STOCK PRICE FORECASTING ANALYSIS – UBER (UBER)

This project focuses on analyzing historical Uber stock price data and building predictive models to forecast future closing prices. The dataset was sourced from Kaggle, containing daily stock prices of various companies, and Uber was chosen for detailed exploration.

Data Exploration & Preprocessing

- **EDA:**

  Visualized historical trends of closing prices and trading volumes. Identified a significant trading volume spike on **December 15, 2023**, linked to after-hours market activity and company updates.

- **Feature Engineering:**

 Created time-series features such as:

  - **Lag features:** Close_lag1, Close_lag2, Close_lag3

  - **Rolling statistics:** 7-day moving average (Close_7d_ma) and standard deviation (Close_7d_std).

  - **High volume flag** to highlight unusual trading days.

- **Data Quality:**

  Checked for missing dates and duplicates. Since stock markets don't trade every day (holidays/weekends), missing dates were left as-is, while duplicates were removed.

Modeling Approach

Three models were built and compared on a **90-day test set**:
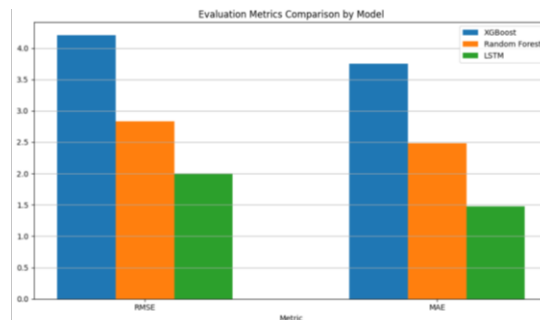
1. **XGBoost** (RMSE ~5.95)

2. **Random Forest** (RMSE ~4.34)

3. **LSTM (Long Short-Term Memory Network)** – Achieved the **best performance** with RMSE ~2.23 and MAE ~1.91.

Stock Close Price: Actual vs Predictions (2025)

The LSTM was chosen due to its ability to capture sequential dependencies and smooth trends in stock price data.

LSTM Architecture & Tuning

- Input features: Lag values, rolling averages, and volume-based flags.

- Architecture: Single LSTM layer with **64 units**, dropout of **0.3**, followed by a dense output layer.

- Tuning experiments (units, dropout, epochs, batch size) led to the selection of the **best configuration**:
  32 units, 0.0 dropout, 50 epochs, batch size 16.

Actual vs LSTM Predictions (Original vs Tuned)
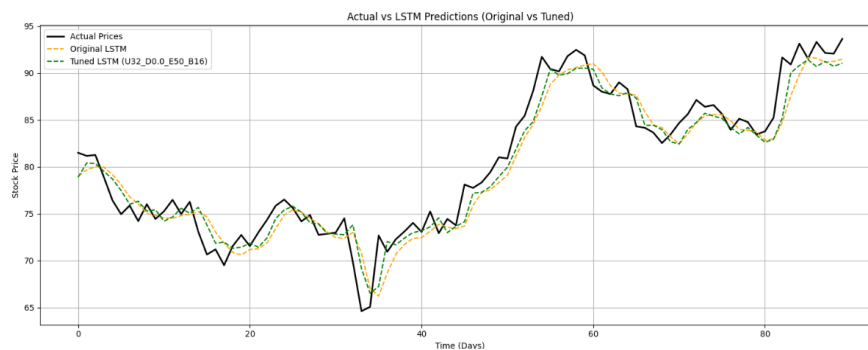
Conclusion

The analysis demonstrated that LSTM outperforms traditional models like XGBoost and Random Forest for predicting Uber's stock closing price, achieving the lowest RMSE and MAE on the test set. The final tuned LSTM model effectively captures sequential trends, making it suitable for reliable short-term (1-day) forecasts.