### **Stock Price Prediction with LSTM Networks Final Assignment Instructions**

#### **1. Install Required Modules**

Ensure that all necessary Python libraries are installed. These libraries are essential for data fetching, preprocessing, model creation, and evaluation. You will need:

* **yfinance**: For downloading stock price data.
* **pandas**: For data manipulation and analysis.
* **numpy**: For numerical operations.
* **matplotlib** and **seaborn**: For data visualization.
* **tensorflow**: For building and training the LSTM model.
* **scikit-learn**: For additional preprocessing and evaluation metrics.

You can install the required modules using pip if they are not already installed.

#### **2. Load Data**

Use the Yahoo Finance API to download historical stock price data. The data typically includes several attributes like Open, High, Low, Close, and Volume, with the focus here being on the **Closing Price**. You need to define:

* **Stock symbol**: The ticker of the stock you want to predict.
* **Start and end dates**: Define the time frame for historical data.

Once the data is fetched, perform a basic check for missing values and ensure data integrity.

#### **3. Exploratory Data Analysis (EDA)**

Before training the model, analyze the stock price data for better understanding:

* **Data inspection**: Look at the first few rows to understand the structure.
* **Missing values**: Check if there are any missing values in the dataset.
* **Visualizations**: Plot the historical closing prices to identify trends and patterns over time.
* **Summary statistics**: Calculate basic statistics like mean, median, min, max, and standard deviation to summarize the data.

This analysis provides insights into the data distribution and helps in preparing it for the model.

#### **4. Data Preprocessing**

Prepare the data for input into the LSTM model:

* **Select relevant features**: Focus on the closing prices, but other features can also be used depending on the scope.
* **Normalize the data**: Use scaling (e.g., Min-Max Scaling) to normalize the values, as LSTM models are sensitive to feature scaling.
* **Create sequences**: Prepare sequences of data (e.g., the last 60 days' data) to predict the next day's stock price.
* **Split the data**: Divide the dataset into training and test sets, typically using an 80/20 split, where 80% of the data is used for training and the remaining 20% is for testing.

#### **5. Build the LSTM Model**

Design and build the Long Short-Term Memory (LSTM) model. The LSTM architecture is ideal for sequential data, such as time series data (stock prices). Follow these steps:

* **Model architecture**:
  + Use **LSTM layers** with suitable units (e.g., 50 units).
  + Add **Dropout layers** to reduce overfitting.
  + The output layer should be a **Dense layer** with one neuron to predict the next stock price.
* **Compile the model**: Choose an optimizer like **Adam** and loss function **Mean Squared Error** for regression tasks.
* **Train the model**: Use the training data and validate the model using the test set. Select an appropriate number of epochs (e.g., 10-20 epochs).

#### **6. Evaluate the Model**

After training, assess the performance of the model using the test dataset:

* **Predictions**: Use the trained model to predict stock prices on the test data.
* **Inverse scaling**: Convert the normalized predicted values back to the original price scale.
* **Plot results**: Visualize the predicted stock prices against the actual stock prices to understand the model's performance.

#### **7. Model Performance Metrics**

Evaluate the model's accuracy and prediction capability using common metrics:

* **Mean Absolute Error (MAE)**: Measures the average magnitude of the errors.
* **Mean Squared Error (MSE)**: Measures the average squared difference between predicted and actual values.
* **Root Mean Squared Error (RMSE)**: A common measure to understand the model's accuracy in predicting stock prices.

These metrics provide a quantitative understanding of the model's error.

#### **8. Final Report and Submission**

* **Document all steps**: Include the methodology, data analysis, model architecture, evaluation, and results.
* **Visualizations**: Present plots and graphs to support your analysis and model performance.
* **Interpretation of Results**: Summarize the results, explaining the implications of the model’s performance, and discuss any challenges or observations during the project.