ISyE 6740 – Spring 2022
Project Proposal

Team Member Names:

1. Saurabh Sinha (GTID: 903747945),
2. Suneet Taparia (GTID: 903661244),
3. Yoages Kumar Mantri (GTID: 903660372)

Project Title:

Stock Classification based on predicted price using analytical models, leveraging Technical and Fundamental factors.

Problem Statement:

Stock Market share price prediction is an age-old problem. Over the years many algorithms and software have been built to solve this problem. There are numerous factors, including and not limited to a variety of subjective/sentiment factors, which can be used to predict stock value. Key challenge is that stock market data has variety, volume, velocity and is time series based. As such this problem poses a lot of analytical potential.

The oldest and most well-known model of stock returns is the Capital Asset Pricing Model (CAPM)[1]. Factors[6] that drive stock returns and have stood the test of time are:

- **Size**: Smaller firms tend to have higher returns on average as compared to larger firms.
- **Value**: Inexpensive stocks tend to outperform expensive ones. It was documented by Fama and Fench in 1993.
- **Momentum**: Stocks that have performed well over the past years continue to perform well.
- **Profitability**: Stocks with robust operating performance tend to outperform those with weak performance.
- **Risk effect (volatility)**: Low beta assets tend to outperform high beta assets.

Another prevalent theory around Stock market price is the Efficient Market Hypothesis (EMH)[2]. This hypothesis states that the market is extremely efficient in reflecting individual information about individual stocks and about the market itself. There are three versions of Efficient Market Hypothesis:

a) **Weak Form**: Future prices cannot be predicted by analyzing historical prices.
b) **Semi-strong Form**: Prices adjust rapidly to new public information.
c) **Strong Form**: Prices reflect all information, public and private.

Due to technological advances, the information gap has reduced to ashes, and weak form is almost non-existent. Most markets have become Strong to Semi-strong. As per EMH, stock performance is thus impossible to predict as future price changes represent random departures from previous prices as information arrives randomly and prices adjust quickly. This makes predicting stock price a complex problem, as we expect a lot of noise and a considerable number of possible factors.

We want to work on this project with an aim to classify stocks predicted price movement into 2 categories.

a) **Up** – After a given time period if the price of a share is greater than or equal to its current price.
b) **Down** – If the price of a share is less than its current price after a given time period.

For this classification, we will be doing 2 types of analysis on the share price historical data.

a) **Technical analysis**: This type of analysis focuses on changes in **price**, **volume** and **related statistics**, with a forward-looking nature through the inferences gathered with technical indicators, developed through heuristics or mathematical calculations. Here the EMH hypothesis plays a vital role. As such, the scope of this analysis will be limited to a short time period (1 – 2 weeks)[3]

b) **Fundamental analysis**, on the other hand, evaluates a stock's intrinsic value using publicly available information. It uses factors based on the overall economy in relation to industry performance and a company's financial factors such as **earnings**, **profit margin**, **assets**, **liabilities** etc. These financial factors will become the variables from which our models classify the stocks. Fundamental analysis can be used for strategic investments[4], since the factors used in this themselves evolve slowly compared to technical factors analysis.

We are attempting to solve this problem to make investment decisions easier. We would use models discussed in ISYE-6740 to find out which combination works best for a given problem.

(Optional) Data Source

As this is a well-known and worked-upon problem, there are a lot of data sources available. We are planning to use data for Indian and US Markets. Data needed for both Technical and Fundamental analysis can be downloaded from various credible sources, some of which are mentioned below:

1) Kaggle
2) Yahoo Finance
3) NSEPY and other libraries in Python
4) Web scraping from financial websites like moneycontrol.com

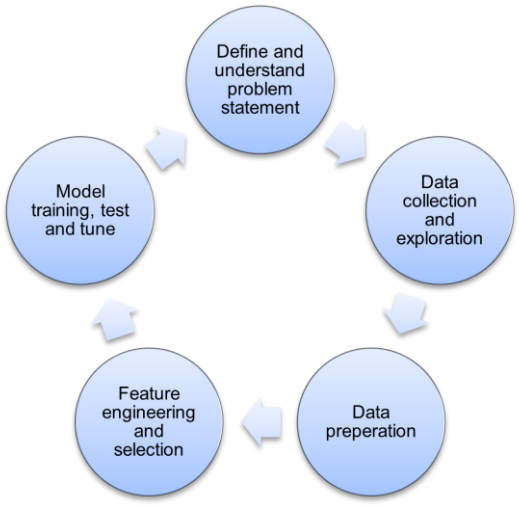
Methodology

The following are the proposed "initial" features that will be used for solving the problem:

	FEATURE 1	FEATURE 2	FEATURE 3	FEATURE 4	FEATURE 5	FEATURE 6
TECHNICAL	Previous Close	Open	High	Low	Close	Volume
FUNDAMENTAL	Debt to Equity	Beta	EPS	Sector	Dividend yield	Return on Equity

We would use classification algorithms (Logistic Regression, KNN, Linear / Kernel SVM etc.) to identify and bucket stocks into Up and Down categories.
Next, we will use predictive models [5] [7] [8] or a subset (Lasso, Ridge, Multiple Linear Regression, ARIMA, SARIMA, CNN, LSTN etc.) to predict stock price.

We would follow this below mention lifecycle for Data Science project:



1. Define and understand problem statement – The team would continuously work on refining initial problem statement and develop expertise in the subject matter.
2. Data collection and exploration – We would collect data from various credible sources (as mentioned above) and use visual and statistical exploration of the data.
3. Data preparation – Extracted data would go through standardization and imputation. For Time-series Significance tests, (ADF) T-Statistic test, stationary test would be done before data is made available to ML models for training and testing.
4. Feature engineering and selection – Based on all the features available in the dataset we would perform feature selection (PCA, Lasso, Elastic Net etc.) also define new features if needed.
5. Model training, test, and tune – Once we have the right set of features, we would train, test, and tune the ML models.

Evaluation and Final Results

Classification models to label the stocks between “Up” and “Down” categories will be evaluated using the following metrics - Confusion matrix, Accuracy, True-positive, True-negative, Precision and Recall.
Regression models for predicting stock price will be evaluated[5] using the following metrics - Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Average Percentage Error (MAPE), R-Squared and Adjusted R-Squared.

Based on the observed values for these metrics, we will find out the best model for classification and predict stock price direction.

References

1. Eugene F. Fama and Kenneth R. French, The Capital Asset Pricing Model: Theory and Evidence
2. Alexandra Gabriela Ţiţan, The Efficient Market Hypothesis: Review of Specialized Literature and Empirical Research
3. Credit Suisse, Technical Analysis - Explained
4. Silpa K S, Arya Mol J, Dr. A S Ambily, A study on fundamental analysis of selected IT companies listed at NSE
5. Marwa Sharaf, Ezz El-Din Hemdan, Ayman El-Sayed & Nirmeen A. El-Bahnasawy , "StockPred: a framework for stock Price prediction", 12th Feb 2021
6. Jennifer Bender, Remy Briand, Dimitris Melas, Raman Aylur Subramanian, "Foundations of Factor Investing", Dec 2013
7. Mehr Vijh, Deeksha Chandola, Vinay Anand Tikkiwal, Arun Kumar, Stock Closing Price Prediction using Machine Learning Techniques, International Conference on Computational Intelligence and Data Science (ICCIDS 2019)
8. Sreelekshmy Selvin; R Vinayakumar; E. A Gopalakrishnan; Vijay Krishna Menon; K. P. Soman, Stock price prediction using LSTM, RNN and CNN-sliding window model