DEBANJAN SAHA

19CS30014
AI60003: ASSIGNMENT-1 < **STAGE-2** >

PROBLEM STATEMENT

OVERVIEW

Stock market investments have proven to be one of the best ways to grow long-term wealth. Over several decades, the average stock market return is about 10% per year. However, it comes with its own set of risks and insecurities. Hence, it is very important to **analyze** the risk factors involved in any investment in the stock market and get a **detailed prediction or future analysis** of the stocks in which someone is interested to invest. **Visualization** of different parameters such as the **average price of various stocks, daily ups and downs, the correlation between different stocks, daily returns, long term investment returns** etc. using some user-friendly and easily interpretable diagrams using pre-existing algorithms and also making predictions of future behaviour of such stocks in a structured manner or a graphical format using several artificial intelligence algorithms is required. This not only enables us to get a report card on the stock in which we are investing based on the past data but also shows us the long-term effect of our investment by getting the predicted trend of those stocks in future.

SPECIFICATIONS

Given the overview of the problem that I am trying to solve. I would like to list down a few specifications of the problem and how I am breaking it down into different parts to solve it. The central idea of this problem statement revolves around the fact that stock market analysis enables investors to identify the intrinsic worth of security even before investing in it. This **forecasting problem** leads to the proposed solution which is analysis and prediction. To analyze and predict something, we need some data. Hence, I will highlight these parts here and also a specification of how to use that data to do an analysis and prediction are stated as follows:

SELECTION OF DATA

My choice of the dataset for this problem is the collection of data of the best-performing stocks in the software industry i.e stock prices of FAANG companies (an acronym for American tech companies Facebook, Amazon, Apple, Netflix and Google). The reason behind choosing these companies is due to the publicly available genuine repositories containing the historic stock price data of these companies. Also, considering the performance of the stocks of these companies over the years, we could get a really good analysis of the historic data and we will be able to predict the future values of the stock prices of these companies and see how they relate to each other or how the risk factors are found during investment in these types of stocks.

DATA-SOURCES

Various open-source APIs are available these days to extract the stock prices of a large set of companies. One of them is **YahooFinanceAPI**. Several developers have used these types of

APIs to or crawled the internet to create some datasets for the convenience of others. Some such data sources are given as follows:

1. https://www.kaggle.com/aayushmishra1512/faang-complete-stock-data : The dataset present in this repository consists of all historic data from the beginning of stock trading of the FAANG companies up to 2019/2020.

2. Kaggle: FAANG Historical Stock Price Data: This dataset lists the FAANG stocks data for the past 5 years (it was updated recently in 2021, hence it has the latest updates in the stock prices)

GOALS

Given the data source. the overall goal of my problem is to do a robust data analysis and prediction using AI algorithms and as sub-goals, I will try to achieve the following things :
1. Generating different data analysis figures such as 'Trend' etc to visualize the data
2. The change in the price of the stocks over time and analyzing the time series
3. The daily return of the stocks on average
4. The moving average of the various stocks
5. Correlation between different stocks
6. How can we attempt to predict future stock behaviour? (Predicting the stock opening/closing price of one/more companies using algorithms described below)
7. How much value do we put at risk by investing in a particular stock?
8. An attempt to answer the question of whether it is beneficial to invest in these stocks.

INPUT/OUTPUT

In computational terms, it is always a good idea to formulate the problem into a series of inputs and outputs. Here, the input to the problem is various **Opening/Closing Prices** or High/Low Stock Prices for each day retrieved from the datasets, which is fed to a few data analysis algorithms and AI-based prediction algorithms and the outputs will be several values and **plots of analyzed data**, trends and analysis of detailed historical data along with the **prediction values** and a plot corresponding to that and analysis of risk factors etc.

ALGORITHMS TO USE

Time-series analysis and Time-series forecasting are the two most important methods in stock price analysis and prediction. Time series analysis can be useful to see how a given asset, security, or economic variable changes over time. We can decompose(**Time-series decomposition**) the data into level, trend, seasonality, and noise components using **seasonal decompose library methods** to analyze the data well. **Augmented Dickey-Fuller test (ADF test)** is a common statistical test used to test whether a given Time series is stationary or not.

Artificial Neural Networks (ANN) are widely used for the prediction of stock prices and their movements. Hence, I would like to use the **Long Short-Term Memory (LSTM)** Model to predict stock prices. LSTMs are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price. The hidden states of this neural network act as storage/memory of the past data improving which are used to predict future values.