**Overview**

Time Series forecasting has gained more recognition in recent times. With people understanding forecasting’s intrinsic financial values and availability of data becoming greater with time, the prominence of the topic is only going to increase.

The recent advancements in Machine Learning and the ability of it to predict complex relationships to predict future has been a hot topic in forecasting research. In this project, we aim to perform 2 tasks:

1)- Best Machine Learning Model for Stock Price Prediction

2)- Study if sentiment information extracted from various sources can help us predict the stock price(adjusted closing price) movement(Up or Down) better.

In the first part of project, we concluded following works:

1. Did Literature analysis of past work to get a deeper understanding for doing something novel.
2. Collected stock prizes data and tweets from twitter.
3. Performed sentiment attitude(positive, negative) and mood(joy, sadness etc) analysis on tweets using “**p-senti**”.
4. Checked/pre-processed stock prizes data to convert it into a stationary time series(if not).
5. Performed statistical tests like Granger-Causality to check if sentiment data helps leverage stock price movement.
6. Figured out technical indicators that can act as a feature to help predict stock prize movement.
7. Did multi-model multi-case analysis of the stock price movement.
8. Our cases include:
   1. Using technical Indicators only
   2. Using technical Indicators and Sentiment attitudes
   3. Using technical Indicators, Sentiment attitudes and sentiment moods
9. Our models included standard Machine Learning Algorithms like LSTM(Long Short Term Memory), SVM(Support Vector Machine), AdaBoostClassifier etc.
10. Studied the results.

A brief summary of all our work

1. **Literature Analysis**
2. **Collection of Data**

Stock Price Data Source: quandl

Sentiment Data Source- Twitter

(Our primary aim is to increase amount of data in future works to improve performance of Deep Learning Techniques, whose performance primararily rests on amount of data it is trained on.)

1. **Seniment analysis using p-senti**
2. **Checked/pre-processed stock prizes data to convert it into a stationary time series(if not).**
3. **Performed statistical tests like Granger-Causality to check if sentiment data helps leverage stock price movement.**
4. **Figured out technical indicators that can act as a feature to help predict stock prize movement.**
5. **Machine Learning Analysis of Sentiment behavior on our stocks:**

Our primary aim was to study effect of sentiment analysis on stock prize movement. So for comparison, we first build a baseline Machine Learning Model(technical indicators only). Then we include other sentiment attributes. For first, we only include sentiment attitudes. Later, we include sentiment moods as well to study its effects.

All the above study is carried with multiple Machine Learning Models with a 0.75-0.25 training-validation split. Multiple models helps us figure out the best model for such studies and the split helps us evaluate how our models shall perform with unseen data.

(Full results of our studies have not been included in this report as it needs further reasoning and tuning, many of which have been excluded due to bad performance and to keep our report crisp. Still, the results can be accessed from \_\_\_\_\_\_\_\_)

The companies being studied separately includes Apple(AAPL), Google(GOOG), Hewlett Packard(HPQ) and JP Morgan(JPM).

[In our results, “Basic Lstm” is vanilla LSTM while “LSTM” is stacked LSTM]

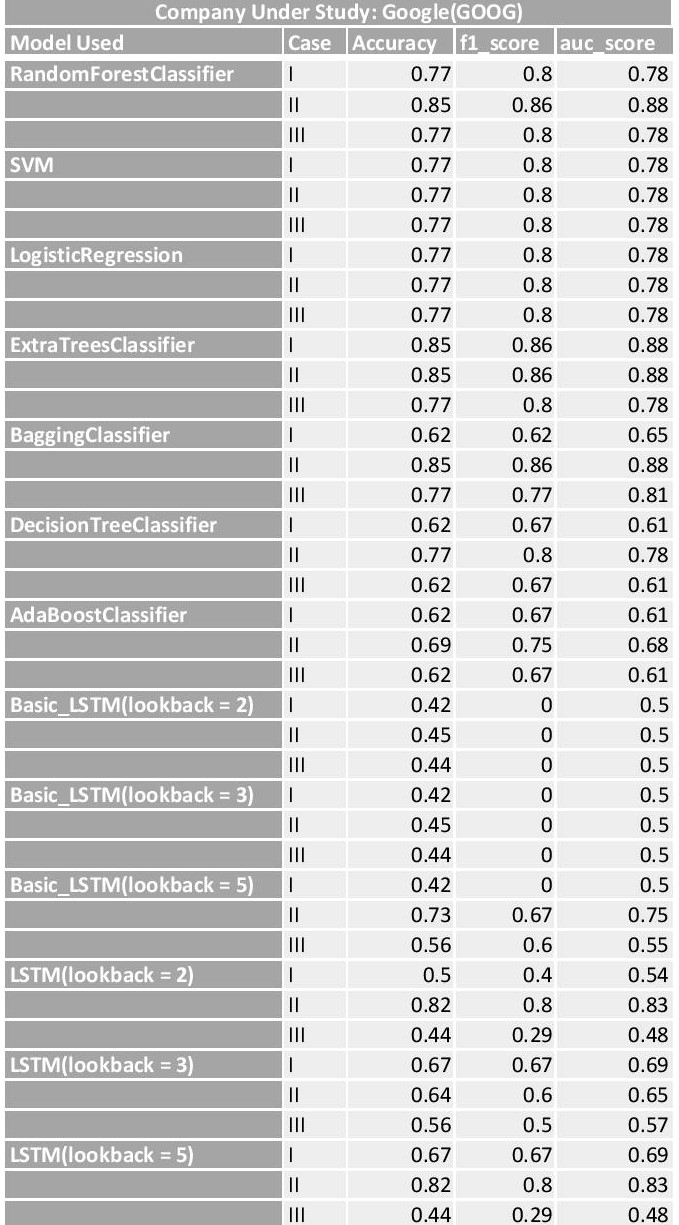
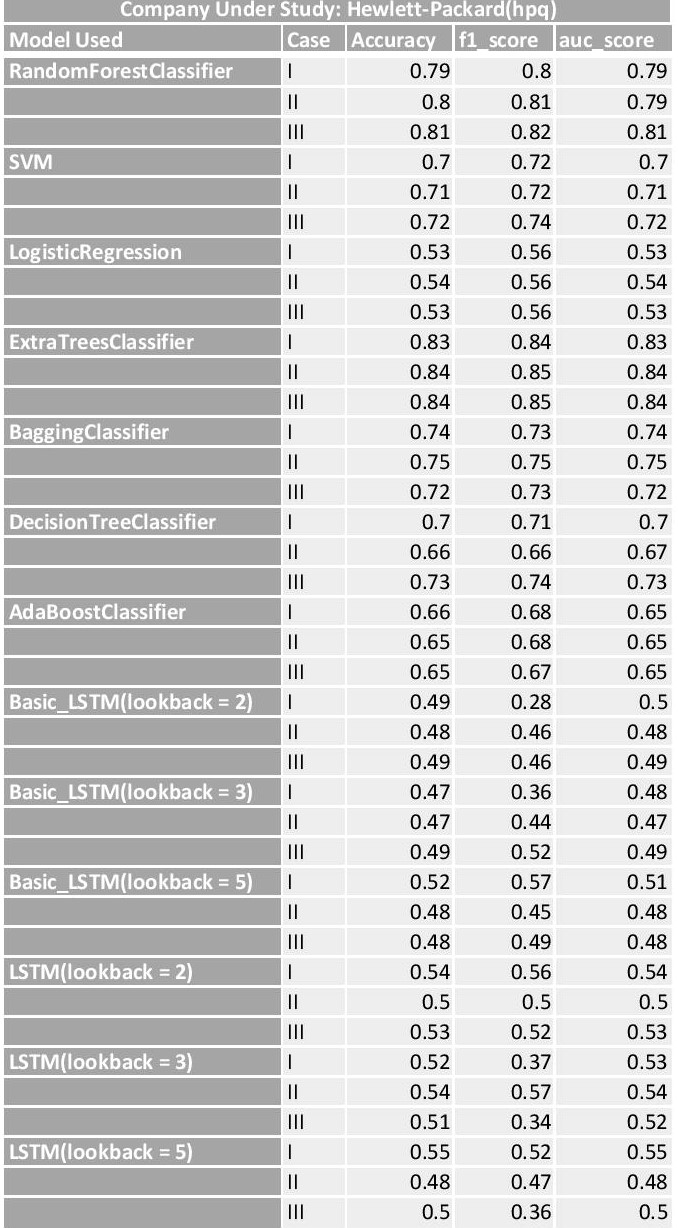
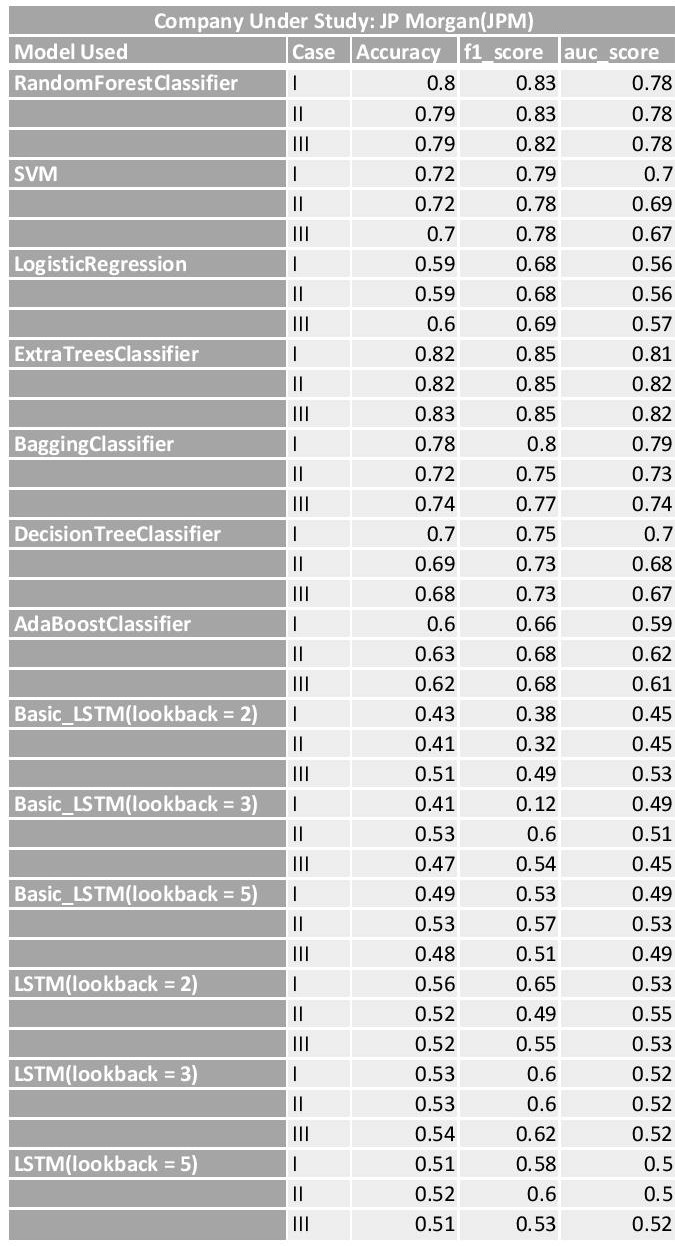
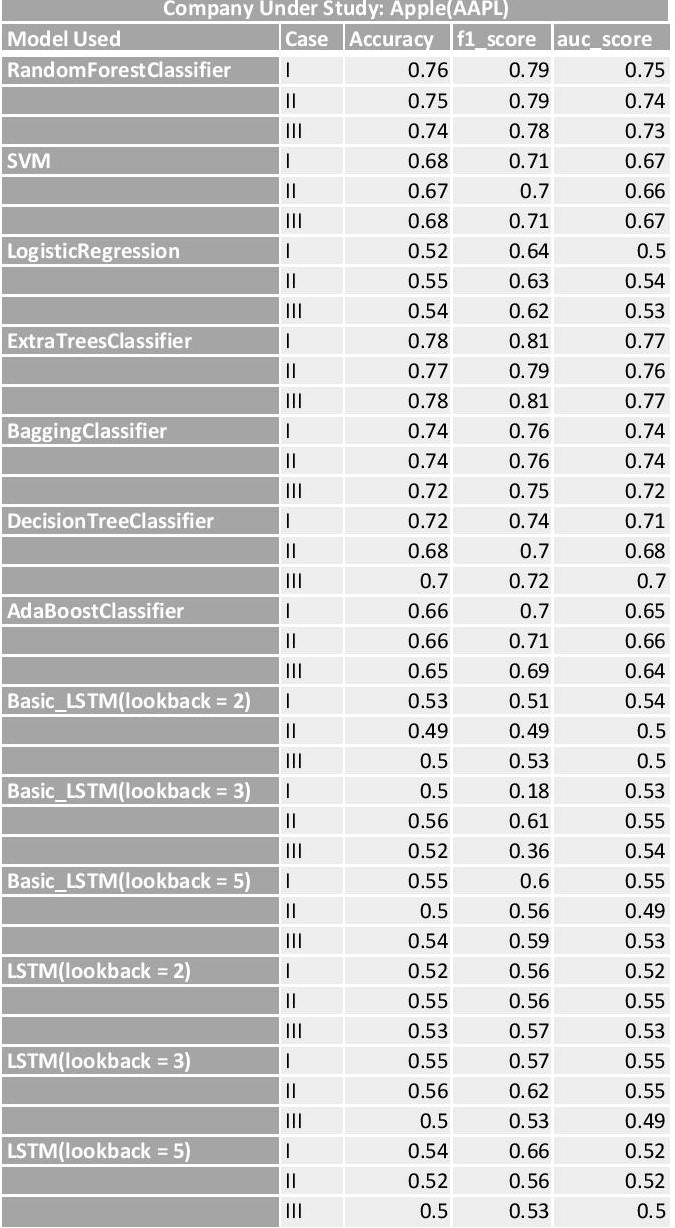


Table \_\_ :The summarized results of the performance of different Machine

Learning Models, when applied on different cases for different companies.

Case I: technical indicators only

Case II: technical indicators + sentiment attitudes

Case III: technical indicators + sentiment attitudes + sentiment moods

**Studying Results:**

1)- As expected, Deep Learning Techniques(LSTM) perform poorly due to presence of less data.

1. - As expected Ensemble methods outperform other techniques, due to their ability to model much more complex relationships then other techniques.
2. - The contribution of sentiment does not come out clear from the results as there are negative, positive and neutral correlations observed by using sentiment as a feature. While Google generally has positive or neutral improvement of results on including sentiment, JP Morgan has the opposite. What decides the movement shall be studied by us in the second part of the project.
3. - While stock prize movement is hailed by many experts as a very random process, the achievement of accuracy as high as 84 percent is a notable work.

**Probable Future Works:**

1. Collect more data to better our Deep Learning Results.
2. Use other sentiment analysis techniques with data from sources like redditt or Newspapers.
3. Try to predict what dictates the different contribution of sentiment in different companies.
4. Induce more technical indicators

References: