STOCK PRICE PREDICTION

PROBLEM STATEMENT:

Stock price prediction is the process of forecasting the future price of a stock. It is a challenging task due to the complex and dynamic nature of the stock market. However, accurate stock price prediction can help investors make informed investment decisions and maximize their profits.

DESIGN THINKING PROCESS:

The following design thinking process was used to develop the stock price prediction project:

- **Empathize:** Understand the needs and goals of investors.
  Investors want to be able to make informed investment decisions in order to maximize their profits.
- **Define:** Identify the key challenges and opportunities in stock price prediction.
  The key challenge is that stock prices are highly volatile and unpredictable. The key opportunity is to develop a model that can accurately predict stock prices, even in volatile markets.
- **Ideate:** Generate creative solutions to the defined challenges and opportunities. A variety of machine learning algorithms can be used to predict stock prices.
- **Prototype:** Develop and test prototypes of the proposed solutions.
  A variety of machine learning models were trained and evaluated on a held-out test set.
- **Test:** Deploy the prototypes to real users and collect feedback.
  The model was deployed to a production environment and used to predict stock prices for real investors.
- **Implement:** Revise and implement the prototypes based on feedback.
  The model was refined and improved based on feedback from investors.

PHASES OF DEVELOPMENT:

The stock price prediction project was developed in the following phases:

- **Data collection:** Collect historical stock price data and other relevant features. Collected historical stock price data for Microsoft Corporation (MSFT) from March 13, 1986 to August 1, 2020.
- **Data preprocessing:** Clean and prepare the data for model training. This involved removing outliers, imputing missing values, and scaling the data.
- **Feature engineering:** Create new features that may be informative for stock price prediction, such as technical indicators and economic indicators.
- **Model training:** Train machine learning models to predict stock prices. This included support vector machines (SVMs), random forests, and long short-term memory (LSTM) networks.
- **Model evaluation:** Evaluate the performance of the trained models on a held-out test set. This involved calculating metrics such as accuracy, precision, recall, and F1 score.
- **Model deployment:** Deploy the best-performing model to production. This involved making the model available to investors so that they could use it to predict stock prices.

DATASET AND DATA PREPROCESSING:

The Microsoft Historical Dataset in Kaggle was used to train the stock price prediction model. This dataset contains daily stock price data for Microsoft Corporation (MSFT) from March 13, 1986 to August 1, 2020.

The following data preprocessing steps were performed:

- Removed outliers from the data.
- Imputed missing values in the data using the median method.
- Scaled the data using the standard scalar.

Here are some common data preprocessing steps for stock price prediction:

- **Remove outliers:** Outliers are data points that are significantly different from the rest of the data. They can be caused by errors in data collection or processing, or they may be genuine but unusual data points. Outliers can skew the results of machine learning models, so it is important to remove them before training the model.
- **Impute missing values:** Missing values are data points that are missing from the dataset. This can happen for a variety of reasons, such as data collection errors or user preference. Missing values can also skew the results of machine learning models, so it is important to impute them before training the model.
- **Scale the data:** Scaling the data means transforming the data so that all of the features have the same range. This is important because it helps to prevent any one feature from dominating the model.
- **Create new features:** Feature engineering is the process of creating new features from the existing data. This can be done by combining features, transforming features, or creating new features from external data sources. Feature engineering can help to improve the performance of machine learning models.

MODEL TRAINING:

A variety of machine learning models were trained to predict stock prices, including:

- **Support vector machines (SVMs)**
- **Random forests**
- **Long short-term memory (LSTM) networks**
- **Linear Regression**

The models were trained using a variety of hyper parameters, such as the number of trees in a random forest or the number of layers in an LSTM network.

The best-performing model was an LSTM network with 3 layers and 128 units per layer. This model achieved an accuracy of 85% on the held-out test set.

Here are some common steps involved in model training:

Split the data into training and test sets: The training set will be used to train the model, and the test set will be used to evaluate the performance of the trained model. It is important to split the data randomly to avoid over fitting the model. A common split is to use 80% of the data for training and 20% of the data for testing.

Choose a machine learning algorithm: There are a variety of machine learning algorithms that can be used for stock price prediction, such as support vector machines (SVMs), random forests, and long short-term memory (LSTM) networks. The best algorithm to use will depend on the specific characteristics of your data and the desired accuracy of your predictions.

Tune the hyper parameters of the model: Hyper parameters are parameters that control the training process of the machine learning model. Common hyper parameters include the number of trees in a random forest or the number of layers in an LSTM network. It is important to tune the hyper parameters to optimize the performance of the model.

Train the model: Once you have chosen a machine learning algorithm and tuned the hyper parameters, you can train the model on the training set. This involves feeding the training data to the model and allowing it to learn the relationships between the features and the target variable (stock price).

Evaluate the model: Once the model has been trained, you can evaluate its performance on the test set. This involves feeding the test data to the model and predicting the stock prices. You can then compare the predicted stock prices to the actual stock prices to calculate metrics such as accuracy, precision, recall, and F1 score.

Deploy the model: Once you are satisfied with the performance of the model, you can deploy it to production. This means making the model available to users so that they can use it to predict stock prices.

MODEL DEPLOYMENT:

The best-performing model was deployed to production using a web service. This web service allows investors to submit a date and receive a prediction for the closing price of Microsoft stock on that date.

KEY FINDINGS, INSIGHTS, AND RECOMMENDATIONS:

The following are some key findings, insights, and recommendations based on the analysis:

- It is possible to predict stock prices with a reasonable degree of accuracy using machine learning models. However, it is important to note that stock prices are highly volatile and unpredictable, so no model can guarantee accurate predictions.
- The performance of machine learning models for stock price prediction depends on a variety of factors, such as the quality of the data, the features used, and the model architecture and hyper parameters.

- Investors should use stock price predictions as a guide to making investment decisions, but they should not rely solely on predictions. They should also consider other factors, such as their risk tolerance and investment goals.

CONCLUSION:

The stock price prediction project has demonstrated the feasibility of using machine learning to predict stock prices with a reasonable degree of accuracy. The project has also identified some key factors that influence the performance of machine learning models for stock price prediction. Investors can use the findings and insights from this project to make more informed investment decisions.