STOCK PREDICTION WITH APPLIED DATA SCIENCE

PHASE 3: DEVELOPMENT PART-01

To Do:

- ❖ Collect historical stock price data from a reliable source.
- ❖ Clean the data, handle missing values, and create features.
- ❖ Visualise data using line charts, histograms, and explore trends.
- ❖ Choose and develop a predictive model.
- ❖ Evaluate the model's performance using appropriate metrics.
- ❖ Plan for future enhancements, including feature expansion and advanced modelling techniques.

Overview:

- ❖ Focusing on the initial steps of the data science pipeline, including collecting historical stock price data, preprocessing it to prepare for modelling, and conducting a basic exploratory data analysis to understand the data's characteristics.

Steps to be followed:

❖ Data Collection: - Retrieve data for the specific stock or index you intend to predict.

❖ Data Preprocessing: - Handle missing data, if any, using techniques like interpolation or imputation.

❖ Exploratory Data Analysis (EDA): - Visualize the historical stock price data to uncover trends and patterns.

❖ Model Selection: - Choose an appropriate modeling approach based on the problem (e.g., linear regression, time-series forecasting, machine learning models like Random Forest or LSTM).

❖ Data Splitting: - Divide the data into training and testing datasets to assess model performance accurately.

❖ Model Development: - Develop and train the initial prediction model using the training dataset.

❖ Model Evaluation: - Assess the model's performance using appropriate evaluation metrics (e.g., RMSE, MAE, R-squared).

❖ Fine-Tuning and Improvement: - Fine tune the model by adjusting hyper parameters and making necessary improvements.

❖ Future work: - Consider incorporating additional data sources or features to enhance predictive accuracy.

SAMPLE CODE FOR THE ABOVE PROCESS:(PYTHON).

```python
import pandas as pd

import yfinance as yf

# Define the stock symbol and date range

stock_symbol = "AAPL"

start_date = "2010-01-01"

end_date = "2020-12-31"

# Download historical stock price data

data = yf.download(stock_symbol, start=start_date, end=end_date)

# Drop rows with missing data
```

```python
data = data.dropna()
# Set the date as the index
data = data.set_index('Date')
# Calculate 50-day and 200-day moving averages
data['50MA'] = data['Close'].rolling(window=50).mean()
data['200MA'] = data['Close'].rolling(window=200).mean()
import matplotlib.pyplot as plt
# Visualize the closing price over time
plt.figure(figsize=(12, 6))
plt.plot(data.index, data['Close'], label='Close Price', color='blue')
plt.title('Stock Price Over Time')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.show()
# Create histograms, scatter plots, or any other relevant visualizations
for EDA.
```