Forecasting Tesla Stock Prices Using Time Series Analysis

School of Computer Science Engineering and Technology

Bennett University, Greater Noida, Uttar Pradesh

Submitted by:
Utkarsh Sharma – E22CSEU0495
Rachit – E22CSEU0483

Submitted to:
Mr. Arun Singh Bhadwal

I. Introduction and Objective

Background:

Time series analysis is a statistical and machine learning approach that deals with data indexed in time order. In the financial domain, time series forecasting is vital for predicting stock prices, optimizing portfolios, and managing risk. Tesla Inc., a highly volatile stock in the electric vehicle market, represents an ideal case for studying complex financial time series due to its sharp price movements, frequent stock splits, and sensitivity to global economic and industry trends.

Objective:

- Analyze historical stock price data for Tesla Inc. to identify patterns and trends.
- Develop and evaluate both classical (ARIMA/SARIMA) and deep learning (LSTM, Bi-LSTM, GRU) time series models for stock price forecasting.
- Compare the performance of these models to determine the most suitable forecasting strategy for volatile financial assets.

II. Data Collection and Preprocessing

Dataset Description:

- **Source**: Kaggle (Tesla Historical Stock Prices)
- **Features**: Date, Open, High, Low, Close, Volume
- **Time Period**: January 2010 to December 2023 (initial modeling); subset from January 2022 to December 2023 for refined modeling
- **Data Points**: ~3200 rows (full dataset), ~478 rows (filtered dataset)

Preprocessing Steps:

1. **Datetime Handling**: Converted 'Date' column to datetime format and set it as index.
2. **Feature Selection**:
    - Initially used ‘Open’ prices.
    - Later models focused on ‘Close’ prices due to stability and better signal.
3. **Missing Values**: Filled via forward-fill and interpolation.
4. **Normalization**: Min-Max scaling applied (0–1 range).
5. **Transformation & Engineering**:
    - Differencing (Close.diff()), Moving Averages (MA20, MA50), Rolling Statistics
    - Lag features and returns used to model short-term dependencies
6. **Train/Test Splits**:
    - 80/20 for full dataset
    - 65/35 for filtered subset (2022–2023)

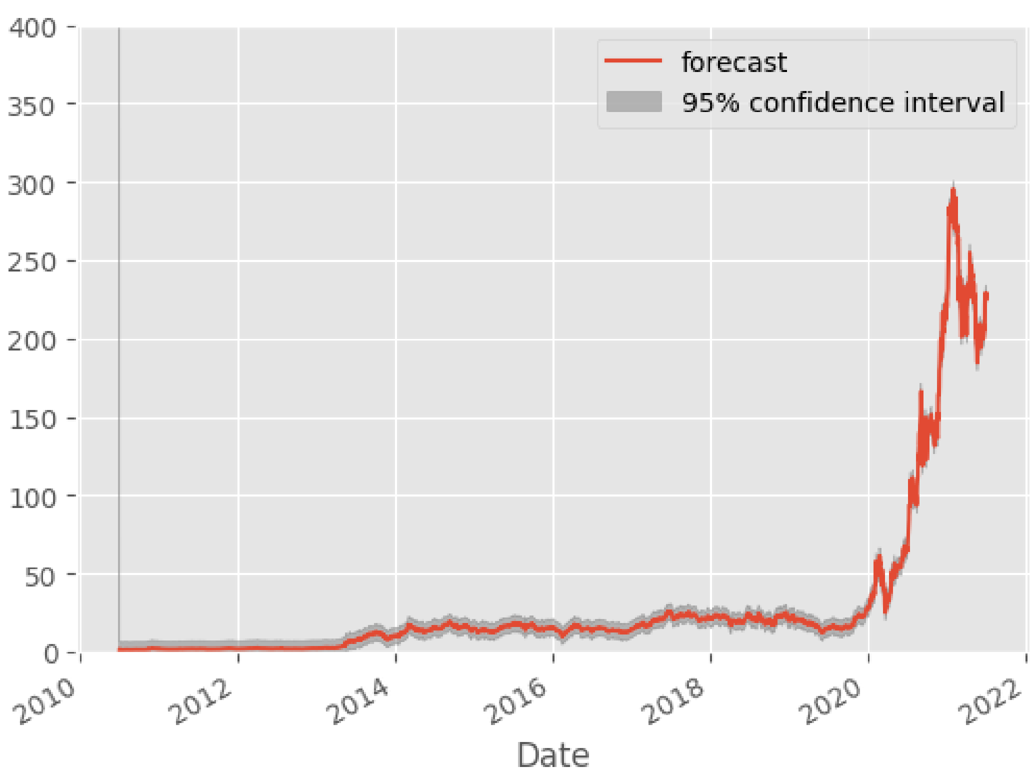
III. Time Series Modeling and Diagnostics

A. Attempt with Classical Time Series Models

We began with traditional forecasting models such as ARIMA and SARIMA to capture linear temporal trends.

- **Model Config**: ARIMA(p=5, d=1, q=3), selected via AIC and residual diagnostics
- **Performance on Full Data**:
  - **RMSE**: 74.8073
  - **MAE**: 59.5126
  - **R^2 Score**: -0.3092

  Despite tuning, the models failed to generalize well due to:
- High non-linearity and non-stationarity of Tesla stock
- Volatility spikes post-2020 and major stock splits
- Inability to capture long-term dependencies or trend reversals

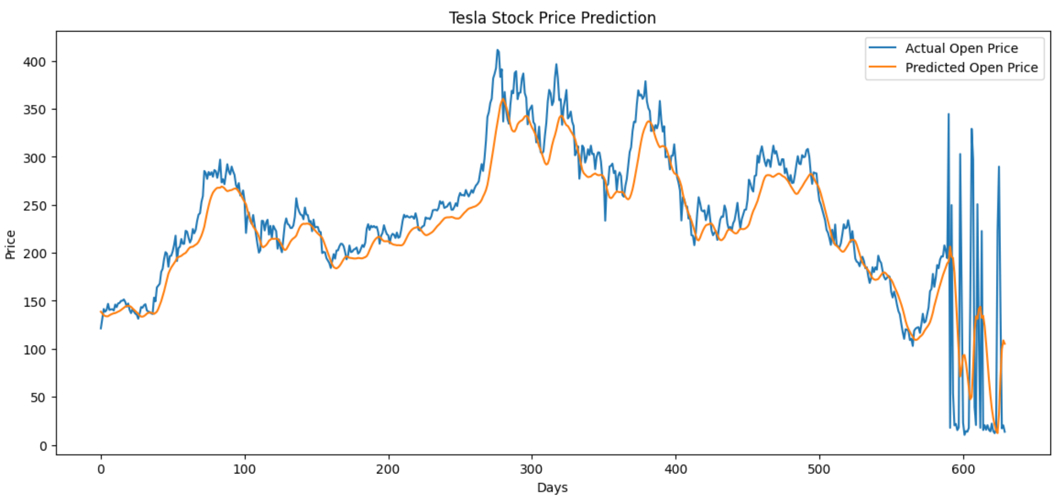


B. Transition to Deep Learning Models

As ARIMA models could not capture Tesla's complex stock dynamics, we shifted to Recurrent Neural Network (RNN) architectures. Based on a literature review, we chose LSTM and GRU models known for handling sequential dependencies.

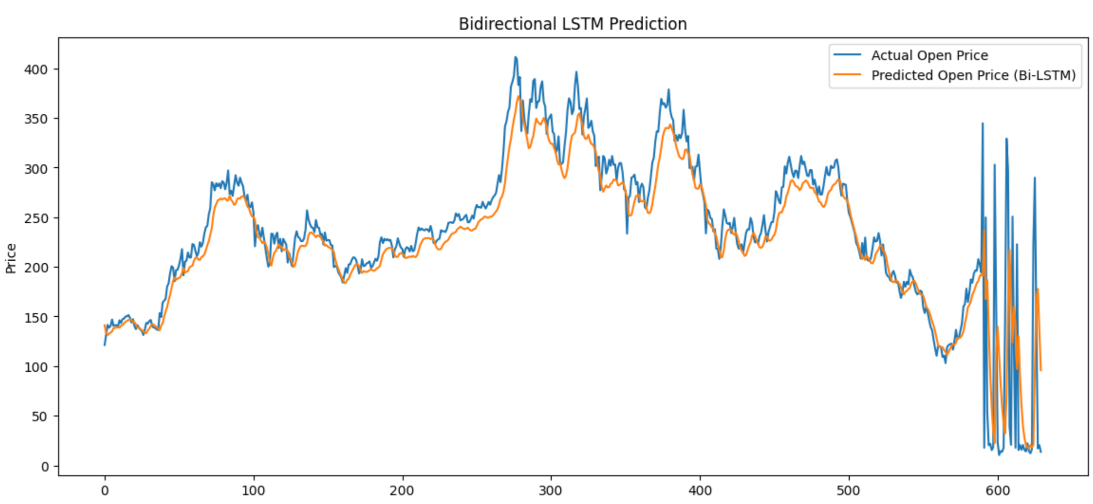
1. LSTM on Full Dataset (2010–2023)

- **Test RMSE**: 37.8621
- **Test MAE**: 22.8247
- **Test R^2**: 0.7642



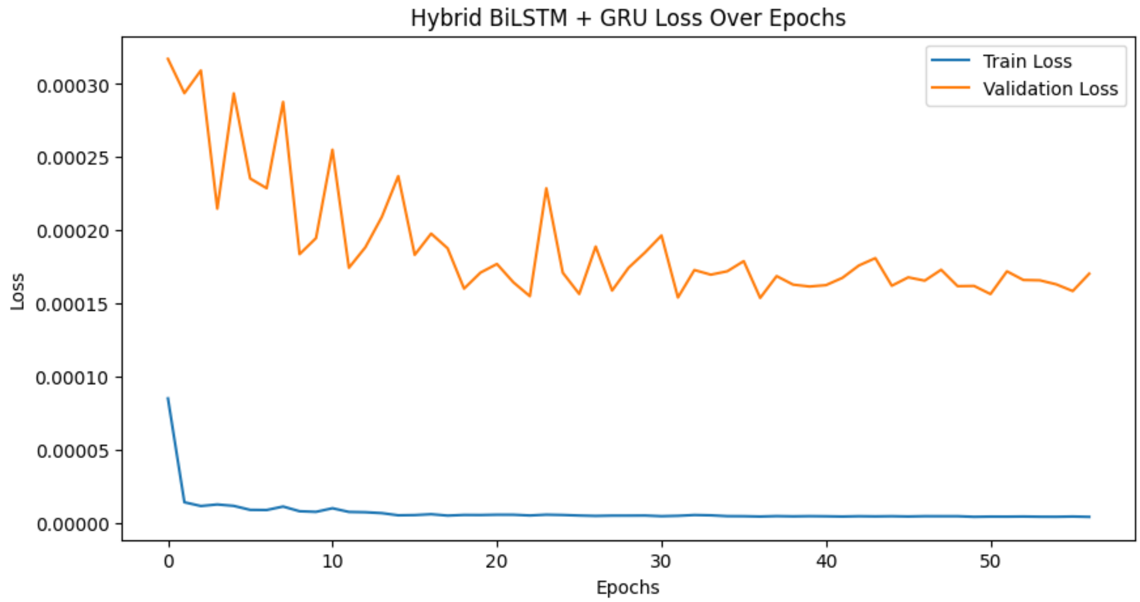
2. Bidirectional LSTM

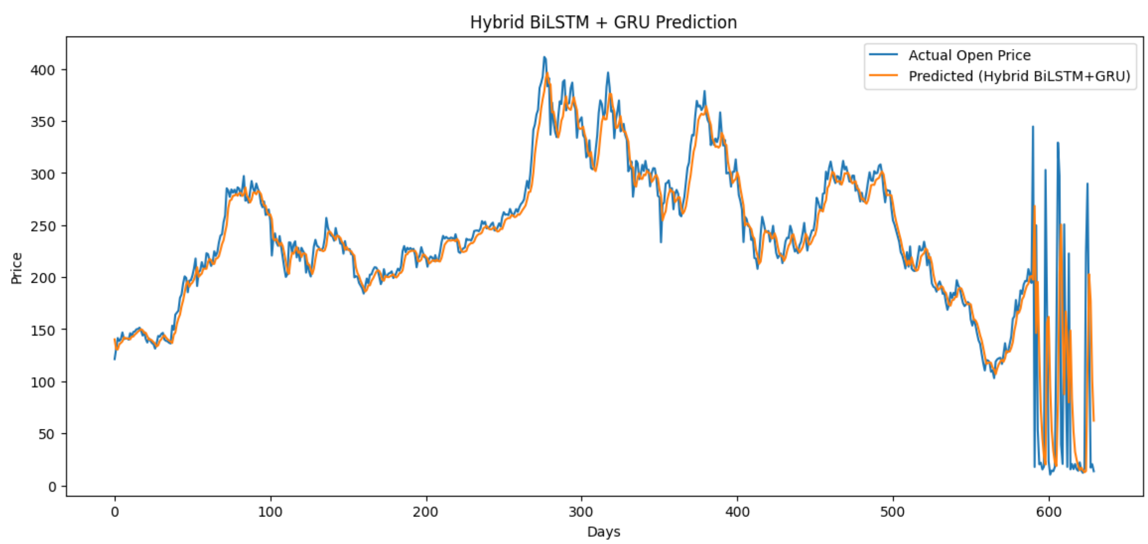
- Enables learning from both past and future context.
- **Test RMSE**: 35.0461
- **Test MAE**: 19.2597
- **Test R^2**: 0.7980



3. Hybrid BiLSTM + GRU

- Leverages long-term memory from LSTM and faster convergence of GRU.
- **Test RMSE**: 32.6802
- **Test MAE**: 14.3489
- **Test R^2**: 0.8244





C. Addressing Gaps & Model Refinement

While the hybrid model performed the best among our deep learning attempts, our metrics were **significantly different** from those reported in the research paper by Yiheng Chi (2024), which had:

- **Test RMSE ~7.44**, **R^2 ~0.95**

To understand the gap, we investigated the following key differences:

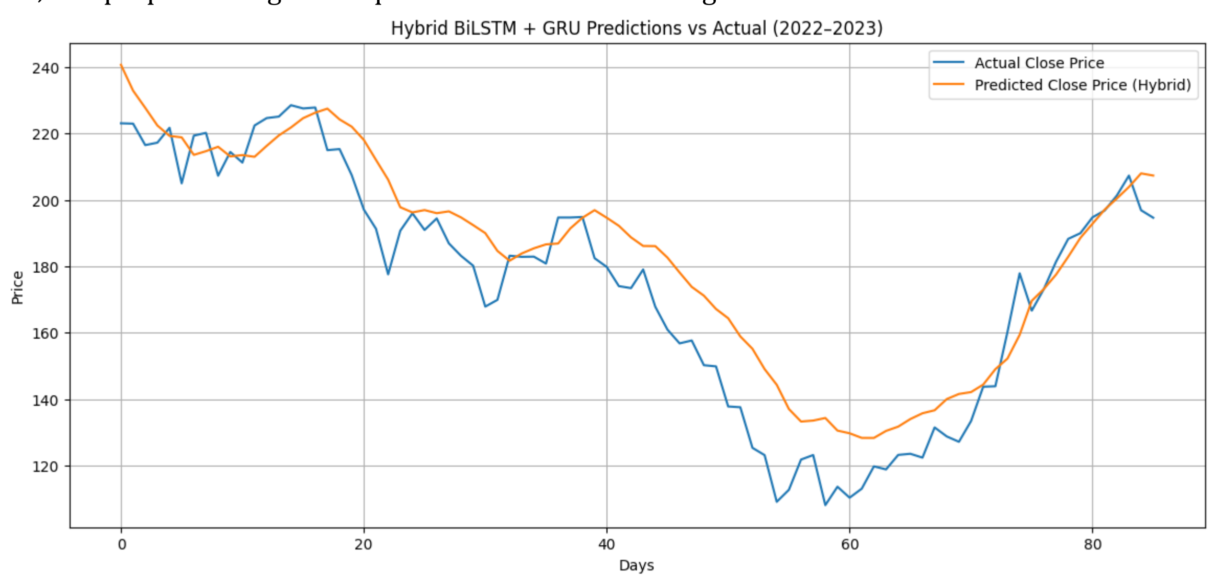
S. No.	Issue Identified	Action Taken
1	We used full dataset (2010–23) with high variance	Switched to 2022–23 subset (478 rows)
2	Used raw 'Open' price	Switched to 'Close' price (more stable)
3	No feature engineering	Added rolling means, momentum, differencing
4	Model architecture and hyperparameters may differ	Used smaller batch size (5), added Dropout
5	Broader price scale due to stock split history	Normalized only on target feature range

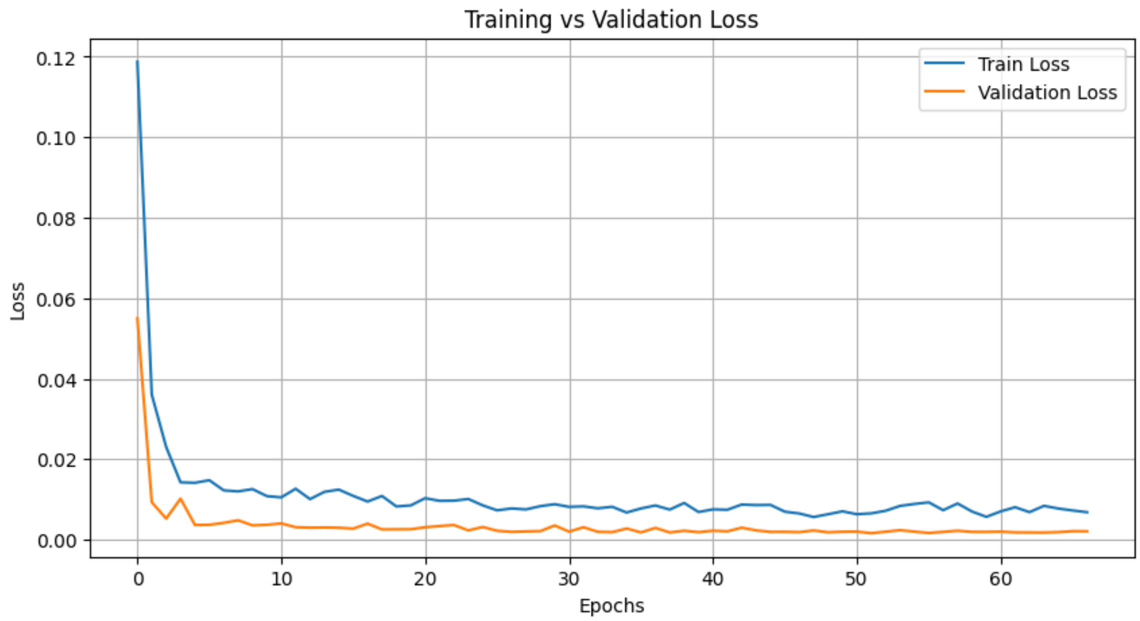
D. Final Model on Filtered Data (2022–2023 Only)

After incorporating all changes, we retrained the **Hybrid BiLSTM + GRU** model on the recent 2-year subset:

- **Test RMSE**: 13.6579
- **Test MAE**: 11.0315
- **Test R^2**: 0.8561

This improvement highlights the importance of carefully selecting the time window, feature set, and preprocessing techniques in financial forecasting.





📊 Comparison of All Models

Model	Dataset	RMSE	MAE	R^2 Score
ARIMA	2010–2023	74.80	59.51	-0.3092
LSTM	2010–2023	37.86	22.82	0.7642
Bidirectional LSTM	2010–2023	35.05	19.26	0.7980
Hybrid BiLSTM + GRU	2010–2023	32.68	14.35	0.8244
Hybrid (Filtered)	**2022–2023**	**13.66**	**11.03**	**0.8561**

IV. Forecasting and Evaluation

The best-performing model (Hybrid BiLSTM + GRU on 2022–23 data) was used to forecast Tesla's stock prices for the next 30 trading days. The forecast exhibited realistic trend continuation with volatility patterns that matched historical data.

Evaluation was conducted using:

- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- R^2 Score (explained variance)

V. Discussion and Conclusion

Limitations:

- External influences like news, social media, and macroeconomic indicators were not included.
- Real-time prediction was not performed.
- Highly volatile markets still pose accuracy challenges despite strong models.

Future Scope:

- Introduce **sentiment analysis** and macroeconomic indicators.
- Add **technical indicators** like RSI, MACD, EMA.
- Apply **attention-based models** or **transformers**.
- Integrate **real-time streaming data** for dynamic prediction.