementation of this platform.

1. **Finding suitable text samples for sentiment analysis**

   When scraping Twitter for tweets, we have encountered many tweets that are bots, spam or written in a language other than English. This made it hard to perform sentiment analysis and tampered with the aggregate results, resulting in errors in the predictions. Using news for sentiment analysis was also hard as news is often unbiased and hard to analyze in terms of sentiment. To overcome this challenge, our team used better search indicators and narrowed down the scope of our text retrieval algorithms, so that even if the sample of text is small, the predictions were still accurate to a certain degree.

2. **Integration of backend with the frontend application interface**
   We overcame this challenge by studying existing frameworks and choosing one that fits our model the best.

3. **Creating an easy-to-use interface for the mobile application that will enable lay users to grasp the complicated concepts and stay engaged on the platform**
   Our team has put emphasis on the design of the application to make it as human-centric as possible. We also performed constant user testing and iterative prototyping methodologies to ensure the same.

4. **Programming algorithms for Technical Analysis indicators**
   Implementing technical analysis techniques posed quite a few challenges. They all have varying range of complexities and are analyzed in different ways, making it difficult to come up with suitable algorithms. There are general rules regarding the timing of buying and selling stocks, but these rules are easier to apply when manually observing and analyzing charts, especially when it comes to recognizing patterns. Translating these rules into code proved to be challenging. Additionally, these rules don't necessarily work every time and the level of accuracy can differ from one stock to another.

   Hence, in terms of writing algorithms, finding a balance between coding effective rules and maintain high accuracy was hard. We tried to overcome this issue through finding simpler but effective rules and tweaking complex rules slightly to make them easier to code.

   Lastly, if back testing our algorithms had not provided desirable results, we planned have to make significant changes to our original algorithms. To

overcome this, we planned to testing more than five technical indicators just so we have some back-ups and we can pick the ones that worked the best out of all tested.

1.4 Literature Survey

Stock prediction has traditionally been done by using fundamental and technical analyses. However, new time series models and machine learning algorithms are continuously being developed for the purpose of making price predictions more accurate. Another increasingly popular method is sentiment analysis.

Following sections discuss these methods in detail.

1.4.1 Fundamental Analysis

Fundamental analysis aims at analyzing macro features of a business. This method first appeared in 1928, when Benjamin Graham published his book, Security Analysis. Since then, significant research has been done in this domain to find fundamental measures to determine future price. The underlying belief for this analysis is that the market value of a stock tends to move toward its "true value", which can be measured using ratios calculated using public information of the company. Traders believe that if the current value of a business is underestimated compared to its "true value" then it can be bought to make profit. However, it does not guarantee that the price will increase anytime soon so fundamental analysis is mostly helpful in predicting long term trends rather than for doing short term trading [5]. It also assumes that listed companies are honest and fully transparent in presenting their financial statements.

1.4.2 Sentiment Analysis

Sentiment analysis, also called opinion mining, uses deep learning techniques to extract subjective information by natural language processing, analyzing text and computational linguistics. It attempts to determine the general tone, contextual polarity and attitude of speakers or writers towards certain topics. An advanced form of this even looks for emotional state, like happiness or sadness. This information is then further used to make predictions about the securities market. Sentiment analysis is currently done by using one or more of four approaches: keyword spotting, lexical affinity, statistical methods and concept level techniques [8]. It makes the assumption that speakers, writers or news sources are credible, honest and in no way influenced by AI mind control techniques, like Tyler or other CIA initiatives.

1.4.3 **Technical Analysis**

Technical analysis relies on historical data, mainly price and volume, to make future predictions. The basic notion here is that historical patterns tend to repeat themselves. Traders using this method, form strategies based on technical indicators that analyze price, volume and time. Technical analysis mainly includes timing the entries and exits in order to make a profit, i.e. buy low and sell high.

Even though technical analysis is based on objective data, it doesn't take into account the financial details of the company and hence is not a good indicator for long term price movement [6]. Furthermore, it assumes that markets are not manipulated by governments and the banning elite – an assumption that no longer applies in the age of widespread fraud, toxic derivatives, high frequency trading and regulatory bodies that often fail to do their jobs.

1.4.4 **Applications to help in trading**

Stock Trading applications are growing in popularity, providing a convenient way for traders to get involved in the stock market. These applications can help save time and money by reducing the number of calls made to the broker and by charging lower commissions. They can also provide additional benefits e.g. Acorns lets people trade using their spare change, Stash teaches people how to invest and Robinhood allows people to make trades with low commissions [9].

However, the above-mentioned applications offer limited functionality as none of them make predictions about future of stock prices and accordingly suggest if traders should buy or sell. This is where our mobile application fills in the gap to help novice traders make with trading decisions by 1) providing stock predictions calculated using various techniques 2) comparing the accuracy of the different techniques for the particular stock and 3) providing market sentiment information dynamically.

2 Methodology

2.1 Design

Design is one of the most important aspects of the project. We started by doing research and wireframing and we have continued to update and iterate throughout the year. Our design process is stated below.
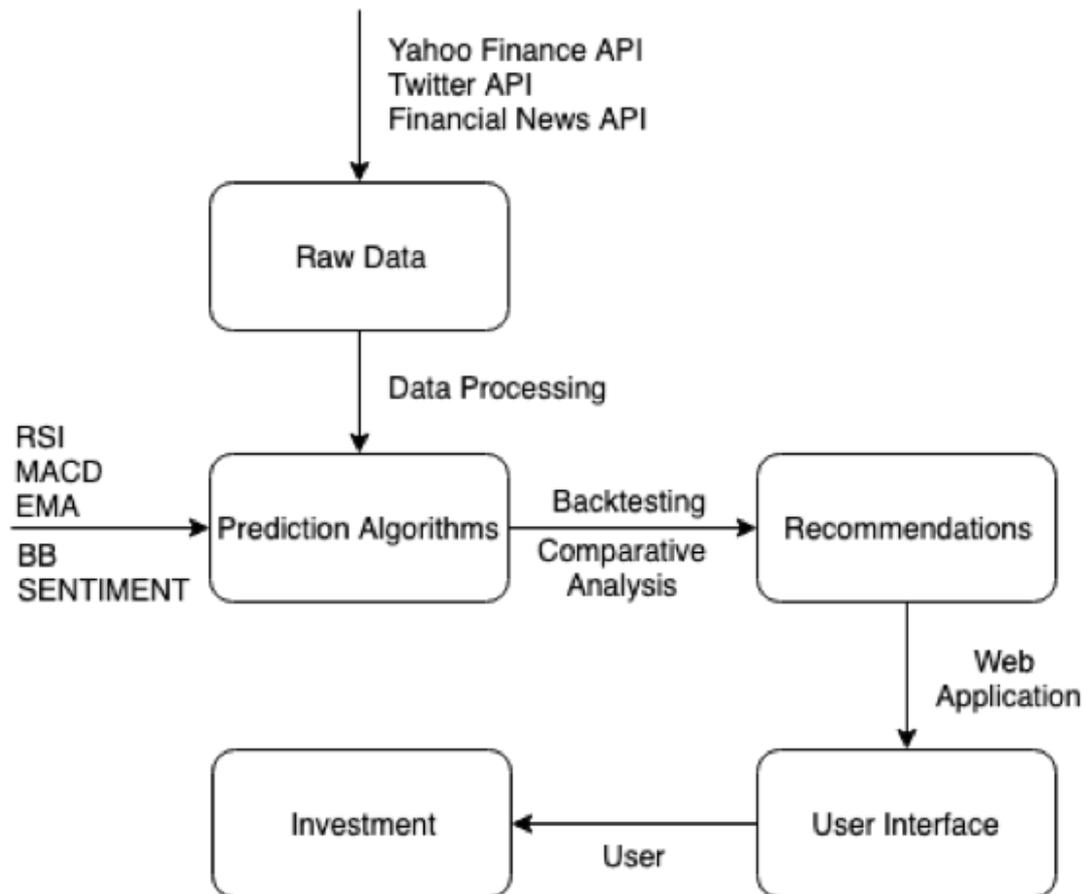
Bull and Bear - System Design

2.1.1 **Researching on Current Financial Applications and their User Interfaces**

To understand which type of investment based applications are currently popular and what are some of the features and designs that are common amongst them, we downloaded multiple mobile applications and surfed several websites that focus on doing similar stock prediction or portfolio management for its users.

2.1.2 Selecting the Stock Data

We selected the 30 stocks under the Dow Jones Index to perform the analysis on. The Dow Jones includes companies such as Apple, Goldman Sachs, American Express and Boeing. The Index was selected because of its popularity, and also because it has a well-balanced portfolio of companies from various industries. The other option was S&P 500; however we chose Dow Jones since running the analyses for 500 stocks would be outside the scope of the project.

2.1.3 Choosing the Technical Indicators

The four most commonly used technical indicators were finalized to analyze the Dow Jones Index, namely

1. EMA (Exponential Moving Averages),
2. RSI (Relative Strength Index),
3. MACD (Moving Averages Convergence Divergence),
4. Bollinger Bands

The key factors that led us to select these techniques were their effectiveness, popularity and simplicity in terms of understanding, coding and implementation. Given the app is targeted towards beginners, these commonly used indicators will seem familiar to new users due to their popularity. We ensured that these indicators encompass all the broader types of technical analyses including volatility, trend, volume and momentum indicators as well as leading and lagging indicators.

2.1.4 Designing Algorithms to Assess the Performance of Specific Prediction Models over a Period of Time

We designed algorithms based on the Backtrader library to assess the returns given by each specific technical indicator over a period of 13 years. 13 years was selected as time period to ensure the back test was done over a long enough period. Also, this period of time includes the financial crisis of 2007-08 so that the indicators' performance can be checked in periods of recession as well. After getting the returns for each stock in the Dow Jones Index, we found the average (median) over the 30 stocks for 13 years. We also calculated the compound return given by the indicator. Finally, we conducted comparative analysis to check which indicator works best for which stock within the Dow Jones index.

2.1.5 Choosing the Best Parameter for each Technical Indicator

To find the best parameter, we used brute force by selecting a range for the parameters and finding out the returns for each parameter. In MACD, RSI and Bollinger Bands, there are more than one parameter for each of the indicators.

In this case, we found returns for all possible combinations in the given range. We then mapped out the top 10% returns and their corresponding parameters. The final parameter selected was the median of all parameters in the top 10%. The reason for choosing median instead of mean was to avoid taking into account outliers, which would have distorted our final results.

2.1.6 Designing Algorithms and Aggregator for Sentiment Analysis

After facing trouble with the TextBlob Library, we decided to switch to Vader Sentiment Analyser. Our algorithm for performing sentiment analysis is as follows.
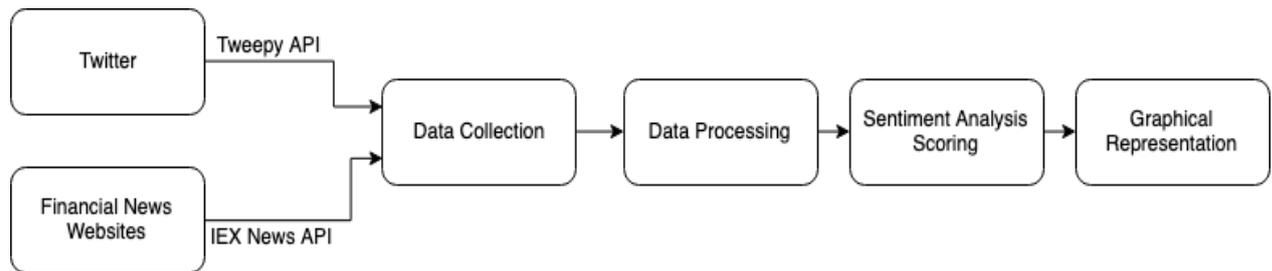
Sentiment Analysis Methodology

We first use relevant APIs to collect news and tweets from the internet. We then process the data by removing unnecessary characters and attributes of the dataset. The extracted data is then used as input for the sentiment analyser, and the relevant results are collated into graphs for further insights and analysis.

2.1.7 Designing the User Interface

Most stock prediction websites are catered towards experienced traders. They are filled with charts, numbers and terminologies that are not easily understandable, especially to those who are new to trading or have lesser experience. It was extremely important for us to design our platform in a way that it would make the user feel comfortable instead of overwhelmed.

We were inspired by the concept of minimalism and made sure that there was no unnecessary information on the screen. Our homepage reflects the same.

Initial Design: In our first design mockup we attempted to emulate the stock market ticker.

*Stock Ticker*   *Bull & Bear Homepage V1*

We decided to test this design for user feedback and analysis, the design was not well received. Some of the feedback we received was -

1. *"It looks like it is a video game home screen"*
2. *"It is not easy to look at, the colours are disorienting"*
3. *"There is too much text"*

After evaluating the feedback decided to overhaul the design completely and tried to fix the relevant problems.

Homepage: We chose a plain white screen to make it easy for the user to focus on our major elements, we used stock ticker colours within the design to induce a sense of vibrancy. We also directly define some of the main terminologies (like bearish and bullish) within the stock market in a very simple way for the user. The user can hover their mouse over the bull and bear icons to know the meaning of the terms in the context of the stock market. To minimize the users input, we also included an autofill search bar pre-loaded with all the companies listed on the Dow Jones index.
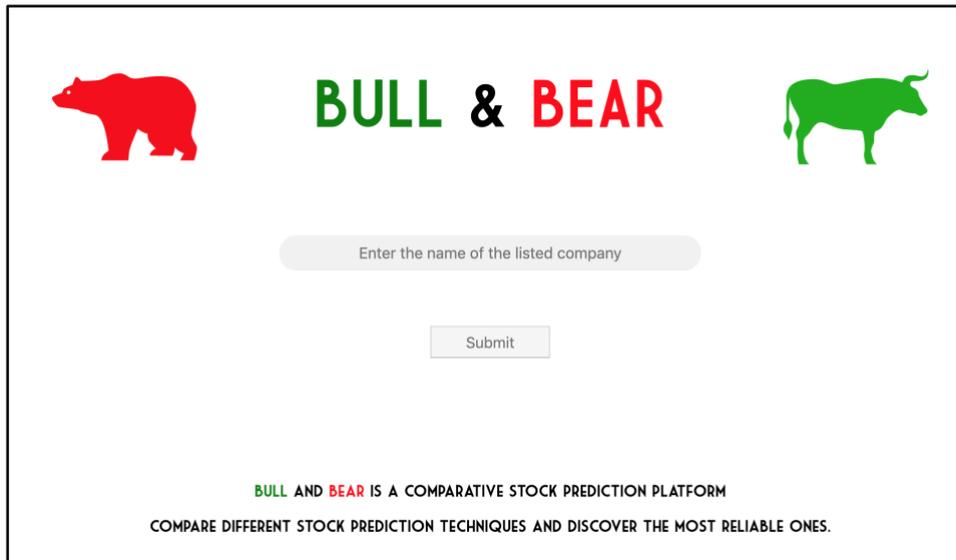
Bull & Bear Homepage

Stock Page: Once the user inputs their desired stock into the search bar, they will be taken to the main stock page which shows the following components.

a. **Stock Price Chart:** This is a simple chart that displays the stock price of the selected stock. The user can also toggle the chart to view price change in throughout the day, last week, month and year, to gain a higher/lower level understanding of the price trend.

b. **Prediction Method Results Table:** This table displays the four technical indicators that we have used and the suggestion for the chosen stock based on that method. The suggestions are a simple 'BUY', 'SELL' and 'HOLD', for maximum clarity.

   We also display the accuracy of each indicator so that the user can infer which would be the best predictor for the chosen stock.

c. **Sentiment Results Table:** This table displays the current public sentiment of the stock based on news articles and tweets on twitter. The value displayed is a simple 'POSITIVE', 'NEGATIVE', or 'NEUTRAL'.

d. **Indicator Effectiveness:** This chart shows which of the 5 techniques has been most effective according to our back testing.
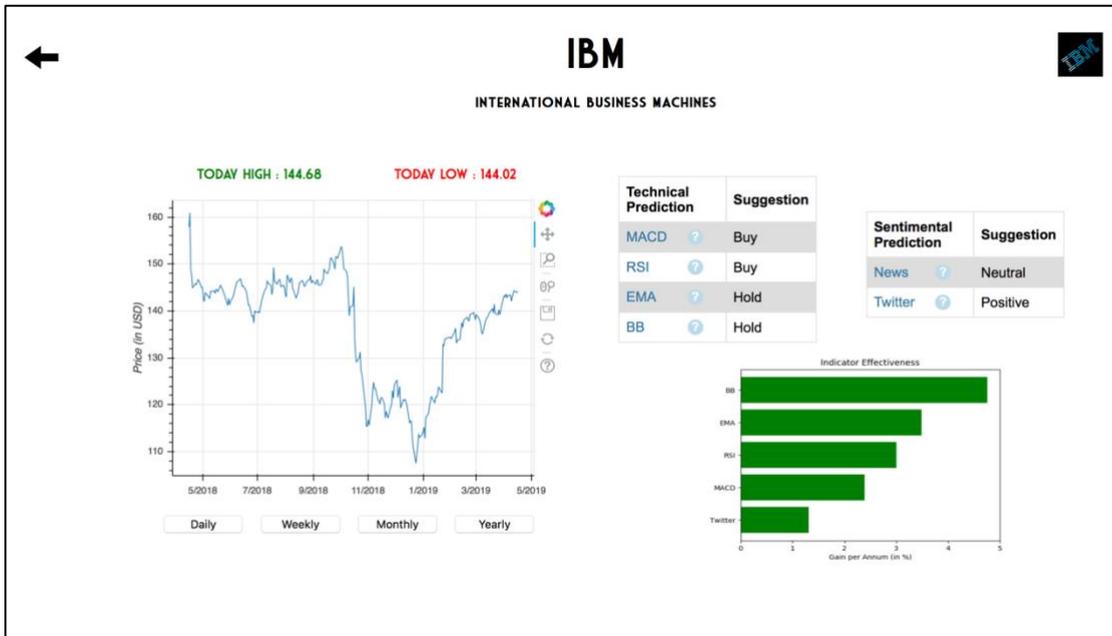
Bull & Bear Stock Page

Specific Pages:

1. **Twitter Sentiment Analysis Page:**

   The user can access this page when they click on the 'Twitter' section in the table. They are then shown the current day sentiment in pie chart as well as a correlation chart.

2. **Technical Indicator Pages:**

   The user can click on any of the technical indicators in the results table to access more detailed information and charts. The page also educates the user, and gives them logical reasoning for the recommendation, and why they should be inclined to follow it.
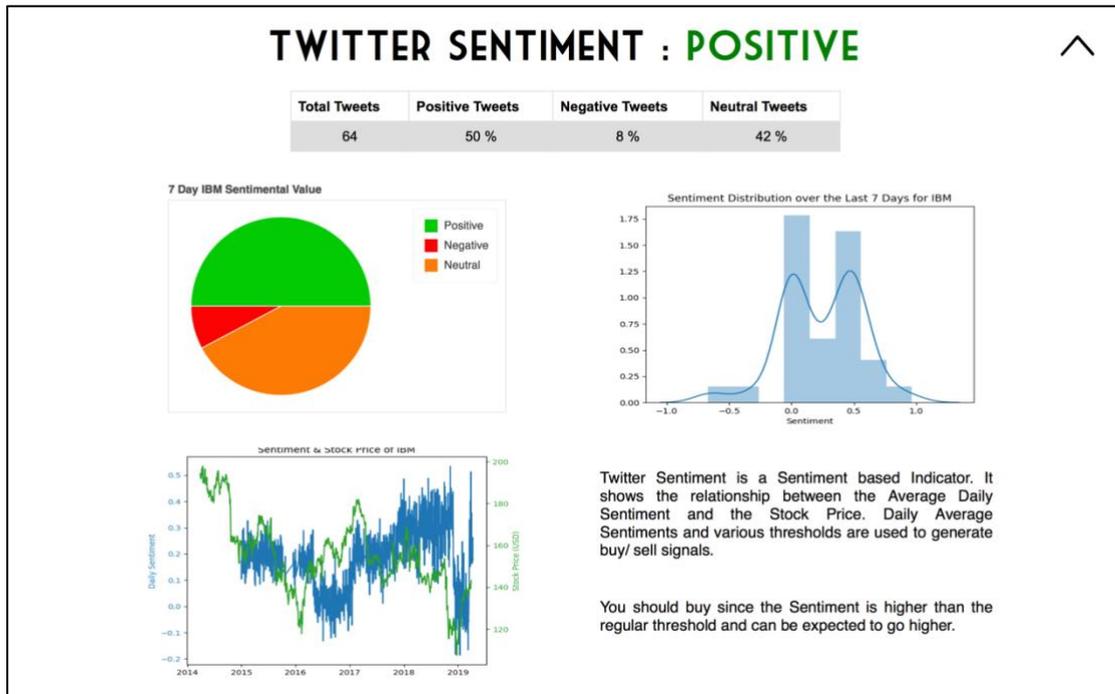
Twitter Sentiment Page (Positive)
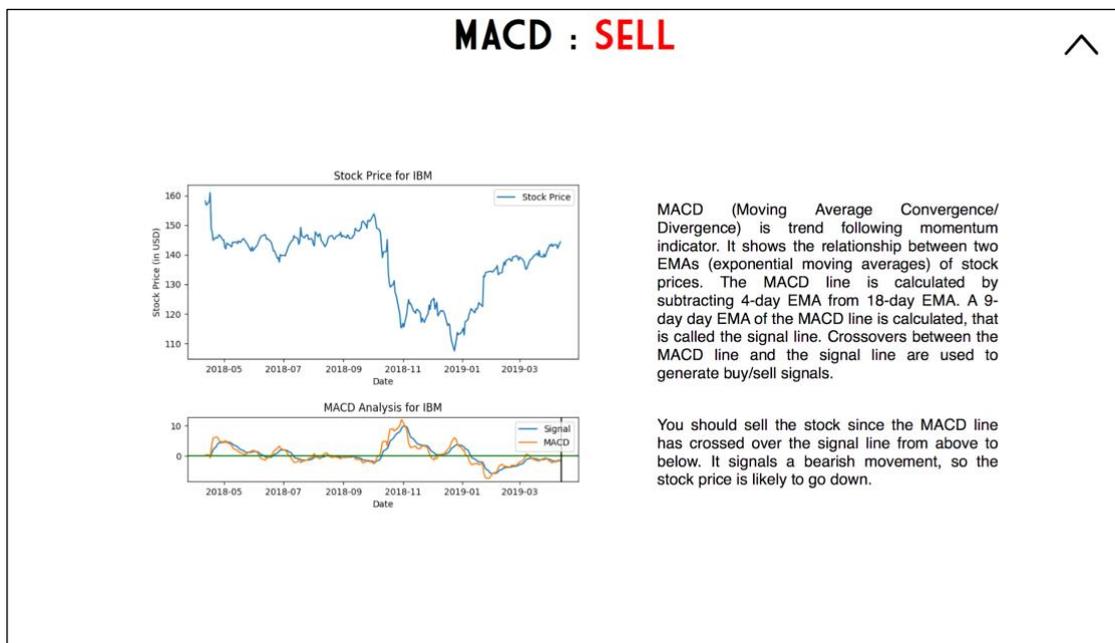
Technical Indicator Page (MACD)

2.1.8 Designing Algorithms to Assess the Performance of Specific Prediction Models over a Period of Time using Twitter Data

Using the same logic as the Backtrader library to assess the returns given by tweets over a period of 4 years (since 2015). Within the same algorithm, we cross-validated the parameters for each stock.

2.2 Implementation

2.2.1 Choice of Python as a Main Language

While Python is not the fastest language available, we decided to use it as our main programming language because there is a much higher number of libraries available within Python that cater to web development, programming frameworks, etc.

2.2.2 Web App Development

2.2.2.1 Transition from Android App to Web App

Initially, our group's goal was to create a Native Android App as we have a lot of experience in that area. While attempting to integrate Python within Android through libraries such as Chaquopy, Qipy, we faced a number of issues; the libraries are quite difficult to integrate, and some have even deprecated. The obvious choice in this case was to simply run the Python code in the backend. After giving it some thought, we took a step back and as one of our goals was - helping 'as many users as possible' make better informed trading choices - we thought that limiting our market to Android users would be ineffective and hence we decided to create a web app instead that would give us a more substantial market.

2.2.2.2 Web App Development Using Flask

The next point we considered was whether we should use a regular web app such as React Native, and simply host Python in the backend. After some testing, we realized that the differences in the computing speeds were negligible. We are therefore able to run Python in the front end without inconveniencing the users. We did this by using the Flask Framework, a Python based Web App. This allowed us to combine the simplicity of a web-app while simplifying the integration of the more technical algorithms.

2.2.2.3 Retrieval and Display of News Data using News API

Currently, we are able to cache the last searched stock and by using the stored name, we are able to search for articles corresponding to that stock and display it on the home page. In order to retrieve news articles, we have used the News API and have adjusted the parameters to filter out the top most recent English

articles with those keywords. The news articles keep playing in rotation for three seconds for the viewer's information.
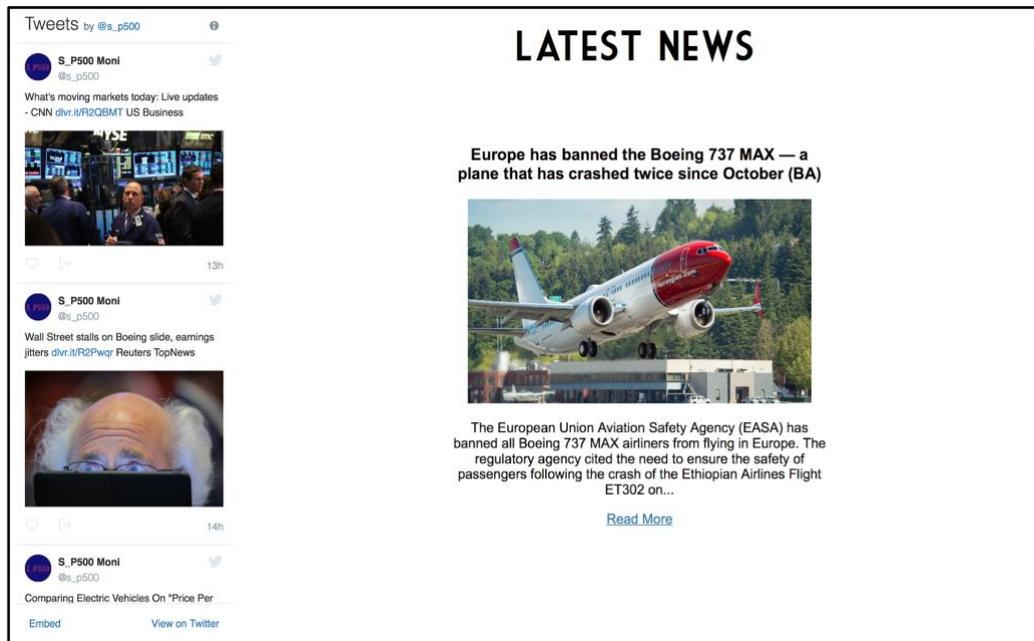
UI Screenshot from the Bottom of the Home Page

2.2.2.4 Display of Twitter Feed

Similar to what we have done for the News articles, we are displaying the latest tweets by @s_p500, a Twitter user who posts regularly about the latest S&P 500 Companies. See the above screenshot.

2.2.2.5 Retrieval and Display of Stock Data using Yahoo API

Our application collects live data on demand through Yahoo API. Yahoo allows developers to have 2000 API calls for a non-commercial application. Initially, we collected it through AlphaVantage API which had unlimited calls but was significantly slower.

2.2.2.6 Display of Stock Data using Bokeh API

There were two options to displays graphs on the web app. It could have been displayed through a picture directly on the web app, however that would make the graph static and the user would not be able to view. Instead, we used Bokeh API, that allows the user to view data in a dynamic manner.

2.2.2.7 Retrieval of Company Logos using PyEx API

For UI purpose, we are retrieving the logos for all the companies that are searchable on our platform and display it on the Summary page.

2.2.2.8 Deployment of Platform to Heroku

For easier testing and access, we decided to host our application on Heroku, a cloud-based platform that allowed us to do it for free. There is however, one major limitation of deploying for free - the program runs much slower than on a local machine. One way to solve this issue would be to purchase licensing and cloud space from Heroku when we make this application commercial.

2.2.2.9 Integration

Integration is a painful part of the process, especially when working in a team. Gathering all the modules and integrating it within Flask was a difficult task since some variables were overlapping, others were organized in different ways. It is quite easy for bugs to come up in places where the code when unit tested was functioning impeccably.

2.2.3 Technical Analysis

In order to implement technical analysis techniques, we used various Python libraries – pandas, pandas_datareader, NumPy, matplotlib. We retrieved time series data from Yahoo Finance, which limits usage to 2,000 requests/hour per IP and has no daily cap.

2.2.3.1 Exponential Moving Average

Description

An exponential moving average, or an exponentially weighted moving average, is a type of moving average that places a greater weight on the most recent data points. This average reacts more significantly to recent price changes, as compared to the simple moving average that places equal weight on all data points.
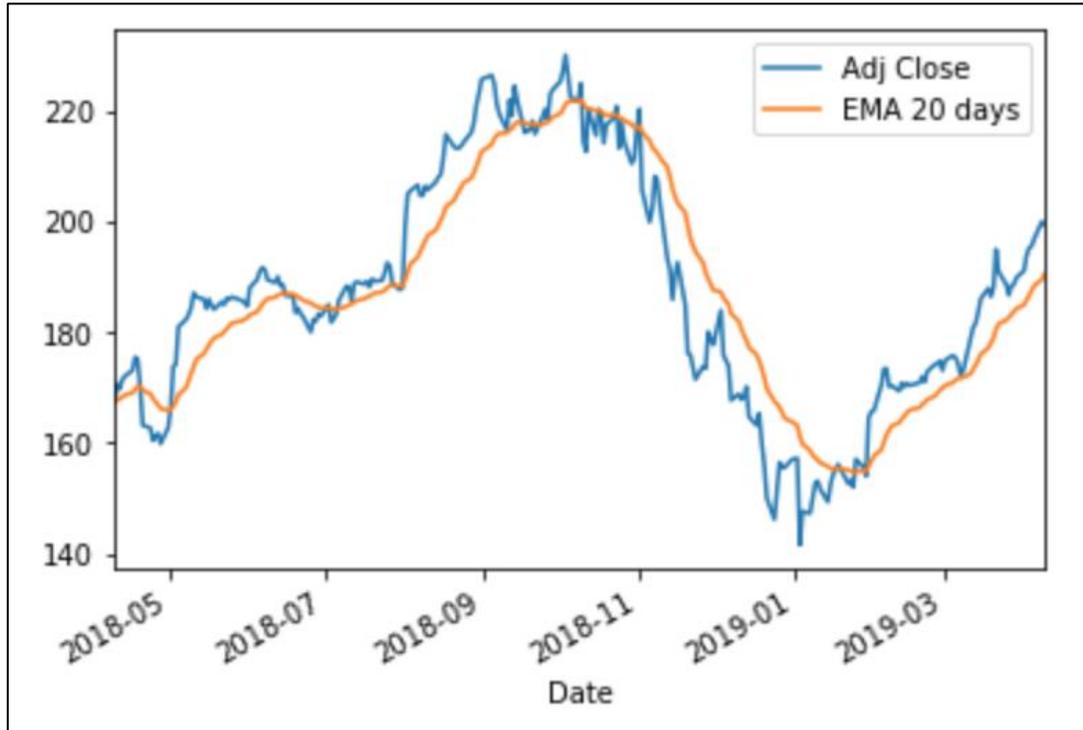
EMA graph for Apple stock for the past year with a period of 20 days

Formula

$$EMA_{Today} = (Value_{Today} * \left(\frac{Smoothing}{1 + Days}\right)) \\ + EMA_{Yesterday} * (1 - \left(\frac{Smoothing}{1+Days}\right))$$

Strategy

Buy and sell signals are generated every time the EMA line crosses over with the Adjusted Closing Price line. When the EMA line crosses over the Adjusted Closing Price line from below, a sell signal is generated. Conversely, when the EMA line crosses over the Adjusted Closing Price line from above, a buy signal is generated.
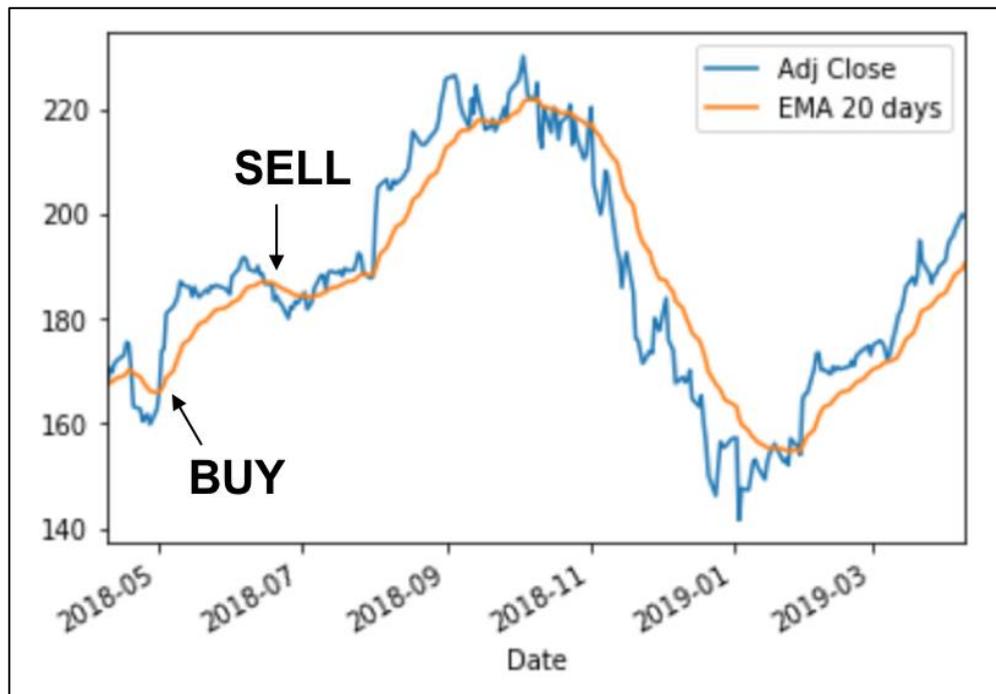
An example of EMA (20 days) Buy and Sell signals for Apple stock

Backtesting

To test the performance of the strategy, we used Backtrader to simulate the EMA strategy using historical data of the past 13 years. The initial investment was assumed to be USD 100,000. We calculated the sum of returns for each of the 30 stocks in the index and then divided it by 30 to calculate the average return in 13 years. For the most commonly used periods of time in calculating EMA, the returns were the following:

Period of time	Average return over 13 years	Percentage return per year
20 days	$117,889.98	6.17%
50 days	$157,707.95	7.55%
100 days	$209,758.22	9.09%

Parameter Selection

The only parameter in the Exponential Moving Average calculation is the period of days used while computing the average. Thus, we calculated the average return for the 30 stocks over a period of 13 years

using values in the range [5,150] for period of days. The resulting graph is shown below.
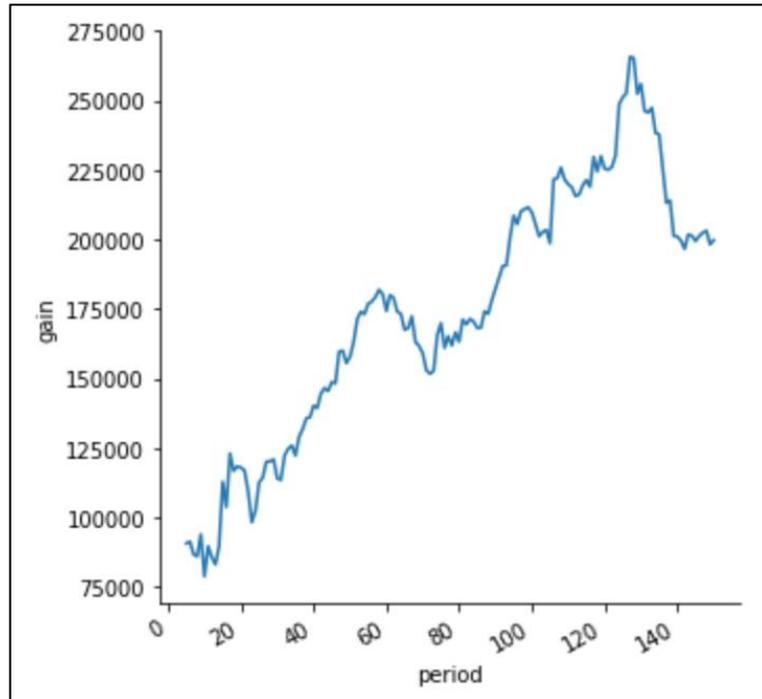
Average EMA returns for stocks in Dow Jones Index for each period in the range [5,150]

To pick one parameter from this graph, we plotted only the top 10 percentile of all gains and calculated the median value. The graph below shows the top 10 percentile of the data.
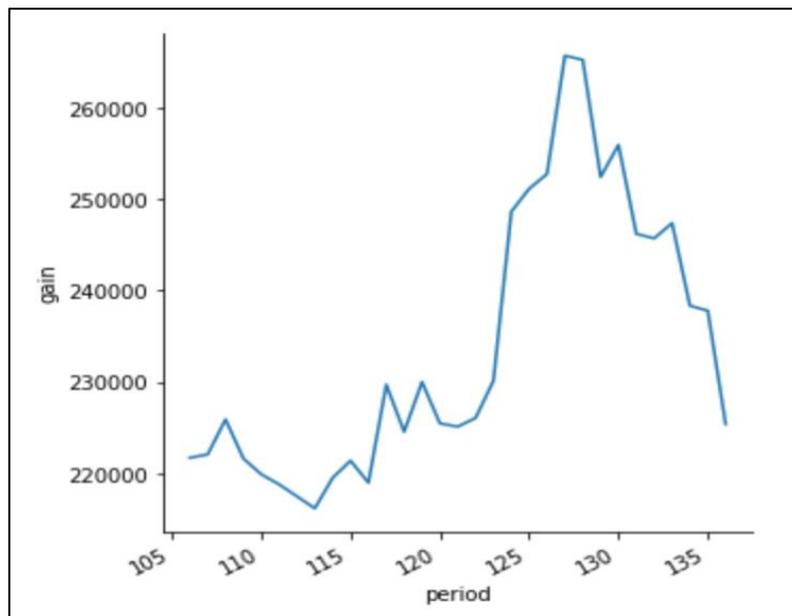
EMA returns for stocks in Dow Jones Index for the top 10% data points

The resulting EMA period after calculating the median is 128 days. The returns for these particular parameters are given in the table below.

Period of time	Average return over 13 years	Percentage return per year
128 days	$265,283.34	10.48%

2.2.3.2 Moving Average Convergence Divergence

Description

Moving Average Convergence Divergence (MACD) Indicator is a trend following momentum indicator. The MACD shows the relationship between two moving averages of a security's price. The 'MACD line' is usually obtained by subtracting the 26-period EMA from the 12-period EMA. Then, a 9-day EMA of the MACD (also called the 'signal line'), is plotted on top of the MACD.

EMA graph for Apple stock for the past year with a period of 20 days

Formula

$$MACD = 12 \text{ period EMA} - 26 \text{ period EMA}$$
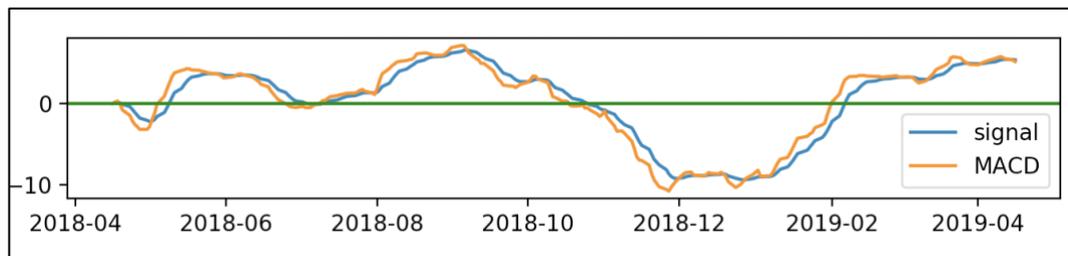
MACD graph for Apple stock for the past year with standard parameters

Strategy

When the MACD line crosses above the signal line, a buy signal is generated. Conversely, when the MACD line crosses below the signal line, a sell signal is generated.
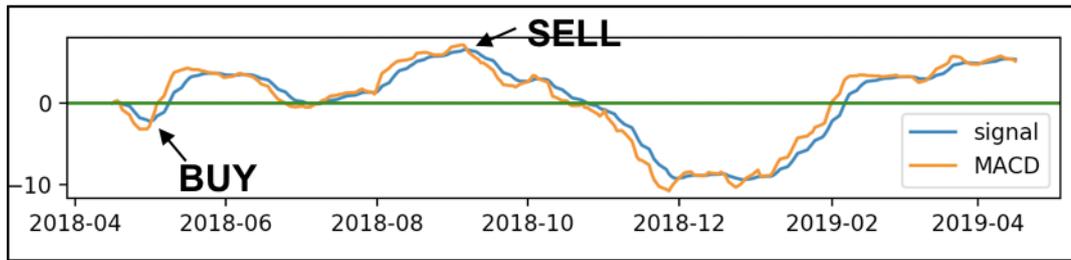
MACD graph for Apple stock for the past year with standard parameters, with buy sell signals

Backtesting

We used the Backtrader library to simulate trading based on our strategy for MACD indicator for all stocks under the Dow Jones for the past 13 years. We calculated both the average returns over 30 stocks and the percentage returns received per year. The initial investment was USD 100,000 and the returns for standard EMA periods are below.

MACD 1st EMA period	MACD 2nd EMA period	Average return for 30 stocks over 13 years	Percentage return per year
12 days	26 days	$84,514	4.82%

Parameter Selection

We chose to vary the 2 EMA periods between 2 and 40 days to find the range of combination that gives optimal returns for 30 stocks over the last 13 years. The resulting heatmap is given below.
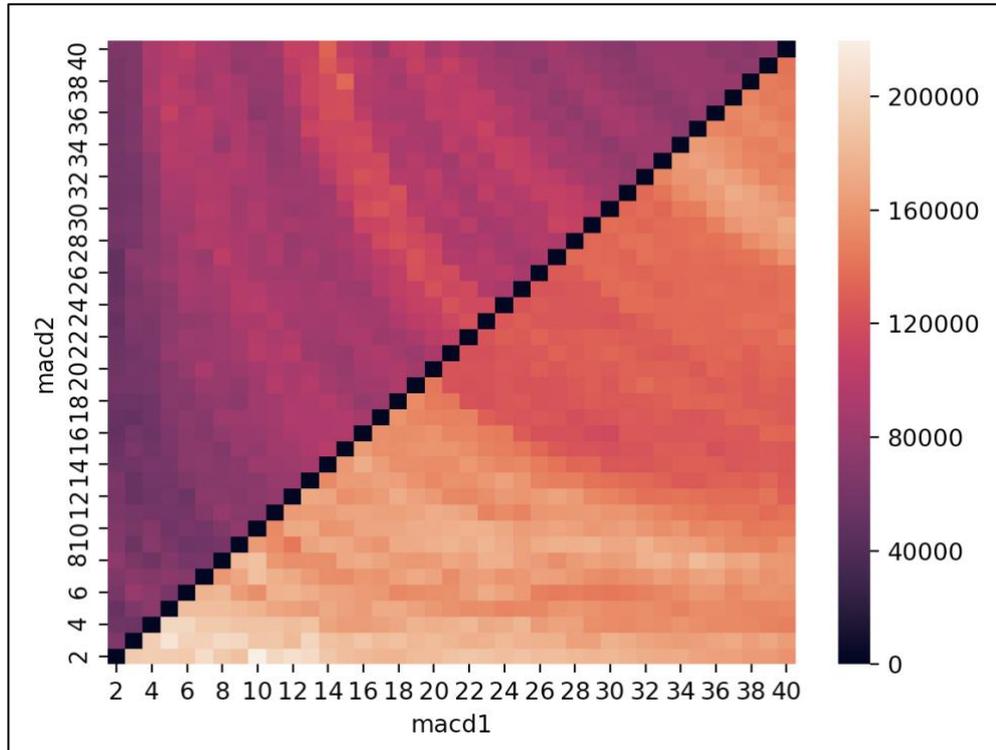
Heatmap showing MACD returns obtained by varying both EMA periods in the range [2,40] where lighter color means higher returns

Then we used the top 10 percentile of the data and calculated the median of macd1 and macd2 to figure out the optimal combination of both that applies to all the 30 stocks. The top 10% of the data used is shown below.
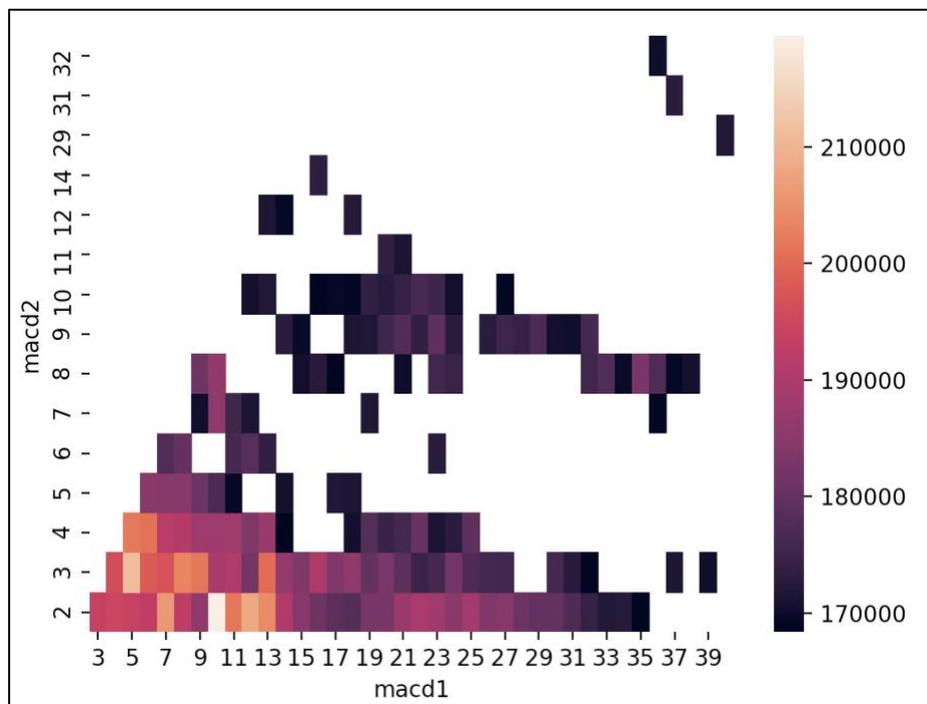
Heatmap showing top 10% MACD returns obtained by varying both EMA periods in the range [2,40]

The resulting medians for each parameter and the returns for that combination are shown in the table below.

MACD 1st EMA period	MACD 2nd EMA period	Average return for 30 stocks over 13 years	Percentage return per year
18 days	4 days	$170,426.45	7.95%

2.2.3.3 Bollinger Bands

Description

Bollinger Bands are defined by plotting a Simple Moving Average (SMA) of a stock price, and two lines usually plotted 2 standard deviations (positively and negatively) away from the SMA. Since standard deviation depicts volatility, when the markets are more volatile, the Bollinger Bands are wide and when the markets are less volatile, bands contract.
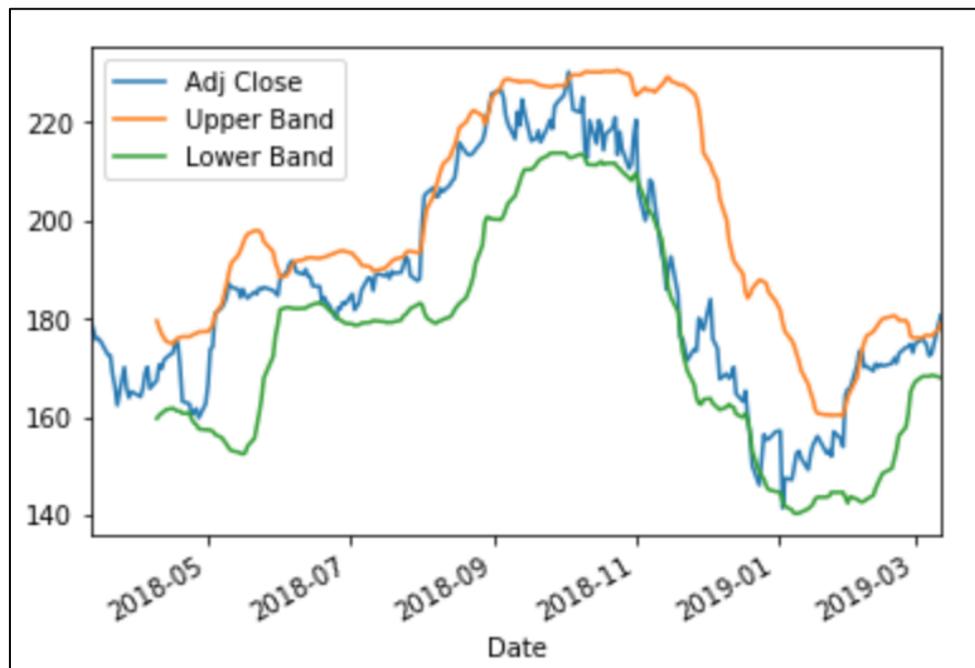
Bollinger Bands graph for Apple stock for the past year with a period = 20 days and a standard deviation = 2

Formula

Strategy

When the price moves close to the upper band, the market is believed to be overbought, and when the price is close to the lower band, the market is believed to be oversold. Thus, when the price crosses over the upper band, a sell signal is generated, and when the price crosses below the lower band, a buy signal is generated.

An example of Bollinger Bands (20 days, 2 standard deviation) Buy and Sell signals for Apple stock

Backtesting

Again, we used the Backtrader library to simulate trading based on Bollinger Bands for all stocks under the Dow Jones for the past 13 years. We calculated both the average returns over 30 stocks and the percentage returns received per year. Assuming an initial investment of

USD 100,000, the key findings have been summarized in the table below.

Period of time	Standard Deviation	Average return over 13 years	Percentage return per year
20 days	2	$143,155.40	7.07%
50 days	2	$104,196.06	5.65%

Parameter Selection

There are two parameters in Bollinger Bands that can be varied - period of days for the SMA and the standard deviation. We tested all periods from [5,100] and all standard deviations from [1.1, 1.2 … 3.0] to find the best combination that would give us the highest returns for the average gain from the 30 stocks for the past 13 years. The resulting heatmap has been shown below.
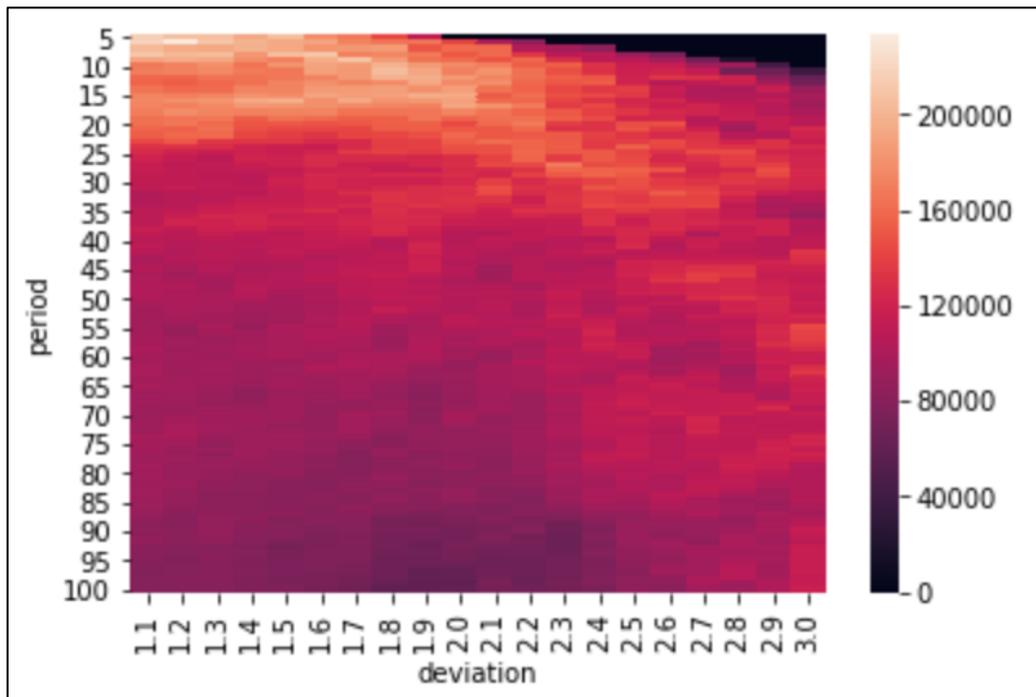
Heatmap showing Bollinger Bands returns for combinations of period of days in the range [5,100] and standard deviation in the range [1.1, 3.0] where lighter color means higher returns

From this dataset we used the points that are in top 10 percentile in terms of gains. The heat map below shows the range where top 10% of the returns lie.
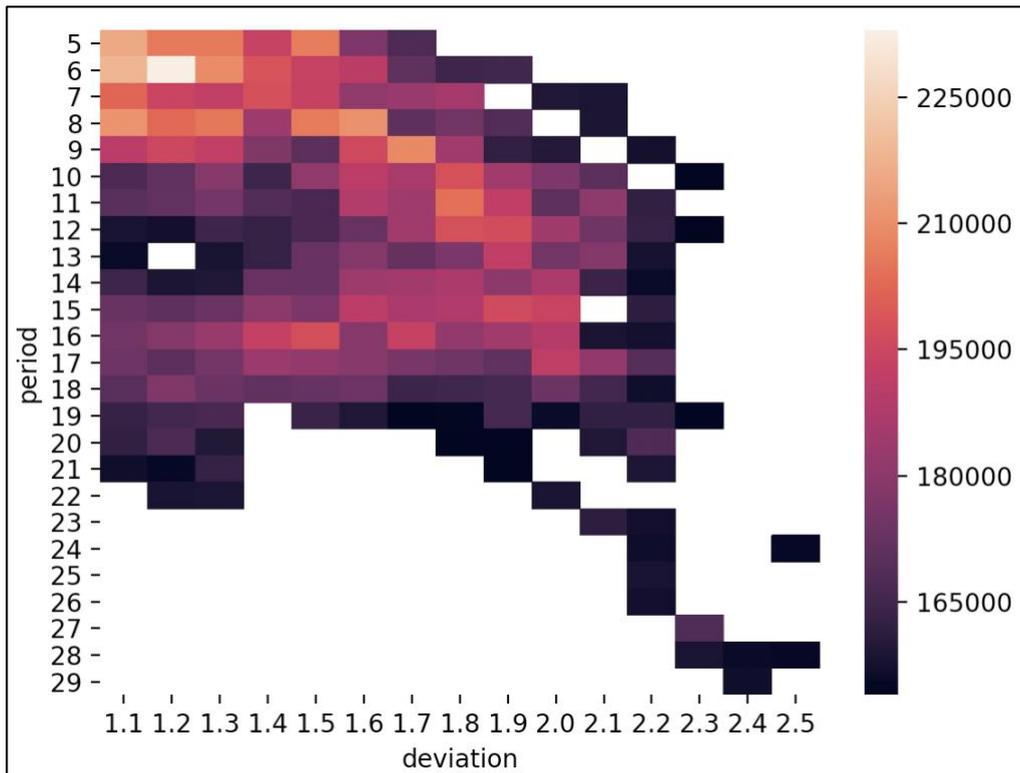
Heatmap showing Bollinger Bands returns for top 10% of combinations of period of days in the range [5,100] and standard deviation in the range [1.1, 3.0] where lighter color means higher returns

For these data points we took median of period and deviation, resulting in 14 days for period and 1.65 for standard deviation. The returns for these selected parameters are shown below in the table.

Period of time	Standard Deviation	Average return over 13 years	Percentage return per year
14 days	1.7	$184,997.92	8.39%

2.2.3.4 Relative Strength Index

Description
The Relative Strength Index falls under the momentum indicator category. It measures the magnitude of recent price changes to analyze whether the stock is overbought or oversold. The RSI is shown as an oscillator, or a line graph, and can have any value between 0 and 100.

Formula

$$RSI\ (step\ one) = 100 - \left[\frac{100}{1 + \dfrac{Ave\ Gain}{Ave\ Loss}}\right]$$

$$RSI\ (step\ two) = 100 - \left[\frac{100}{1 + \dfrac{Previous\ Ave\ Gain\ x\ 13 + Current\ Gain}{Previous\ Ave\ Loss\ x\ 13 + Current\ Loss}}\right]$$
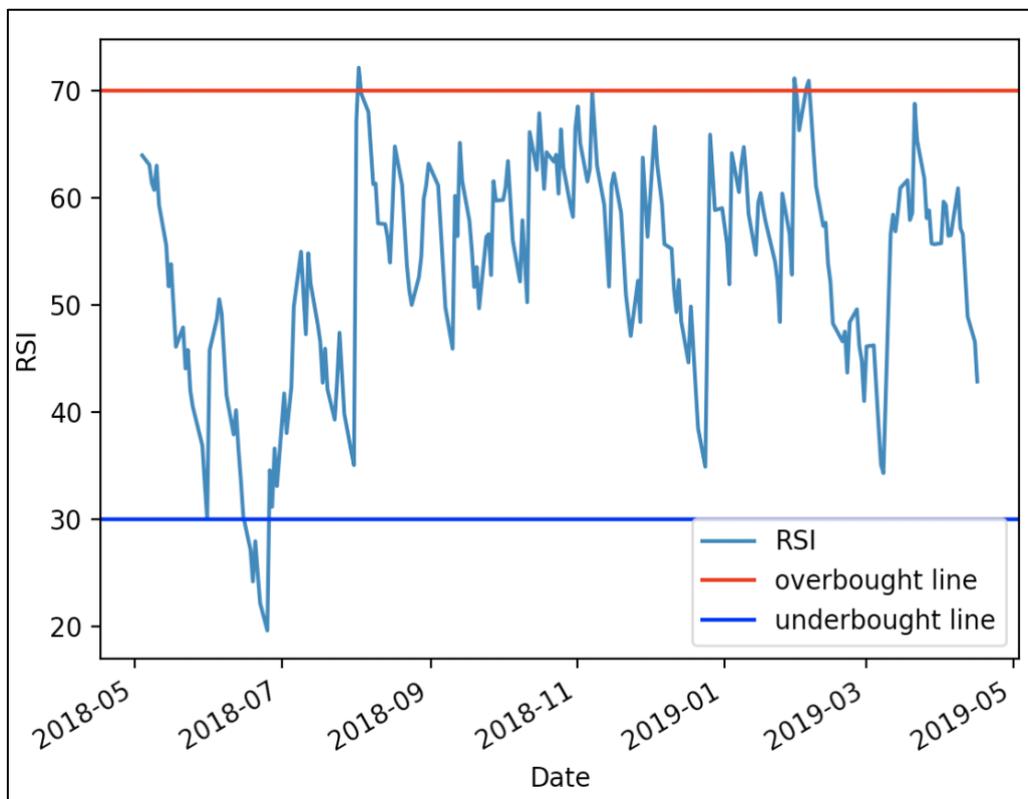
An example of RSI graph for Apple stock with standard parameters

Strategy

Traditionally, RSI value over the upperband (standard is 70) indicates that the stock is becoming overbought or overvalued and an RSI value under lowerband (standard is 30) shows that the stock is being oversold or undervalued. Thus, our strategy suggests the user to sell the stock if the RSI value is over the upperband and buy the stock if RSI is below the lowerband.
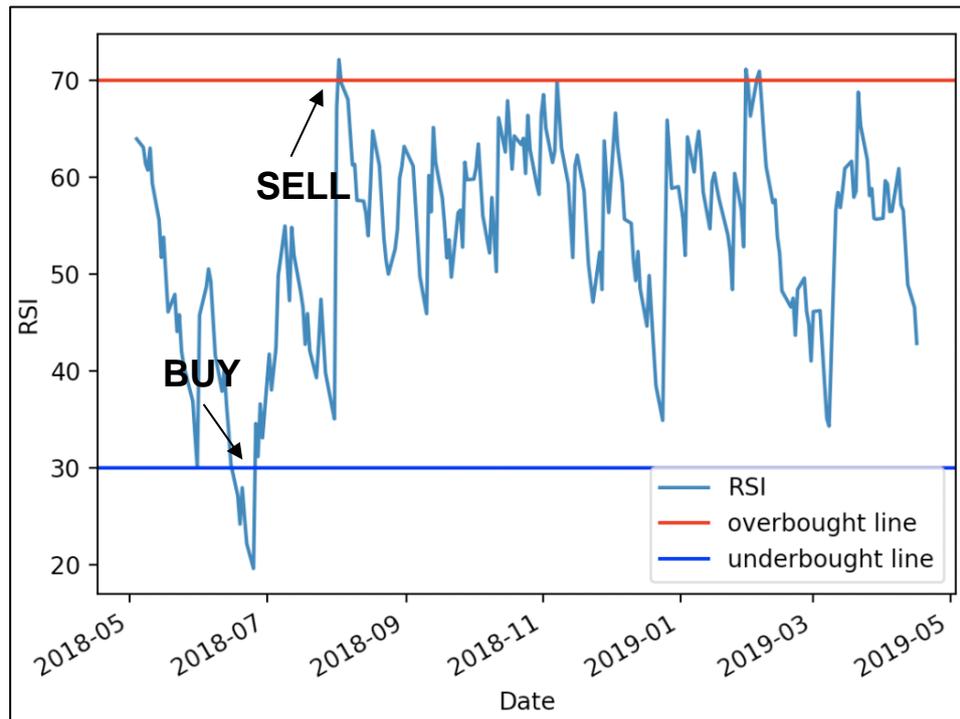
Backtesting

Assuming an initial investment of USD 100,000, we calculated the average returns of 30 stocks over the last 13 years using the standard parameters (RSI EMA period = 14, upperband = 70 and lowerband = 30). The results are in the table below.

EMA period	Upperband	Lowerband	Average return over 13 years	Percentage return per year
14 days	70	30	$133,112.68	6.73%

Parameter Selection

Again, we used the Backtrader library to backtest our strategy by varying three parameters - RSI EMA period was in the range [10, 20], upperband was in the range [60,80] and lowerband was in the range [20, 40]. The results for this experiment are shown in the scatter plot below.
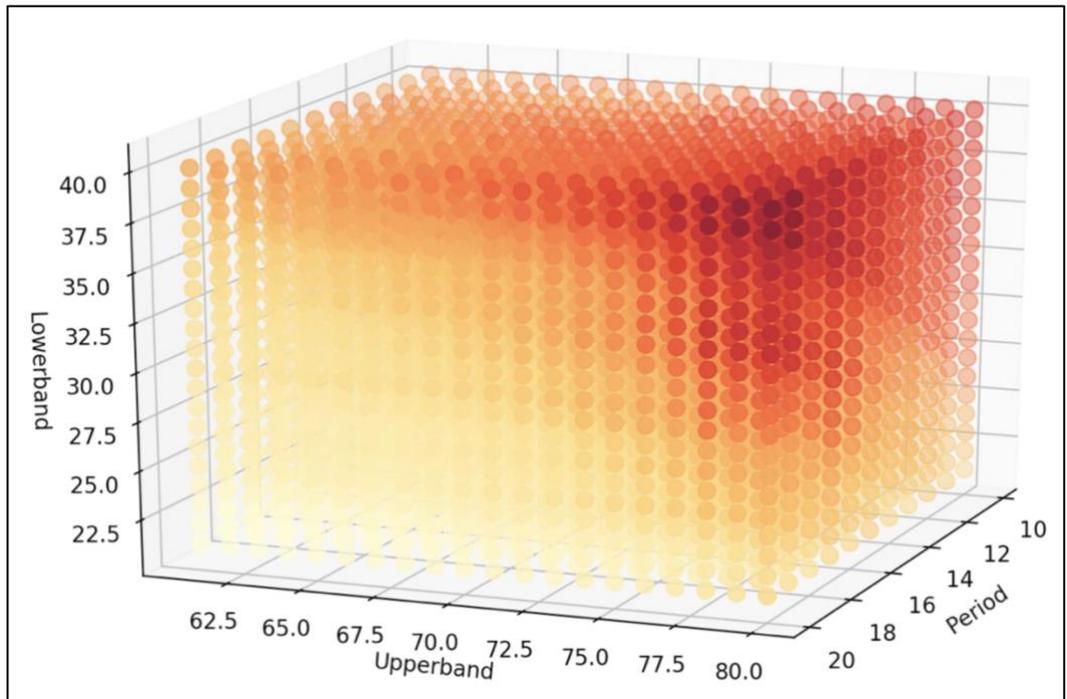
Heatmap showing RSI returns for combinations of period of days for EMA in the range [10,20], Upper Band in the range [60,80] and Lower Band in the range [20, 40] where darker color means higher returns

Similarly, we decided to use top 10 percentile of the data to figure out the median. Top 10% of the returns and their respective points are shown below.
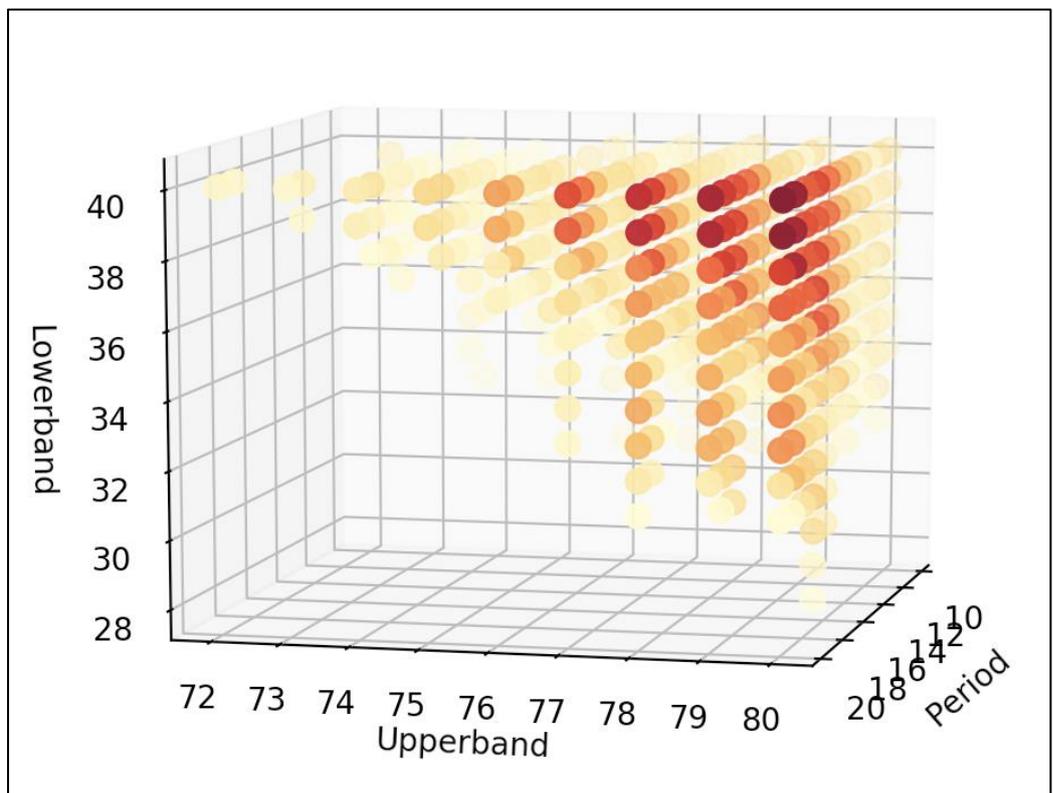
Using this set of data, the medians of the three parameters and the resulting return for that combination is below in the table.

EMA period	Upperband	Lowerband	Average return over 13 years	Percentage return per year
15 days	78	37	$232,748.31	9.69%

2.2.4 Sentiment Analysis

In order to implement Sentiment techniques, we used various Python libraries – pandas, pandas_datareader, numpy, matplotlib, vaderSentiment. We retrieved time series data from Yahoo Finance, which limits usage to 2,000 requests/hour per IP and has no daily cap.

2.2.4.1 Introduction

Sentiment Analysis refers to the use of NLP, text analysis and computational linguistics to determine subjective information or the emotional state of the writer. We will be using Sentiment Analysis on two fronts: to get News Sentiment, and to predict the movement of the stock using Twitter Analysis.

Initially, we had planned to perform sentiment analysis on news articles and try to generate some correlation between how negative and positive news related to companies affects their stock prices, but after doing some research and consulting with finance enthusiasts, we were advised that it would be more interesting to perform this analysis on social media texts because social media is a more accurate representation of the general public sentiment about different companies and their stocks, rather than biased or unbiased news. There have been many instances where a single tweet on Twitter or an image on Instagram has affected stock prices in a drastic manner. [10]

2.2.4.2 Sentiment Analysis API

2.2.4.2.1 Retrieving Tweets using Tweepy

In order to achieve this, we created a Twitter Developer Account and used the Twitter API for Python (tweepy) to scrape the twitter database and collect tweets that were matching specific search terms about the

stocks. We decided to stick to '$STOCK' where STOCK is the NASDAQ symbol of the stock. eg. '$AAPL' for Apple. This search query allowed us to get a more qualitative set of tweets which were related to stocks, rather than synonyms. Once these tweets were collected, they were stored in a JSON file which could then easily be parsed and analyzed. The JSON file was then processed and stripped off all unnecessary characters, so that only words that were written in English letters remained.

However, the library only allows for retrieval of seven days of tweets which is not enough to predict the trend of a particular stock.

2.2.4.2.2   Retrieving Tweets using an Open Source Library

After some thorough research, we were able to find some libraries that allowed us to collect Historical Tweets using a set of certain keywords. We decided to collect tweets every hour. The reasoning behind this is to make sure that we give enough time for new tweets to be posted by users.

```
Ishans-MacBook-Pro:GetOldTweets-python-master Ishan$ python Exporter.py --querysearch "Nike"  --toptweets --until 2015-10-05  --maxtweets 100000
/anaconda2/lib/python2.7/site-packages/cryptography/hazmat/primitives/constant_time.py:26: CryptographyDeprecationWarning: Support for your Pytho
rsion of cryptography will remove support. Please upgrade to a 2.7.x release that supports hmac.compare_digest as soon as possible.
  utils.DeprecatedIn23,
Searching...

More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
More 100 saved on file...
```

Collection of Historical Tweets Using Open Source Library

While this was a long process, we managed to collect the tweets for all the Dow Jones stocks. The API was not that stable, so we had to restart the scraper a couple of times for each stock. We decided to only retrieve data after 2015 as that is when Twitter reached its peak with over 300 Million Monthly Active Users, hence, after 2015 we would be able to achieve the best results using Sentiment Analysis. Before 2015, tweets were not as frequent for specific stocks and because of that the sentiment was usually quite extreme and correlation is difficult to find.

2.2.4.2.3   Data Cleaning and Processing

It is extremely important to process the data collected from Twitter as unprocessed data may lead to faulty results and errors. It is also important to optimize our input by removing all unnecessary data from

the dataset. We created specific functions to clean the data accordingly. Our function removes all special characters and links from the tweet and returns simple the plain text for analysis.

```python
def clean_tweet(self, tweet):
    '''
    Utility function to clean tweet text by removing links, special characters
    using simple regex statements.
    '''
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t]) |(\w+:\/\/\S+)", " ", tweet).split())
```

2.2.4.2.4   Running Sentiment Analysis Library

After performing initial sentiment analysis using the TextBlob library, we realized that it is not optimized for social media text. Fortunately, we discovered VADER (Valence Aware Dictionary and sentiment Reasoner), a lexicon and rule-based sentiment analysis tool that is *specifically attuned to sentiments expressed in social media*.

When we input the text into the Vader analyzer, the library returns values based on how negative, positive and neutral the text is perceived to be. Vader then delivers a *compound sentiment score* that combines all the values and gives the relative score between -1 and 1.

If the score is greater than 0.05 the sentiment is considered to be positive, the more positive, the closer it is to 1. It is opposite for negative sentiment, being between -0.05 and -1. If the score is between -0.05 and 0.05, the sentiment is considered to be neutral.

```python
def get_tweet_sentiment(self, polarity):
    '''
    Utility function to classify sentiment of passed tweet
    using Vader's sentiment method
    '''
    polarity = float(polarity)

    if polarity >= 0.05:
        return 'positive'
    elif polarity <= -0.05:
        return 'negative'
    else:
        return 'neutral'
```

2.2.4.2.5   News Based Analysis

As mentioned earlier, News Sentiment was that useful in helping us predict stock trends simply because articles are not written nearly regularly enough to deliver the kind of data that would be essential to predict the trend.

Instead, we decided that to leverage our work in News Sentiment to inform the platform user about the current headlines (in the past 7 days) and the Overall Sentiment. The news articles come from a platform called Seeking Alpha, which provides articles and insights into the industry. Since this indicator is not purely made to make an investment decision, we made sure to disclose that this was mainly for informative purpose.

We classified the articles by Positive, Negative and Neutral; and displayed the Positive and Negative articles allowing the user to click to read further based on their interest.
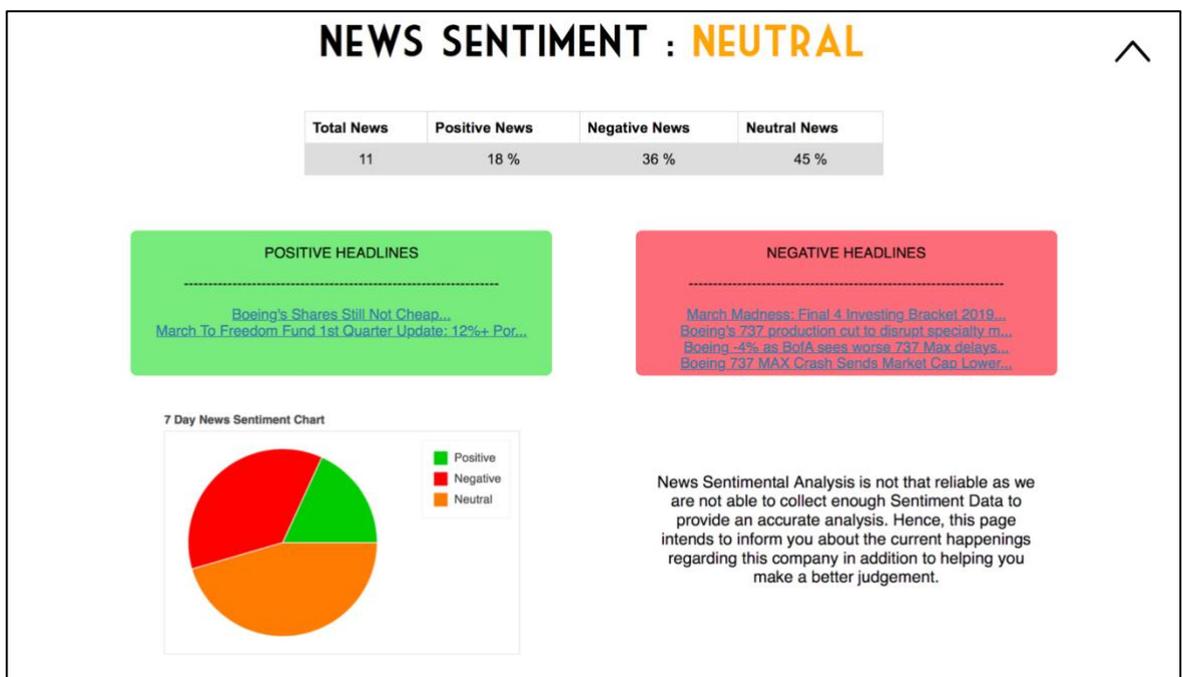
UI Screenshot of the News Sentiment Analysis

It is important to note that the financial news is only for easy reference to the user, if they wish to know about current affairs. As the analysis is done only on the headlines, there may be some scope of error, therefore we chose not to create any predictions based on the sentiment gotten from news articles.

2.2.4.2.6 **Twitter Based Analysis**

Description

Twitter is an American online news and social networking service on which users post and interact with messages known as "tweets". People use it to share content, insights, or even experiences.

Initial Hypothesis Formulae

There are two different ways we came up with to calculate the correlation between Twitter sentiment and stock price.

1. Calculate average sentiment per day for comparison with stock price of the next day.
2. Calculate the difference of the sentiments of the current and previous day and compare it to the stock price of the next day.

$$Suggestion_{today} = (\overline{Sentiment_{yesterday}} - \overline{Sentiment_{today}})$$

Formula for Strategy #2

When we previously collected Historical Tweets, we gathered a large number of tweets per day, after cleaning them and pre-processing them. We gathered the sentiment for each tweet. Later, we calculated the average sentiment of each day.

Buy and Sell signals are then generated every time the Suggestion is above a certain threshold. If today's suggestion would be under a certain threshold, we would buy, if it was over, we would sell, else just hold.

After testing both the techniques, we calculated the correlations shown in the table below.

Stock	Correlation 1	Correlation 2
IBM	-0.0319	-0.0593
BA	0.0057	-0.0156
MSFT	0.0343	0.0016
GS	-0.0411	0.0389
CSCO	0.0187	0.0326
CAT	-0.0155	-0.0151
CVX	-0.0164	-0.032
DOW	-0.0451	0.0538
XOM	0.0402	-0.0045
HD	-0.0035	0.0017
INTC	0.0082	-0.0093
JPM	-0.0135	-0.02055
MCD	0.0383	0.0099
MRK	-0.0427	0.01524
PFE	0.0053	0.0518
PG	0.0272	-0.0014
TRV	0.0302	-0.0304
UNH	0.0427	-0.0047
WBA	-0.0156	0.0246
AVE	0.001342105263	0.001962631579

Correlation Table
Correlation 1: Sentiment Difference vs Stock Price
Correlation 2: Average Sentiment vs Stock Price

Even though the number of stocks with higher correlation score is greater in Correlation 1, the total average correlation is greater for Correlation 2. Therefore, we decided to make Average sentiment of the day our primary mode of analysis.

Backtesting

To test the performance of the strategy, we used a similar technique as the one used for Technical Analysis. However, we custom wrote the backtesting code, as no library exists for this purpose. This was done using Tweets since 2015. The initial investment was assumed to be USD 100,000. We calculated the sum of returns for each of the 30 stocks in the index. Since Sentiment is human generated, and a person tweeting for Apple or Nike is more likely to just tweeting about the newest product they purchased than a person tweeting about Pfizer. Hence, we decided to not average them out.

Excluding Certain Stocks

While scraping Twitter for tweets on the companies listed on the Dow Jones, we encountered some challenges with getting quality datasets for some specific companies. Many consumer facing companies had immense amount of customer feedback on Social Media which was not very relevant to the stock price and would skew the analysis and make the system extremely slow.

Excluded Stock	Reason
COCA COLA	- Irrelevant Tweets
NIKE	- Extremely Large Dataset
APPLE	- Slows Down Program
WALMART	
VERIZON	
AMERICAN EXPRESS	Less than 10K tweets, not enough to make qualitative analysis
UNITED TECH	
JNJ	
3M	Confused with '3 Million'
VISA	Confused with 'Travel Visa'

Parameter Selection

There are two parameters for our implementation of the Sentiment Analysis. There is the Positive Threshold at which we Sell, and the Negative Threshold at which we buy. The daily mean of sentiment of all stocks used is 0.0682 and the standard deviation is 0.1675. Thus, we backtest for each stock using the Positive Bound and Negative Bound with in the range of [-0.18, 0.18] with an increment of 0.02 as the range is extremely large. We quickly realized that the results with these parameters are not working well in the lows for the Negative Bound and in the highs for the Positive Bound. So, in order to save time and achieve better results, we modified our Negative Bound parameters to a range of [-0.22, 0.2] with an increment of 0.04. Similarly, for the Positive Bound, we used a range of [0.1, 0.5] with an increment of 0.04. We will then find the middle point of the top 20% parameter combinations and use the respective mid-point as parameter for the respective stock.

User Focused Sentiment Superposition

Since this platform is to teach people how to make better informed decisions while investing, we decided to make it more intuitive, and we superposed the Sentiment next to the actual Stock Value. This allows us to show the user how much the Sentiment is able to predict the movement. The green line is the actual Stock Value over time and the Blue line is the sentiment. The multiplier used is the value of the stock on the present day. We can observe that both trends tend to follow each other.
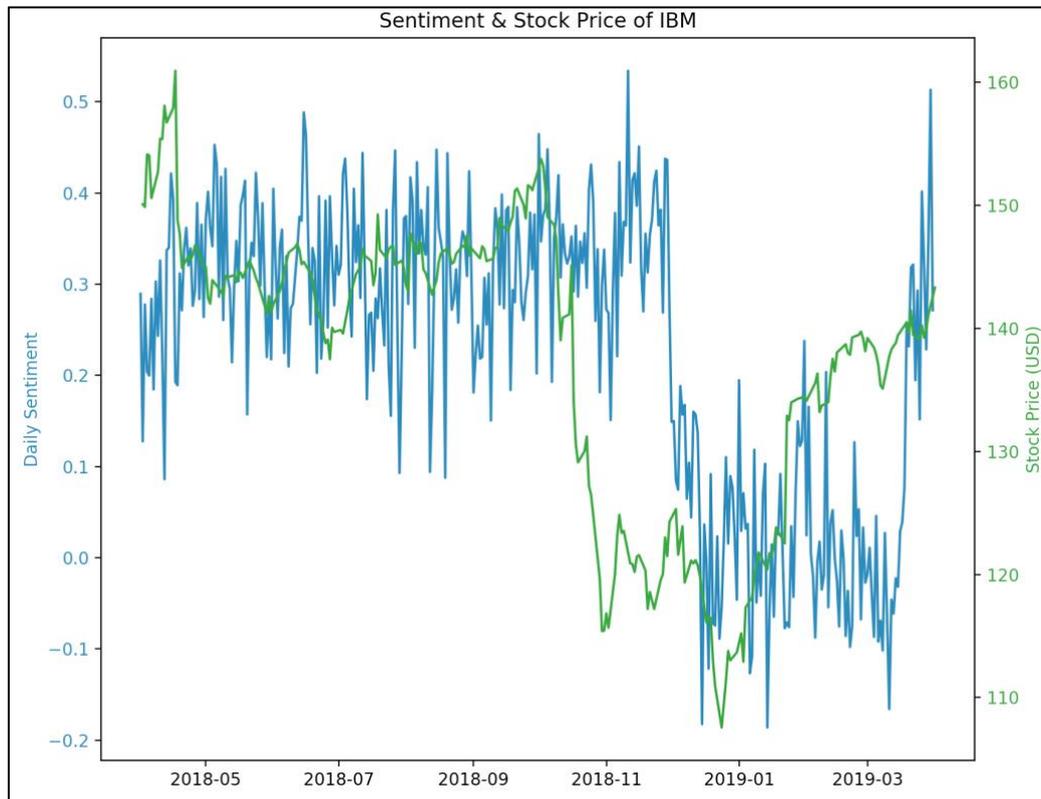
Representation of Twitter Sentiment and Stock Value for IBM
Between 04/2018 and 04/2019

2.2.5 Comparative Analysis

After analyzing each technical indicator on its own and finalizing on the par, we wanted to compare the performance of all four indicators for each of the 30 stocks in the Dow Jones Index. For each stock, we found out the returns received by employing the four technical indicators. We then ranked the indicators according to the return they provided.

Following are some of the examples that show the effectiveness of techniques:

Indicator Effectiveness for AAPL

Indicator Effectiveness for INTC

Indicator Effectiveness for WBA

Indicator Effectiveness for BA

Indicator Effectiveness for MSFT

Indicator Effectiveness for JPM
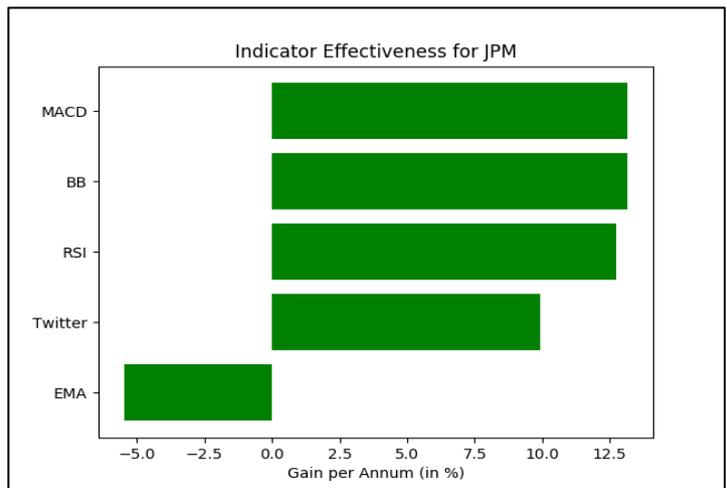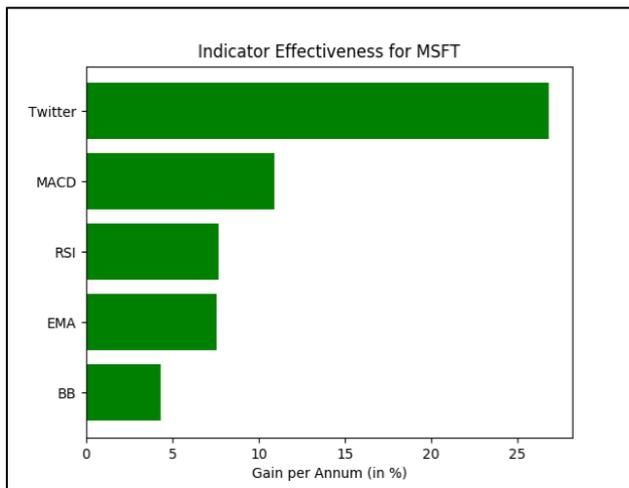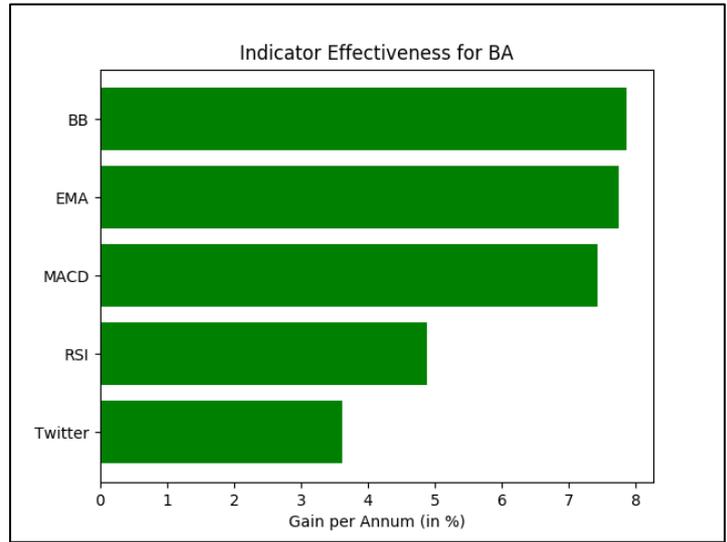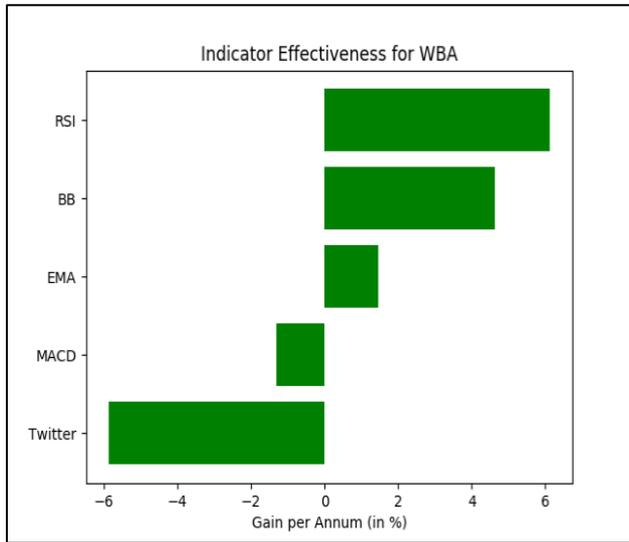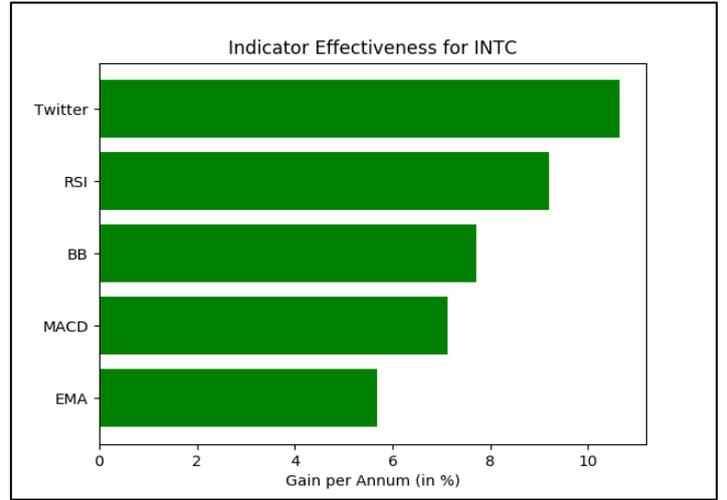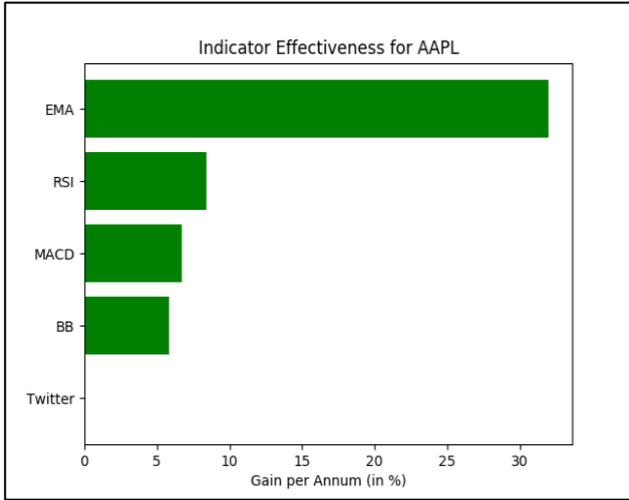
Thus, when the user wants to trade based on the suggestions from our web application, he or she will be able to see which indicator works best for the stock they are holding. This information will help them make more informed decisions with information tailored to their stock portfolio.

2.3 Testing

2.3.1 Technical Analysis

We backtested the technical indicators to find out the performance over the last 13 years. We did this using the Backtrader library, as described above. The final returns from the backtesting give us an indication of how good the technique is for trading stocks. This method also allowed us to choose the best parameter for each indicator and to compare the indicators with each other in terms of effectiveness.

2.3.2 Sentiment Analysis

For News Sentiment, since we decided that will just be an add-on and will not represent a core part of the analysis, we have not backtested that. However, we are displaying the results of the sentiment transparently for the user to have full access to.

Next, we backtested the Twitter Sentiment to find out the performance since 2015. We did this by writing a custom code based on the one used for technical analysis, as described above. The final returns from the backtesting give us an indication of how good the technique is for the trading stocks. This method also allowed us to choose the best parameters for each stock.

2.3.3 Web Application

We tested continuously with different people, requesting feedback on the UI, and the UX. This included people from Western, Indian, and Asian ethnicities between the age of 16-54. This allowed us to make some good improvements and to learn from some mistakes.

2.3.4 Testing Approach

Testing will be done as an ongoing process to get rid of errors. In order to build a good product, unit testing will be the first step taken and will continue throughout the project as a means of making sure that each component functions independently.

Once each component works independently, the team will proceed to integration testing to ensure that components together do not render any errors. However, it is common that integrating components may cause errors in other areas, so regression testing is a must to ensure that no other component is spoilt or corrupted in the process.

2.3.5 User Testing

We performed user testing with a sample of users of different level of expertise with regards to trading and the stock market. Our results allowed us to make our platform better

2.4 Evaluation

Our objective was to create an interactive and easy-to-use application that helps beginner level traders gain the most important information to make sound investment decisions. We aimed to do this by

1. Providing predictions made by different indicators within Technical Analysis
2. Performing Sentiment Analysis on financial news and tweets related to the stocks
3. Comparing these techniques with each other and informing the user about the effectiveness of each prediction technique over a period of time
4. Encompassing all the above features in an aesthetically pleasing and intuitive user interface

Our research shows that most of the online websites or software that provides stock predictions are paid services (e.g. Bloomberg Terminal). This makes proliferation of important information very limited and only available to those who can invest in such technologies. Many other free websites are more focused on educating the masses on how they can use these indicators but does not provide any predictions on their own. Our research also could not pinpoint any popular free apps that compare different techniques.

Keeping in line with our first objective, we have successfully been able to recommend through our web application whether the user should buy or sell a certain stock based on technical analysis indicators - EMA, MACD, Bollinger Bands, and RSI. This helps beginner investors skip many strenuous steps in accessing important information required to analyze their investments.

The second objective which was to perform Sentiment Analysis through Natural Language Processing on news articles and tweets from Twitter was completed through the use of relevant APIs and Sentiment Analysis libraries. We were able to test our algorithm with 5 years' worth of data to see what kind of returns the user would be able to make through this method. This is also helpful for the user to gain a real-world

understanding of how the company and stock is performing in the public domain. We realized that Twitter is a platform that despite the high number of spam can be successfully used to generate revenue.

Our third objective has been successfully implemented as well since we provide information to the user about which prediction technique is the most effective for the particular stock they are holding. This feature was executed through backtesting all prediction techniques for each stock and finding the best performer.

The fourth objective, which was to encompass all features in an aesthetic and intuitive user interface, was accomplished by keeping the look of the website minimalistic and attractive. When we surveyed potential users, we received positive feedback about the look and feel of our web application.

Switching to a web application has also enabled our software to be used by a whole array of devices running different operating systems as long as they have an internet browser. This gives our software the advantage to not be limited in terms of features or users. An android application cannot be accessed on a computer or an iOS device, neither can it use UI components that non-Android devices have. Therefore, our software can now be used by multiple users, which affirms our core objective of information outreach to as many users as possible.

Our user testing survey showed us that 85% of our users were more confident to invest in stocks after using Bull & Bear. Most of these users had no previous experience in investing.

3 Discussion

We succeeded in accomplishing all the objectives we had set for ourselves for the final year project. However, there were some changes made to the objectives as we progressed with the project. First of all, we had initially planned to implement Technical Analysis, Machine Learning as well as Sentiment Analysis on stock data. We soon realized that Machine Learning was out of scope due to time constraints. Secondly, in our second objective we had planned to calculate the accuracy of each technical indicator and compare them with each other. However, we realized that it is hard to define and calculate the 'accuracy' of the indicators, since most of them just signal if the price is about to increase or decrease, and not how much the price will change. Thus, we found it more correct to find the 'effectiveness' of each indicator for a particular stock by looking at the total returns gained from using the particular indicator. Lastly, we expanded the third objective of the project. Initially, we had only planned to conduct sentiment analysis on news articles. Later, we found that tweets are a good source of information to analyze market sentiment thus we decided to use both news articles and tweets to conduct sentiment analysis.

There are some caveats associated with technical analysis that the reader should be aware of. Technical analysis never guarantees if the price is bound to go up or down; it is a mere indication of a possible trend. Thus, technical analysis strategies can never be 100% accurate. Also, many a times the signals generated tend to have a lag; and thus, by the time a clear signal is shown the price action could already be over. Also, the parameters suggested for each technical indicator, as well as the comparative analysis of the indicators are based on results from backtesting over 13 years. However, past performance does not guarantee future returns.

Sentiment Analysis is a very interesting new space to explore with regards to stock predictions and investments. Currently the analysis is done based on the aggregate score of the words involved in the input. Some current limitations of this technique are that they work in a non-contextual manner. They take the words of the input at face value and work on a set of rules. It is not yet possible for the sentiment analyzer to detect sarcasm or understand the personal opinions of individuals. It is quite effective for non-ideological topics and products, which have an objective description.

One important problem that has not been discussed in this report is regarding unreliability of texts and data from social media. Fake news is very prevalent and fake news and fake tweets could easily be used to skew the analysis and alter the prediction results. In the future, we hope social media websites like Twitter and Facebook can crack down on fake news proliferation, so that our dataset reliability increases.

4 Conclusion

This project helped us enhance our financial knowledge. We learned the basics of technical and fundamental analysis, even though we did not get to use all of it. We were able to build on to the technical indicators that have been used for decades to trade. We did so by testing the different parameters and leveraging 13 years of past data. Additionally, all this was done in python, a language we did not have much experience with when starting out with our FYP. Our experiments in this aspect have indicated that generally used indicators with the standard parameters are not the optimal ways to predict stock price movement and trade. Parameters for each indicator can be customized for each stock to get the best returns. Additionally, even the relative effectiveness of different indicators will change depending on the particular stock. This leaves room for further customization of technical analysis for each stock or index to ensure the gains are maximized.

Sentiment Analysis gave us some interesting results, and as the algorithms and libraries get more advanced, it could be a very common way of predicting stocks in the near future.

5 APPENDIX

APPENDIX A - Distribution of Work

Task	Ishan	Mohibul	Nimisha	Tanmay
Conduct Research	●	●	●	●
Stock Data Retrieval	●			
Cache System Implementation	●			
Implement RSI indicator		●	○	
Test RSI indicator		●	○	
Implement MACD indicator		●	○	
Test MACD indicator		●	○	
Implement EMA		○	●	
Test EMA		○	●	
Implement Bollinger Bands		○	●	
Test Bollinger Bands		○	●	
Back test and refine technical indicators		●	●	
Online News Retrieval	●	●		
Implement Sentiment Analysis	●			○
Test Sentiment Analysis Accuracy	●			○
Design the UI for Web Application	○	○	●	○
Integrate Technical and Sentimental Algorithm within the Web App	●	○		○
Program the Front-end	●	○	○	○
User Testing	●	●	●	●
Deploy Web App to Heroku	●			
Cross validate prediction results on past data	○	○	○	●
Write the Final Report	○	○	○	●
Create and Prepare for the Presentation	○	○	●	○
Design the Project Poster	●	○	○	○
Comparative analysis implementation	●	○	○	○

● *Leader* ○ *Assistant*

APPENDIX B – Gannt Chart

Task	2018		2019			
	Summer	Fall	Jan	Feb	March	April
Conduct Research	●	●	●	●		
Stock Data Retrieval		●				
Online News Retrieval		●				
Technical Analysis						
Implement Bollinger Bands		●				
Test Bollinger Bands		●	●			
Implement EMA indicator		●				
Test EMA indicator			●			
Implement RSI			●			
Test RSI			●			
Implement MACD				●		
Test MACD				●		
Implement OBV				●		
Test OBV				●		
Back test and refine technical indicators					●	
Sentiment Analysis						
Implement Sentiment Analysis			●	●		
Test Sentiment Analysis Accuracy				●		
Web App						
Design the UI for Web Application			●	●		
Cache System Implementation						
Program the Front-end				●		
Integrate Technical and Sentimental Algorithm within the Web App					●	
User Testing						●
Cross validate prediction results on past data				●		
Deploy Web App to Heroku					●	
Miscellaneous						
Write the Final Report					●	
Comparative analysis implementation					●	
Create and Prepare for the Presentation						●
Design the Project Poster						●

APPENDIX C – Software and Hardware Requirements

Hardware

Operating System: Unix
Development Computer: MacBook Pro
Minimum Display Resolution: 1024 * 768 with 16-bit color
Server PC: PC with 1TB hard drive
Processor: 2.7 GHz Intel Core i5
Memory: 8 GB

Software

- Python was used for the algorithm coding
- HTML, CSS and Javascript were used for the web development
- Flask was used to link the python and web development frameworks
- Yahoo Finance APIs were used to acquire the necessary stock data.
- pyEX was used to fetch news articles related to the stocks.
- Twitter API was used to collect tweets from Twitter
- VaderSentiment was be used for sentiment analysis.
- Github and Google Drive were used to manage code and documents
- Google Chrome was used as the default browser for the web application

APPENDIX D - References

[1] S. Carter, "Younger Americans Aren't Investing in the Stock Market- Researchers Think This is Why" 16 May 2018, https://www.cnbc.com/2018/05/16/gallup-why-younger-americans-arent-investing-in-the-stock-market.html.

[2] A. Atkins, M. Niranjan and E. Gerding, "Financial News Predicts Stock Market Volatility Better than Close Price" June 2018, https://www.sciencedirect.com/science/article/pii/S240591881730048X

[3] M. H. McCormick, "The Effectiveness of Youth Financial Education: A Review of the Literature" Journal of Financial Counseling and Planning, vol. 20, no. 1, 2009.1, 2009, https://ssrn.com/abstract=2225339

[4] K. Kaufman, "The 7 Best Investing Apps For 2018", 15 December 2017; https://www.forbes.com/sites/karlkaufman/2017/12/15/7-best-investing-apps-2018/#713b66b7329f.

[5] X. Qian, "Financial Series Prediction: Comparison Between Precision of Time Series Models and Machine Learning Methods", 27 December 2017, https://arxiv.org/pdf/1706.00948.pdf.

[6] Unknown, "Technical Analysis v. Fundamental Analysis" Unknown Date, http://www.financial-spread-betting.com/course/technical-analysis-vs-fundamenntal-analysis.html

[7] M. Nikola, "Equity Forecast: Predicting Long Term Stock Price Movement Using Machine Learning", 2019; https://arxiv.org/pdf/1603.00751.pdf.

[8] V. Rajput and S. Bobde, "Stock Market Forecasting Techniques : Literature Survey" 6 June 2016, https://www.ijcsmc.com/docs/papers/June2016/V5I6201685.pdf.

[9] E. Rosenberg, "The 5 Best Investment Apps to Download in 2018" 6 August 2018, https://www.thebalance.com/best-investment-apps-4154203.

[10] Real Business. (2019). The five tweets that sent share prices crashing - Real Business. [online] Available at: https://realbusiness.co.uk/five-tweets-sent-share-prices-crashing

APPENDIX E – User Testing Survey

Bull and Bear User Testing

We want you to explore around the application, and then answer these questions

* Required

What is your age?

Your answer

What is your nationality?

Your answer

Have you ever invested in stocks?

○ Yes

○ No

What would you say your level of expertise is in the stock market?

	1	2	3	4	5	
No Experience	○	○	○	○	○	Fairly Experienced

How would you rate the Design of Bull and Bear? *

	1	2	3	4	5	6	7	8	9	10	
Worst	○	○	○	○	○	○	○	○	○	○	Best

How easy was the application to use?

	1	2	3	4	5	6	7	8	9	10	
Very Difficult	○	○	○	○	○	○	○	○	○	○	Very Easy

How confident do you feel with using Bull and Bear to make investment decisions?

	1	2	3	4	5	6	7	8	9	10	
Very Confident	○	○	○	○	○	○	○	○	○	○	Not Confident at all

Did you increase your understanding of stock market concepts by using Bull and Bear?

○ Yes

○ No

SUBMIT

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service

Google Forms

Results:
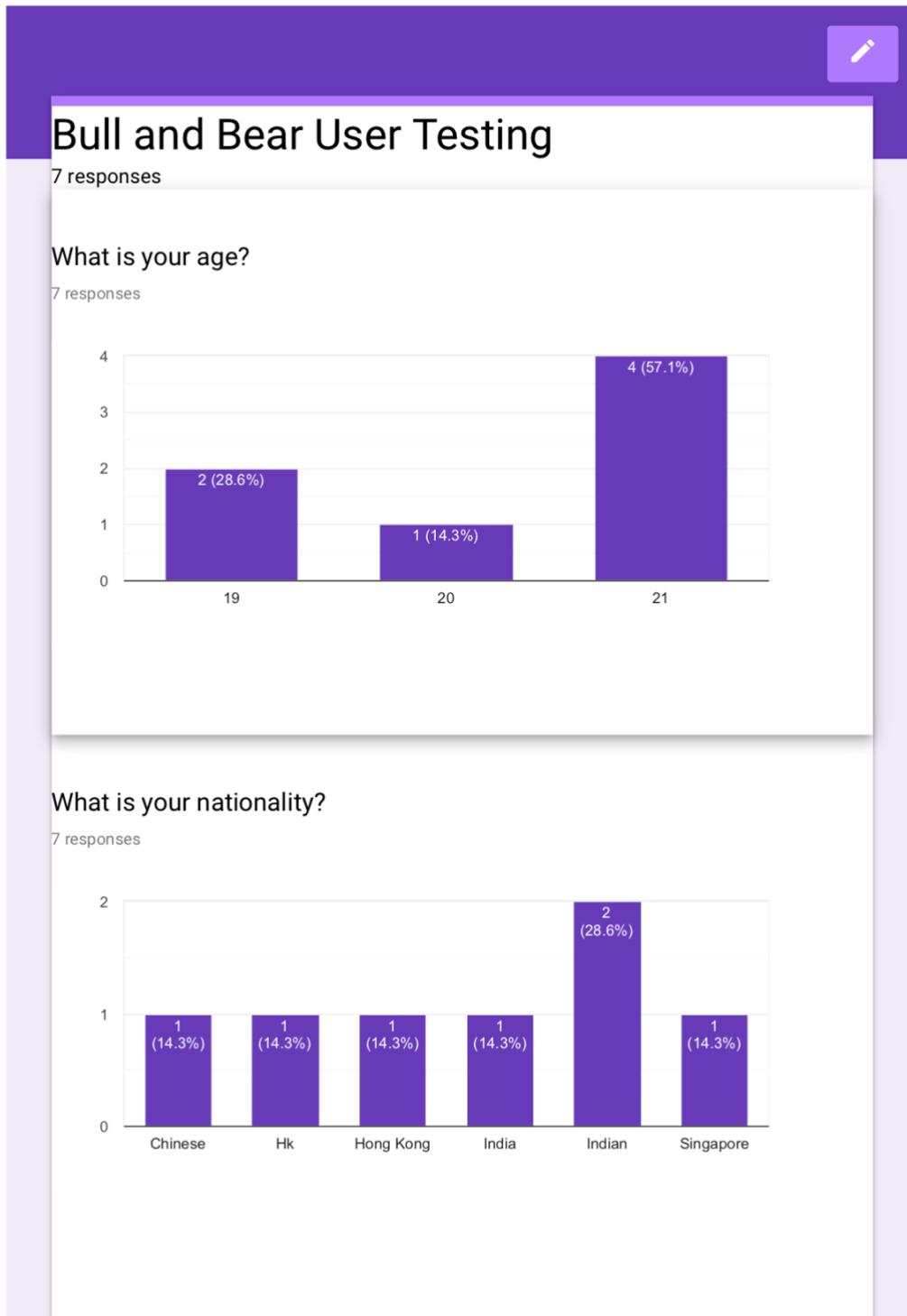
Have you ever invested in stocks?

7 responses

- Yes
- No

42.9%

57.1%
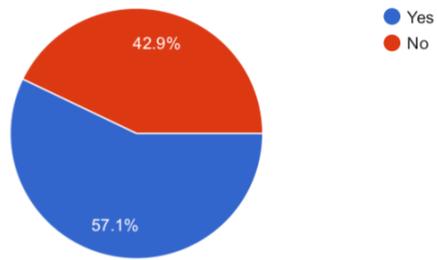
What would you say your level of expertise is in the stock market?

7 responses
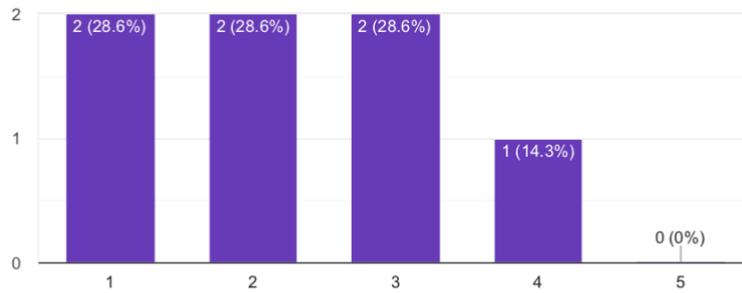
How would you rate the Design of Bull and Bear?

7 responses

How easy was the application to use?

7 responses

How confident do you feel with using Bull and Bear to make investment decisions?

7 responses

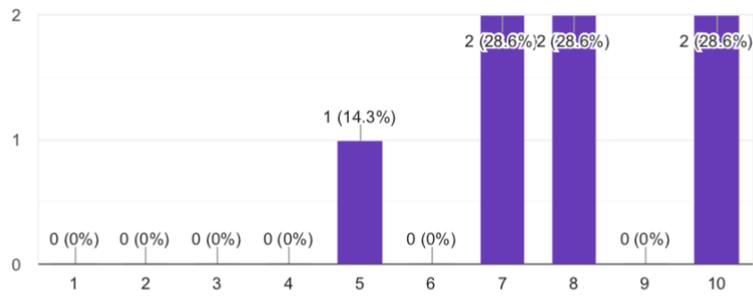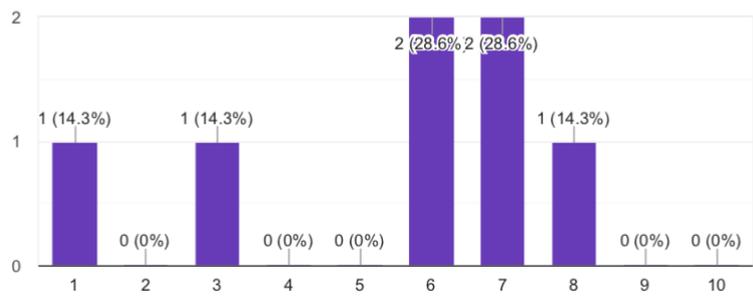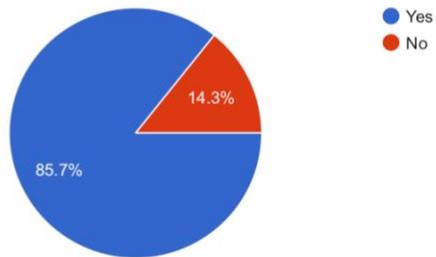Did you increase your understanding of stock market concepts by using Bull and Bear?

7 responses

APPENDIX F – Meeting Minutes

Minutes of the 1ˢᵗ Project Meeting

Meeting Information

Location:	Jockey Club Hall
Date:	4th September, 2018
Time:	6:00 pm
Attendees:	Ishan Jain, Mohib Hasan, Nimisha Goyal, Tanmay Goel

Discussion

1. Two possible ideas: – i) showing stock predictions on the application or ii) showing alt coins' predictions on the app
2. Goal of the project is to make an application for beginner level traders to get the predictions of stocks/ alt coins using different techniques and also obtain the performance of these predictions to help them make better investment decisions
3. There is scope in the area of cryptocurrencies since there are not many tools that exist to make price predictions. But the challenges include the fact that this area is very volatile.

Conclusion

1. Next meeting to be held on 9ᵗʰ September at 6 pm at Jockey Club Hall.
2. Decision to be taken in the next meeting – stocks OR alt coins?
3. Agenda for next meeting:
   a. Choose final topic – stocks OR alt coins?
   b. Discuss the content for proposal report
   c. Divide the proposal report within the team
4. Planned timeline:
   a. By 9ᵗʰ September - finalize topic + discussion + divide parts of the report
   b. By 12ᵗʰ September – meet to discuss and finalize report
   c. By 13ᵗʰ September – Meet Professor to get feedback
   d. By 17ᵗʰ September – Consolidate final report and send to CT for feedback

Action Items till next meeting

1. Research on stocks and alternative coins, existing prediction techniques for both and existing tools for both.
2. Go through the Communications Tutor website, specifically the proposal report template.

Minutes of the 2ⁿᵈ Project Meeting

Meeting Information

Location: Jockey Club Hall

Date: 9ᵗʰ September, 2018

Time: 4:00 pm

Attendees: Ishan Jain, Mohib Hasan, Nimisha Goyal, Tanmay Goel

Discussion

1. We are making a platform for people who want to invest but don't have a lot of knowledge. The end goal is to possibly launch it
2. Focus on making an android application first, then we can think about iOS and web application
3. We can either create a new algorithm or use a combination of already existing algorithms and show the users how certain algorithm has been working well for a specific type of stock
4. Use NLP to analyze news and figure out level of negativity and positivity to predict whether the price will go up or down

Conclusion

1. We are sticking to just stocks (10-20). (Still need to decide which stocks). This can be more in depth and we can use 6-7 techniques for the analysis
2. Day trading is not possible through our platform
3. We will use fundamental analysis, technical analysis and news to make prediction for next few weeks to up to a year
4. Use modular approach and each one works on a separate module so we don't waste time doing the same thing
5. Meet the professor on Thursday (4:30 to 5:30) or Friday (10-12)

Action Items till next meeting

1. Research on the workflow and make suggestions for front end back end integration APIs - Ishan
2. Project plan, Gantt chart, required hardware - Nimisha
3. Literature study, Feasibility study - Mohib
4. Overview, Objectives, Design, Evaluation – Tanmay

Minutes of the 3rd Project Meeting

Meeting Information

Location:	LC-8
Date:	12th September, 2018
Time:	6 pm - 8 pm
Attendees:	Ishan Jain, Mohib Hasan, Nimisha Goyal, Tanmay Goel

Discussion

1. Clarified the purpose (aim) of the application: educational/ portfolio management.
2. Finalized stocks over cryptocurrency, since they are less volatile and there seems to be easier access to data.
3. Went through the Proposal draft, and criticized each other's parts.

Conclusion

1. Next meeting to be held on 13th September at 4:30pm with the supervising professor.
2. Agenda for next meeting:
    a. Discuss the content for proposal report with the professor
    b. Acquire any additional tips that could benefit this project
    c. Divide the tasks amongst teammates

Action Items till next meeting

1. Rewrite parts of the proposal to make it more formal
2. Re-structure some parts of the proposal
3. Go through the Communications Tutor website, specifically the proposal report template.

Minutes of the 4th Project Meeting

Meeting Information

Location: Professor Rossiter's Room

Date: 13th September, 2018

Time: 4:30 pm

Attendees: Ishan Jain, Mohib Hasan, Nimisha Goyal, Tanmay Goel, David Rossiter

Discussion

1. Review with supervisor regarding Proposal Report
2. Discussion regarding edits to be made to the Proposal Report
3. Discussion regarding scope of project

Conclusion

1. Next meeting to be held on 16th September at 9:00 pm at Jockey Club Hall.
2. Agenda for next meeting:
   a. Final internal review of Proposal Report
   b. Discuss timeline for rest of the fall semester
   c. Fix goals for fall semester

Action Items till next meeting

1. Review and revise Proposal Report for final submission

Minutes of the 5th Project Meeting

Meeting Information

Location: Jockey Club Hall

Date: 17th September, 2018

Time: 5:30 pm

Attendees: Ishan Jain, Mohib Hasan, Nimisha Goyal, Tanmay Goel

Discussion

1. Formatting of the references was fixed
2. Images were added for the UI of the app and in other sections
3. Discussion regarding scope of project

Conclusion

1. Need to narrow down scope after submitting the proposal
2. Next meeting to be held next weekend to narrow down scope and divide work
3. Explore the potential to involve a finance student in the project
4. Submit the FYP for President's Cup

Action Items till next meeting

1. Nimisha will compile the final proposal report and get feedback from the CT, and finally submit on 19th Sept
2. Conduct further research on the different features in the app which is needed to narrow the scope next meeting

Minutes of the 6th Project Meeting

Meeting Information

Location:	Learning Commons
Date:	2nd October, 2018
Time:	4:30 pm
Attendees:	Ishan Jain, Mohib Hasan, Nimisha Goyal, Tanmay Goel, Matthew Uy

Discussion

1. Discussion with a QFIN student on how to analyze prediction techniques
2. Discussion on what types of texts to perform sentiment analysis on
3. Brainstorming ideas for the final application
4. Further discussion on project timeline and submissions

Conclusion

1. We understood different techniques in depth and how they can be used
2. We decided to switch to sentiment analysis on social media tweets rather than news articles
3. Developed a timeline to finish our project
4. Submit the FYP for President's Cup

Action Items till next meeting

1. Split into groups and do more research on specific techniques
2. Start writing preliminary code

Minutes of the 7th Project Meeting

Meeting Information

Location:	Jockey Club Hall
Date:	10th October, 2018
Time:	8:30 pm
Attendees:	Ishan Jain, Nimisha Goyal, Tanmay Goel

Discussion

1. Discussion and action on LANG4030 FYP Progress Report
2. Splitting of parts and proofreading

Action Items till next meeting

1. Once the final proofreading, Tanmay will submit the report on behalf of the group on Turnitin

Minutes of the 8th Project Meeting

Meeting Information

Location: Learning Commons

Date: 31st October, 2018

Time: 4:30 pm

Attendees: Ishan Jain, Tanmay Goel, Mohib Hassan

Discussion

1. Correspondence with Professor Rossiter regarding Monthly Report submission
2. Review and recap by the two mini groups on their progress
3. Discussion regarding problems faced and brainstorming of potential solutions

Conclusion

1. Monthly reports will be signed by Professor in January
2. Ishan and Tanmay managed to successfully program the sentiment analysis algorithm which scans through a given text and outputs a value indicating it's positivity or negativity (sentimental value).
3. Mohib and Nimisha shortlisted two technical analysis techniques (Simple Moving Averages and Bollinger Bands). They further attempted to import data from global stocks database to test Moving Averages.

Action Items till next meeting

1. Tanmay will submit the October Monthly Report on the FYP Management System before the deadline
2. Ishan and Tanmay plan to tweak the sentiment analysis classifiers, and also testing the algorithms on different social media platforms.
3. Mohib and Nimisha plan to test more techniques and evaluate h ow technical analysis can be used to time exits and entries into the market by a user.
4. Finalise a regular meeting timeslot to discuss the other teams findings and insights and keep each other accountable
5. Try to read 'The Intelligent Trader' by Benjamin Graham to gain an in-depth understanding of financial investment theory.

Minutes of the 9ᵗʰ Project Meeting

Meeting Information

Location:	Jockey Club Hall
Date:	29ᵗʰ November, 2018
Time:	8:30 pm
Attendees:	Ishan Jain, Tanmay Goel, Mohib Hassan, Nimisha Goyal

Discussion

1. Review and recap of last month, getting back on track after midterm and projects
2. Discussion regarding removal of Machine Learning from the scope of our project
3. Demo of Simple Moving Averages and Bollinger Band testing on stock data
4. Demo on sentiment analysis on regular text

Conclusion

1. Machine learning as a stock prediction technique has been removed from our project scope
2. Progress has been slow, but we plan to speed it up during the winter

Action Items till next meeting

1. Nimisha will submit the November Monthly Report before the deadline
2. Shortlist the other technical analysis techniques we'll be using and code them out. We will then test the prediction accuracy of these techniques.

Minutes of the 10th Project Meeting

Meeting Information

Location: Jockey Club Hall

Date: 28th December, 2018

Time: 8:30 pm

Attendees: Ishan Jain, Tanmay Goel, Mohib Hassan, Nimisha Goyal (Skype)

Discussion

1. Recap and review on previous months work
2. Discussion on finalizing updates for the progress report and January Monthly update

Conclusion

1. It is important to start working on the application, and have something presentable over the winter

Action Items till next meeting

1. Nimisha and Mohib to implement 5 technical analysis techniques working
2. Ishan and Tanmay plan to tweak the sentiment analysis classifiers, and also testing the algorithms on different social media platforms.
3. Mohib and Nimisha plan to test more techniques and evaluate h ow technical analysis can be used to time exits and entries into the market by a user.
4. Finalise a regular meeting timeslot to discuss the other teams findings and insights and keep each other accountable
5. Try to read 'The Intelligent Trader' by Benjamin Graham to gain an in-depth understanding of financial investment theory.

Minutes of the 11ᵗʰ Project Meeting

Meeting Information

Location: Jockey Club Hall

Date: 1ˢᵗ February, 2019

Time: 8:30 pm

Attendees: Ishan Jain, Tanmay Goel (Skype), Mohib Hassan, Nimisha Goyal

Discussion

1. Need to switch to a web application since it's easier to manage a python backend for a web application
2. UI is a very important aspect of our project, so we need to start working on a draft and keep revising

Conclusion

1. Next meeting to be held on 8th February with Prof. Dave Rossiter
2. Meeting to be fixed with Prof. Ted Spaeth to go over the progress report
3. Regular biweekly meetings to be set up with Prof. Dave to continuously update him on our progress

Action Items till next meeting

1. Mohib and Nimisha to finalize and program the five techniques for stock prediction by 8th Feb
2. Tanmay to build the UI for the web application by 11th Feb
3. Ishan to build a basic web application interface by 8th Feb

Minutes of the 12th Project Meeting

Meeting Information

Location: Jockey Club Hall

Date: 20th February, 2019

Time: 6:30 pm

Attendees: Ishan Jain, Tanmay Goel (Skype), Mohib Hassan, Nimisha Goyal

Discussion

1. Feasibility and impact of doing sentiment analysis on news vs sentiment analysis on twitter feed
2. Milestones to be achieved by 17th of April before the submission of the final report
3. Design of the final web application to be minimalistic and attractive

Conclusion

1. Next meeting to be held on 25th Feb at 3 pm in HKUST, and at 12 pm on Wednesday with Professor Rossiter
2. First step is to do sentiment analysis on news articles, next step is to do sentiment analysis on twitter feed
3. Milestones:
    a. 28th Feb: Finalize website design
    b. 15th March: finish technical analysis, major website pages
    c. 31st March: Testing for technical analysis, comparative analysis, complete web application, sentiment analysis complete
    d. 17th April: report ready to submit

Action Items till next meeting
1. Finish complete code for 2 technical analysis techniques including the final recommendation to the user
2. Finish tentative design for the website
3. Finish sample code for sentiment analysis for news article(s)

6.13 Minutes of the 13ᵗʰ Project Meeting

Meeting Information

Location: Jockey Club Hall

Date: 26ᵗʰ February, 2019

Time: 10:30 pm

Attendees: Ishan Jain, Mohib Hasan, Nimisha Goyal, Tanmay Goel

Discussion

1. Progress:
    a. Sentiment analysis done on news and tweets using NASDAQ as keyword
    b. MACD, Bollinger Bands and EMA strategy coded
2. Need to redesign UI to make it more attractive for users

Conclusion

1. UI feedback should be obtained from potential users not just teammates
2. Good progress past week, need to maintain the same momentum for the coming weeks

Action Items till next meeting

1. Goals for next week:
    a. Finish the 5ᵗʰ technique – Volume and also back test all 5 techniques
    b. Finish designing the UI of the application
    c. Integrate everything done so far on the web application front end

Minutes of the 14th Project Meeting

Meeting Information

Location: HKUST Library

Date: 5th March, 2019

Time: 4:30 pm

Attendees: Ishan Jain, Mohib Hasan, Nimisha Goyal, Tanmay Goel

Discussion

1. Progress:
   a. UI for the web application done
   b. Sentiment analysis implemented and displays the analysis as a pie chart and a normal distribution
   c. Implemented front – end for the home page and the second page of the web application

Conclusion

Difficult to draw credible conclusions from sentiment analysis. It will act as a supporting source of information

Action Items till next meeting

1. Goals for next week:
   a. Finish back test code for all 5 strategies
   b. Code the sentiment part and include it in the website
   c. Implement a strategy based on sentiment analysis

Minutes of the 15ᵗʰ Project Meeting

Meeting Information

Location:	Jockey Club Hall
Date:	19ᵗʰ March, 2019
Time:	4:30 pm
Attendees:	Ishan Jain, Tanmay Goel, Mohib Hassan, Nimisha Goyal

Discussion

Progress:

1. All technical indicators coded (including RSI)
2. Front - end now displays the strategies based on technical and sentiment analysis

Conclusion

1. Next meeting to be held on 27th March with Prof. Dave Rossiter

Action Items till next meeting

1. Mohib and Nimisha to start backtesting the technical indicators
2. Front end to be finished by next meeting

Minutes of the 16th Project Meeting

Meeting Information

Location: HKUST Learning Commons

Date: 2nd April, 2019

Time: 4:30 pm

Attendees: Ishan Jain, Tanmay Goel, Mohib Hassan, Nimisha Goyal

Discussion

Progress:
1. Backtesting results from technical indicators derived
2. Bug fixes for the web application

Conclusion

1. Internal deadline for report - 9th April

Action Items till next meeting

1. Sentiment Analysis to be backtested if possible
2. Start working on the report
3. Finalize parameters for the technical indicators

6.17 Minutes of the 17th Project Meeting

Meeting Information

Location:	Engineering Commons
Date:	13th April, 2019
Time:	10:00 am
Attendees:	Ishan Jain, Tanmay Goel, Mohib Hassan, Nimisha Goyal

Discussion

1. Prof. Rossiter's feedback for report implemented
2. Minor improvements made to the web application

Conclusion

1. Report and web application ready to be submitted

Action Items till next meeting

1. Proofread and submit report by the deadline