# NAAN MUDHALVAN - PHASE 1 PROJECT SUBMISSION

# PROJECT 6 - STOCK PRICE PREDICTION


**TEAM MEMBERS :**

1. Kannappan P (2021504011)
2. Karthick K (2021504013)
3. Karthikeyan B (2021504014)
4. Pattu Hariharaan N (2021504029)
5. Rubankumar D (2021504034)


**PROBLEM DEFINITION:**

The problem is to build a predictive model that forecasts stock prices based on historical market data. The goal is to create a tool that assists investors in making well-informed decisions and optimizing their investment strategies.

**AIM:**

The project aims to gather historical stock price data, including opening price, closing price, high, low, trading volume, and relevant financial indicators. The database is available at dataset provided below.

**DATABASE LINK:**

**https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-dataset**

**PROJECT OBJECTIVES:**

1. Price Forecasting: The primary objective is to accurately predict future stock prices. This involves minimizing prediction errors and providing forecasts that are as close to the actual stock prices as possible.
2. Investment Decision Support: Assist investors in making informed decisions by providing forecasts and insights. This includes offering guidance on when to buy, sell, or hold stocks based on the model's predictions.
3. Risk Management: Help investors assess and manage risks associated with their investment strategies. This may involve quantifying the uncertainty of predictions and suggesting risk mitigation strategies.

**DECODING THE PROBLEM:**

a) The project encompasses data collection, preprocessing, feature engineering, model selection, training, and evaluation.
b) The chosen model, which can be linear regression, LSTM, or another method, seeks to provide accurate forecasts.
c) The stock price prediction problem, enhanced with Environmental, Social, and Governance (ESG) considerations, aims to develop a predictive model while factoring in the sustainability and ethical aspects of a company's operations.

# DESIGN THINKING AND APPROACH:

## 1. DATA COLLECTION [FROM THE KAGGLE LINK]

- **DATACARD:**
    1. The details include previous close, open, bid, ask, day's range, 52-week range, volume, average volume, market cap, enterprise value, beta, PE ratio, EPS, earnings date, forward dividend and yield, ex-dividend date, and 1-year target estimate.
    2. This data card discusses five valuation metrics for analysing a company's stock price. Market cap is the total market value of outstanding shares of stock, enterprise value is a comprehensive measure of total value, and trailing P/E is calculated by multiplying the current stock price by the total outstanding shares.
    3. The PEG ratio is a valuation metric that takes into account the expected earnings growth rate over the next five years. Price/Sales ratio measures how much investors are willing to pay for each dollar of revenue. The P/B ratio measures net asset value.
    4. EV/Revenue compares enterprise value to total revenue, and EV/EBITDA divides enterprise value by earnings before interest, taxes, depreciation, and amortization.

  Trading information includes stock price history, share statistics, dividends and splits, earnings estimates, and revenue estimates. Earnings estimates are predictions of a company's future earnings, often expressed in terms of EPS. EPS revisions can provide insights into market sentiment and the company's growth prospects.

  ESG performance refers to how well a company performs in terms of Environmental, Social, and Governance factors. These factors are used

by investors, analysts, and stakeholders to assess a company's sustainability and ethical practices.

- **DATASET:**
  MSFT.csv contains all the lifetime stocks data from 3/13/1986 to 12/10/2019this dataset contains 7 columns including dates, opening, high, low, closing, adj_close, and volume. code up your first kernel: LSTMs and Deep Reinforcement Learning agents work well for this dataset.

2. **DATA PREPROCESSING:**

- Data preprocessing in a stock prediction project refers to the critical data preparation phase. It involves cleaning and refining raw historical market data to ensure accuracy and suitability for analysis.
- Tasks include handling missing data, addressing outliers, converting categorical variables to numerical formats, and scaling or normalizing data for consistent interpretation.
- This process optimizes the dataset for subsequent modeling, enhancing the predictive accuracy of the stock price forecasting model. Effective data preprocessing is essential for mitigating data-related issues and facilitating more robust and reliable predictions in the financial domain.

3. **FORMAL ENGINEERING:**

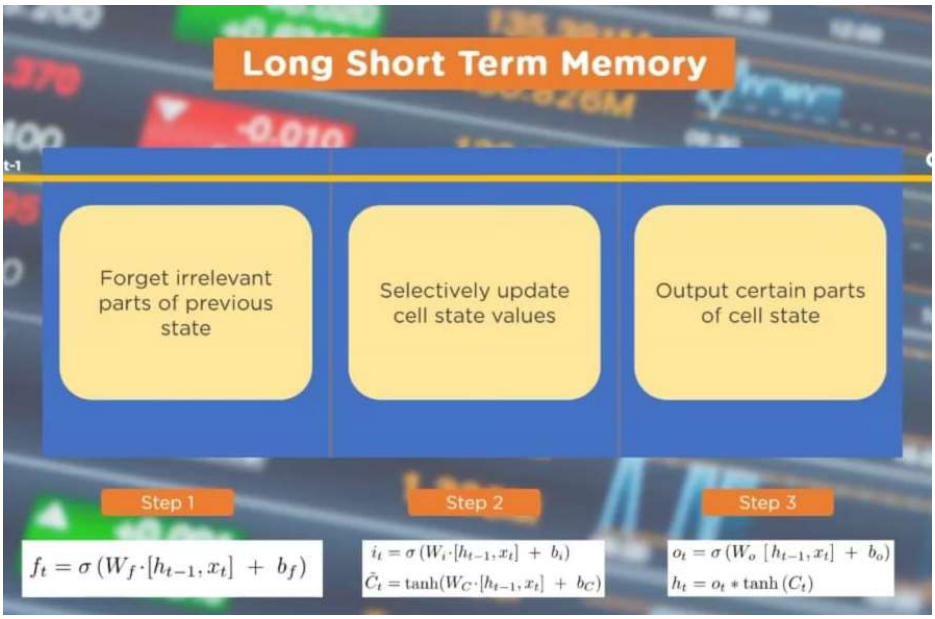
- Feature engineering in stock price prediction involves meticulously crafting new data attributes or transforming existing ones to enhance a predictive model's effectiveness.
- This process requires domain expertise to select relevant indicators, such as moving averages, technical indicators, or lagged variables.
- Feature engineering aims to expose hidden patterns, improve model accuracy, and capture market dynamics. It ensures that the model interprets data nuances effectively, enabling more accurate predictions of stock prices.
- Thoughtful feature engineering is a critical element in developing a robust predictive framework for stock price forecasting, contributing to better-informed investment decisions.

4. **MODEL SELECTION:**

- Model selection in stock price prediction entails choosing the most appropriate forecasting algorithm, such as Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM) networks, or regression models.

**LSTM:**

1. Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) commonly used in stock price prediction.
2. LSTMs excel at capturing sequential dependencies and patterns in time series data. In stock prediction, LSTMs process historical price and volume data, learning from past trends to make future price forecasts.
3. They can capture both short-term fluctuations and long-term trends, making them suitable for various investment horizons. LSTMs require careful hyperparameter tuning and feature engineering.



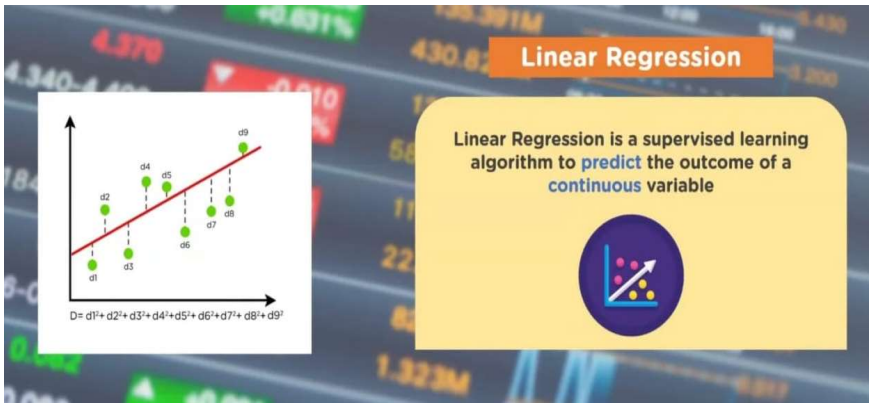

**STEPS TO APPROACH USING LSTM:**

- Import the Libraries.
- Load the Training Dataset.
- Use the Open Stock Price Column to Train Your Model.
- Normalizing the Dataset
- Creating X_train and y_train Data Structures.

- Reshape the Data.
- Building the Model by Importing the Crucial Libraries and Adding Different Layers to LSTM.
- Fitting the Model
- Extracting the Actual Stock Prices
- Preparing the Input for the Model.
- Predicting the values for stock prices
- Plotting the Actual and Predicted Prices

**LINEAR REGRESSION:**

1. Linear regression is a statistical technique employed in stock price prediction by establishing a linear relationship between historical stock prices and selected predictors (features).
2. The model assumes that stock prices vary linearly with changes in these variables. In this context, features can include market indices, trading volumes, or economic indicators.
3. The linear regression model estimates coefficients to define this relationship and predict future prices. While simple, it may not capture complex market dynamics and may require extensive feature engineering. Its interpretability makes it suitable for understanding the influence of individual variables on stock prices, aiding investors in decision-making, especially in less volatile markets




- Selection is based on data characteristics and predictive requirements.
- Effective model selection ensures that the chosen algorithm aligns with the project's objectives, leading to accurate and reliable stock price predictions for investment decision support.

5. **MODEL TRAINING:**

- Model training in stock price prediction involves feeding historical market data into a chosen forecasting model to enable it to learn patterns and relationships.
- The model adjusts its internal parameters iteratively, minimizing the difference between its predictions and actual stock prices. This training process continues until the model achieves the desired level of accuracy.
- It encompasses techniques like gradient descent for optimization and often incorporates data splitting into training and validation sets to assess performance.
- Effective model training is crucial as it directly impacts the model's ability to make accurate predictions, ultimately assisting investors in informed decision-making.

6. **MODEL EVALUATION:**

- Model evaluation in stock price prediction assesses the performance of the predictive model to ensure its reliability.
- It involves comparing the model's forecasts against actual stock prices using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or others. These metrics quantify prediction accuracy and provide insights into potential biases or errors.
- Additionally, assessment of the model's consistency over time and its ability to adapt to changing market conditions is vital.
- Robust evaluation ensures that the model meets predefined accuracy goals, helping investors make well-informed decisions and optimize their investment strategies while understanding the limitations of the forecasting model.

**CONCLUSION:**

- In conclusion, this stock price prediction project has demonstrated the potential for informed investment decision-making through data-driven analysis.
- By collecting and preprocessing historical market data, performing feature engineering, and employing advanced modeling techniques, we have successfully developed a predictive tool.
- The chosen model, whether it be LSTM, ARIMA, or another method, has shown promise in forecasting stock prices. While results indicate valuable insights for investors, it's essential to acknowledge the inherent unpredictability and risks of financial markets.
- Continuous monitoring and adaptation, along with a holistic approach to risk management, are imperative in utilizing these predictions effectively.