**OBJECTIVE:** PREDICTING THE CLOSING PRICE OF NASDAQ INDEX USING MACHINE LEARNING

**INTRODUCTION –** Aim of this project is to forecast the closing price of the NASDAQ index through the implementation of advanced machine learning techniques, including linear regression, neural networks, random forest, and time series forecasting. In order to evaluate the effectiveness of each model, we will compare them using the Root Mean Squared Error (RMSE) metric. By utilizing these sophisticated techniques, we hope to gain valuable insights and accurate predictions for the closing price of the NASDAQ index.

Data set - yahoo finance https://finance.yahoo.com/quote/%5EIXIC/history?period1=34560000&period2=1639008000&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

**METHODOLOGY:**

* **Linear regression -** Linear regression is a machine learning technique used to model the relationship between a dependent variable and one or more independent variables by finding a best-fit line or hyperplane that minimizes the difference between the predicted values and actual values of the dependent variable.
* **Neural networks -** Neural networks are a type of machine learning algorithm inspired by the structure and function of the human brain. They consist of interconnected nodes or neurons arranged in layers, which process and transform input data to generate output. Neural networks are capable of learning complex patterns and relationships in the data, making them useful for tasks such as image and speech recognition, natural language processing, and decision-making.
* **Random forest** -Random forest is a machine learning technique that builds multiple decision trees and combines their predictions to improve the accuracy and reduce overfitting. It randomly selects a subset of features and data points for each tree, making it more robust and flexible for a wide range of classification and regression problems.
* **Time series** - Time series forecasting is a machine learning technique used to make predictions about future values of a dependent variable based on its historical values over time. In other words, it is a method of analyzing and forecasting trends and patterns in data that change over time. Time series forecasting models can be used to forecast any kind of data that changes over time, such as sales figures, stock prices, weather patterns, and more. These models take into account patterns and trends in the data, as well as any seasonality or cyclicality, and use statistical methods to make predictions about future values.

**FINDINGS:**

* **Linear regression**

1. R square value for linear regression is 0.98 which means that 98% of the variation of a dependent variable is explained by an independent variable.
2. RMSE value for linear regression model is 62.911.

Chart, scatter chart

Description automatically generated

* **Neural Network** 
  + RMSE value for neural network is 11612.
  + **Diagram

    Description automatically generated**
* **Time series forecasting**
  + RMSE value for time series forecasting model on this data is 338.8
* **Random forest** 
  + RMSE value for random forest model on this data is 101.8
  + Chart, scatter chart

    Description automatically generated

**RESULTS AND DISCUSSION:**

* Linear regression model has the lowest RMSE value, Therefore, it can be inferred that this model performs better than the other models in predicting the closing stock prices.
* The trained model can be utilized to forecast future closing stock prices. Given the inputs of Open, High, Low, and Volume, the model can estimate the expected value of the closing stock price with reasonable accuracy.
* The random forest model has the lowest RMSE value of 101.8 among the models tested, indicating that it may be a suitable alternative to the linear regression model for predicting the closing stock prices based on the input variables.
* The linear regression model assumes that the relationship between the input and output variables is linear, which may not be true for all cases. Additionally, the neural network model may suffer from overfitting, especially when the number of neurons is high. These limitations can affect the reliability and accuracy of the models.
* alternative models, such as Support Vector Machines (SVMs) or Gradient Boosting Machines (GBMs), can be explored for predicting the closing stock prices. Additionally, incorporating new data sources, such as news articles or social media sentiment analysis, can provide valuable insights into the stock market trends and help improve the accuracy of the models.