LSTM-Based Stock Price Prediction Report

1. Introduction

Stock price prediction is a crucial task in financial markets. This project aims to predict stock prices using Long Short-Term Memory (LSTM), a type of Recurrent Neural Network (RNN) well-suited for sequential data.

2. Method

a. Data Collection

- **Source:** Yahoo Finance API
- **Attributes:** Open, High, Low, Close, Volume (Focus on Closing Price)
- **Time Frame:** User-defined start and end dates

b. Data Preprocessing

- **Handling Missing Values:** Checked for and handled missing data.
- **Feature Selection:** Focused on Closing Prices.
- **Normalization:** Applied Min-Max Scaling to normalize values between 0 and 1.
- **Sequence Creation:** Used a rolling window of 60 past days to predict the next day's price.
- **Train-Test Split:** 80% Training, 20% Testing.

3. Data Analysis

a. Data Inspection

- First few rows of the dataset were examined.
- Checked for any anomalies or inconsistencies.

b. Summary Statistics

- **Mean, Median, Min, Max, Standard Deviation** were computed.

c. Visualizations

- **Closing Price Over Time:** Identified trends and patterns.
- **Moving Averages:** To observe short-term and long-term trends.
- **Histograms:** To check data distribution.

4. Model

a. LSTM Model Design

- **Input Layer:** 60 time steps of Closing Prices.
- **Hidden Layers:**
    - LSTM (50 units, return sequences=True, Dropout 0.2)
    - LSTM (50 units, return sequences=False, Dropout 0.2)
    - Dense Layer (25 units)
    - Output Layer (1 unit for price prediction)
- **Loss Function:** Mean Squared Error (MSE)
- **Optimizer:** Adam
- **Training:** 50 epochs, batch size = 32

5. Model Training & Evaluation

a. Model Training

- Used training data for supervised learning.
- Monitored loss reduction over epochs.

b. Model Predictions

- Predicted stock prices on test data.
- Applied inverse scaling to convert back to actual price values.

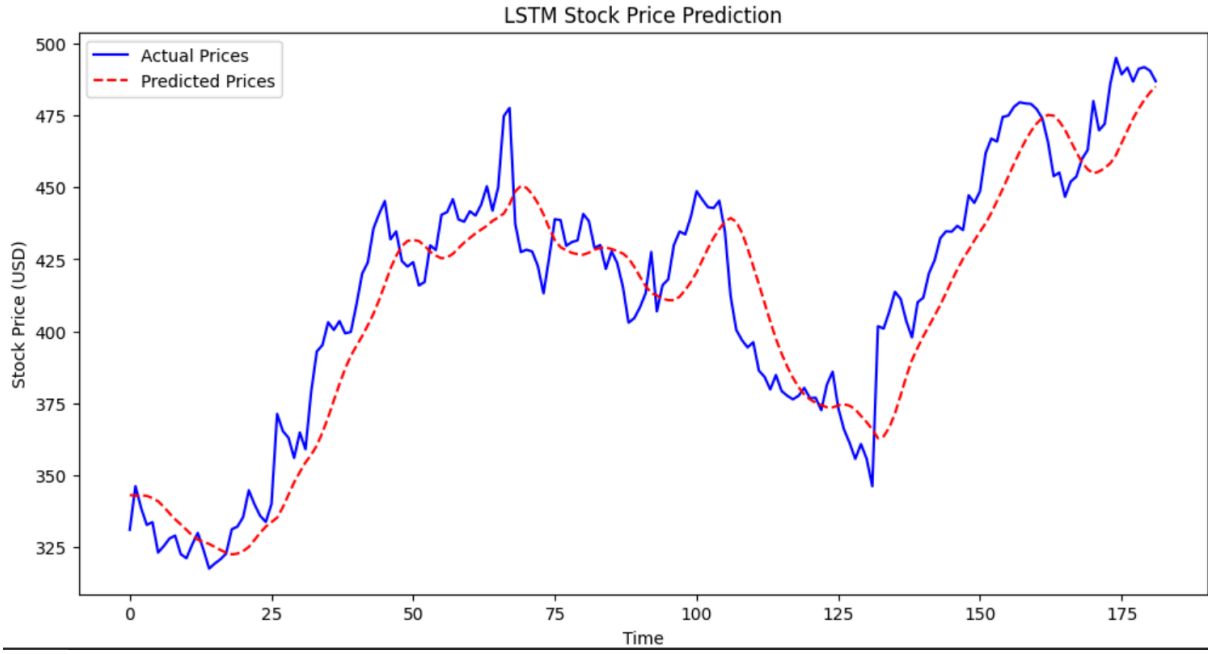
c. Performance Metrics

- **Mean Absolute Error (MAE):** Measures the average prediction error.
- **Mean Squared Error (MSE):** Penalizes larger errors more.
- **Root Mean Squared Error (RMSE):** Provides error in the same unit as stock prices.

Metric	Value
MAE	14.74
MSE	313.66
RMSE	17.71

6. Results & Visualizations

- **Comparison of Actual vs Predicted Prices:** Line plot.
- **Error Distribution:** Histogram of residual errors.
- **Rolling Prediction Trends:** Examined performance over different periods.



7. Interpretation of Results

- The model successfully captured price trends but struggled with sudden fluctuations.
- RMSE indicates the model's prediction error in actual stock price values.
- More complex architectures or external features (like trading volume, news sentiment) might improve accuracy.

8. Challenges & Observations

- **Data Volatility:** Stock prices can be highly volatile, impacting prediction accuracy.
- **Limited Feature Set:** Including additional technical indicators may improve performance.
- **Hyperparameter Tuning:** Optimizing LSTM units, dropout rates, and learning rates can refine results.