**IBM NAAN MUDALAVAN**

**APPLIED DATA SCIENCE**

**PROJECT 6:STOCK PRICE PREDICTION**

**Phase 1: Problem Definition and Design Thinking**

**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. This project involves data collection, data preprocessing, feature engineering, model selection, training, and evaluation.

As the problem definition of the project tells us that , the main motive of our model is to predict prices of a given stock and then predict its price for the next few days. As this will help the investors in understanding the market using appropriate data points and thus giving them a clear vision in making the right and well informed decisions.

**DESIGN THINKING:** This is the phase where in how the problem is going to be approached is discussed

**Data Collection:** In this step which is the primary step for any problem statement is the collection of appropriate data which will help us in making proper predictions based on the relevant data points.

Here , we can take the historical data of the stocks of whom we want to predict the price for and accordingly we will have to take in the right columns that help in the prediction process which would be date, open price, close price, volume, are some of the indicators which can help us in the predicting the price.

**Data Preprocessing:** This is a very crucial part of the project because the accuracy of the model is as good as the data on which it feeds to give us the desired outcome that means if we clean the dataset i.e, removing unwanted data or filling up the null/nan values and then keeping only those factors that actually help in predicting the price is what is to be done in this step.

**Feature Engineering:** There are many factors that affect the price of any stock so we need to use different parameters that help in enhancing the power of prediction of the price like Lag features,moving averages, relative strength index(RSI) ,calendar features,time based features, or more such ideas can be inco-operated.

**Model Selection:** Here we choose the most appropriate or suitable algorithm like linear regression , but then ARIMA or SARIMA which are time series model or even Long short term memory networks(LSTM) can be made use of for predicting the price of the stock.

**Model Training:** This is the point where in the model that we have created is checked against the predefined dataset (70:30 i.e,70 % data for processing/training and 30%data for testing).

**Evaluation:** here the model is evaluated and this part tells us the model's performance using appropriate time series forecasting metrics (e.g., Mean Absolute Error, Root Mean Squared Error).