**Project Proposal: Stock Market Price Prediction Using Machine Learning Models**

**Objective**

The goal of this project is to predict future stock prices using historical stock data from sources like Yahoo Finance. We will apply and compare three different machine learning techniques: **Linear Regression**, **ARIMA**, and **LSTM (Long Short-Term Memory)**. By comparing the results of these models, we aim to determine which method provides the most accurate predictions and is best suited for stock market time series data.

**Motivation**

Stock price prediction is a challenging task due to the non-linear, volatile, and time-dependent nature of financial data. Accurate predictions can be valuable for investors and financial analysts. This project explores multiple approaches to forecast stock prices and aims to understand the strengths and weaknesses of different machine learning models for time series prediction.

**Methodology**

1. **Data Collection:**
   * We will use the yfinance Python library to collect historical stock data (e.g., for Apple Inc., symbol: AAPL). The dataset will include daily stock prices (open, close, high, low, and volume) for a period of 5+ years.
2. **Data Preprocessing:**
   * We will focus on predicting the **closing price** of the stock.
   * The data will be preprocessed by scaling it for the models that require normalization (e.g., LSTM), and it will be split into training and testing sets (80% training, 20% testing).
3. **Models:**
   * **Linear Regression**: A simple baseline model that will use past 30 days of stock data to predict the next day’s closing price. It is easy to implement and serves as a good comparison point.
   * **ARIMA (AutoRegressive Integrated Moving Average)**: A classical statistical model that is often used for time series forecasting. ARIMA will capture trends and seasonality in the univariate stock price data.
   * **LSTM (Long Short-Term Memory)**: A deep learning model well-suited for time series data, especially where sequential dependencies are important. LSTMs can capture more complex patterns in stock price movements.
4. **Evaluation Metrics:**
   * Each model will be evaluated using metrics like **Mean Squared Error (MSE)** and **Root Mean Squared Error (RMSE)** on the test set.
   * We will also visually compare the predicted prices against the actual stock prices to assess each model’s performance.
5. **Tools and Technologies:**
   * **Data Collection**: yfinance
   * **Preprocessing & Analysis**: pandas, numpy, scikit-learn
   * **Modeling**:
     + **Linear Regression**: scikit-learn
     + **ARIMA**: statsmodels
     + **LSTM**: TensorFlow/Keras
   * **Visualization**: matplotlib

**Expected Outcome**

We expect to find that each model has distinct advantages and disadvantages for stock price prediction:

* **Linear Regression** will serve as a simple, interpretable baseline, but may struggle to capture complex patterns.
* **ARIMA** may perform well for short-term predictions where trends and seasonality are evident, but might not handle sudden price changes effectively.
* **LSTM** is likely to provide better results for longer-term predictions by capturing sequential patterns, though it may be more computationally expensive and require more fine-tuning.

**Conclusion**

The comparison of these models will offer insights into which methods are most effective for stock market prediction and under what conditions each model excels. This project will not only demonstrate the application of machine learning in financial data but also highlight the differences between classical statistical models and more advanced deep learning techniques.