Project 6: **Stock Price Prediction**

Problem Definition

The problem at hand is to develop a predictive model capable of forecasting stock prices based on historical market data. The primary objective is to create a valuable tool that empowers investors to make well-informed decisions and optimize their investment strategies. This multifaceted project encompasses several critical steps, including data collection, data preprocessing, feature engineering, model selection, training, and evaluation.

Design Thinking

1**. Data Collection**

Objective: Gather historical stock market data, which includes essential features such as date, open price, close price, volume, and any other pertinent indicators.

Method:

Utilize the Kaggle dataset provided (Dataset Link: Microsoft Lifetime Stocks Dataset).

Explore the dataset's contents and structure to gain insights into the available data.

2. **Data Preprocessing**

Objective: Prepare the collected data for model training by cleaning it, handling missing values, and converting categorical features into numerical representations.

Method:

Check for missing values and decide on an appropriate strategy for imputation (e.g., forward-fill, backward-fill, mean imputation).

Convert categorical features into numerical format through techniques like one-hot encoding or label encoding.

Normalize or standardize numerical features if necessary to ensure that they are on the same scale.

3. **Feature Engineering**

Objective: Enhance the predictive power of the model by creating additional relevant features, such as moving averages, technical indicators, and lagged variables.

Method:

Calculate and add moving averages with various time windows (e.g., 7-day, 30-day) to capture trends.

Incorporate technical indicators (e.g., Relative Strength Index, Moving Average Convergence Divergence) to provide more insight into market behavior.

Create lag features to account for the influence of past prices on future ones.

4. **Model Selection**

Objective: Choose the most suitable algorithms for time series forecasting, such as ARIMA (AutoRegressive Integrated Moving Average) or LSTM (Long Short-Term Memory), to predict stock prices.

Method:

Analyze the nature of the dataset (e.g., trend, seasonality) to determine whether ARIMA or LSTM is more appropriate.

Experiment with different models and hyperparameters to identify the best-performing one.

5**. Model Training**

Objective: Train the selected model using the preprocessed and engineered data.

Method:

Split the dataset into training and validation sets to monitor the model's performance.

Utilize time series cross-validation techniques, like rolling-window cross-validation, to account for temporal dependencies.

6. **Evaluation**

Objective: Assess the model's performance using appropriate time series forecasting metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or others.

Method:

Evaluate the model on the validation set using chosen metrics.

Consider visualizations like time series plots and prediction vs. actual plots to gain a deeper understanding of model performance.

By adhering to these design thinking principles, we will embark on a systematic journey to build a robust stock price prediction model. This will enable us to provide investors with a valuable tool for making informed decisions and optimizing their investment strategies. Throughout the project, we will maintain flexibility to adapt to challenges and iterate on our approach as needed to achieve the best results.

Conclusion :

In conclusion, the initial phase of the Stock Price Prediction project has laid the groundwork for a comprehensive and systematic approach to solving the problem of forecasting stock prices based on historical market data.