StockSight: Stock Price Prediction

Predicting daily stock closing prices(AAPL) using LSTM

Presented by:

- **Ansh Mahajan — 102303380**
- **Ayush Mahajan — 102303087**

Submitted to: Mr. Amardeep Singh

What is StockSight?

- A deep learning–based system designed to forecast daily closing stock prices with high accuracy.
- Leverages Long Short-Term Memory (LSTM) neural networks, renowned for their effectiveness with sequential data.
- Predicts future price trends by analysing intricate patterns within past sequential data.
- Developed using Python, Google Colab, TensorFlow, yfinance for data acquisition, and sklearn for preprocessing utilities.

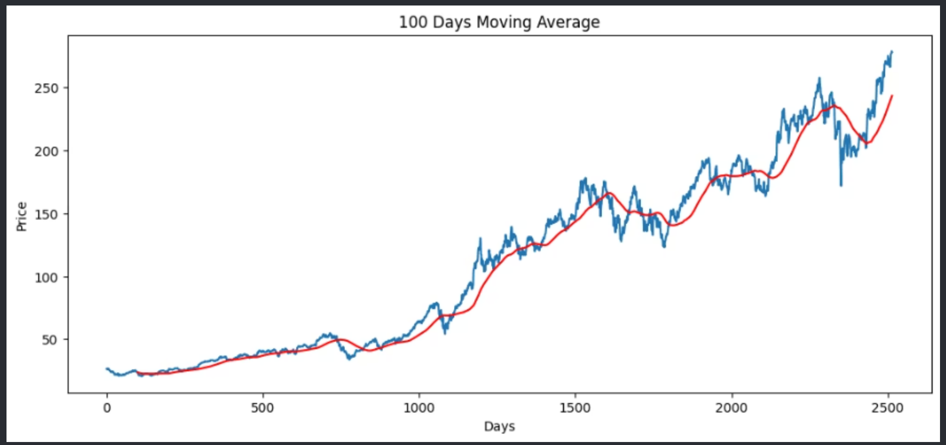
Dataset & Input Features

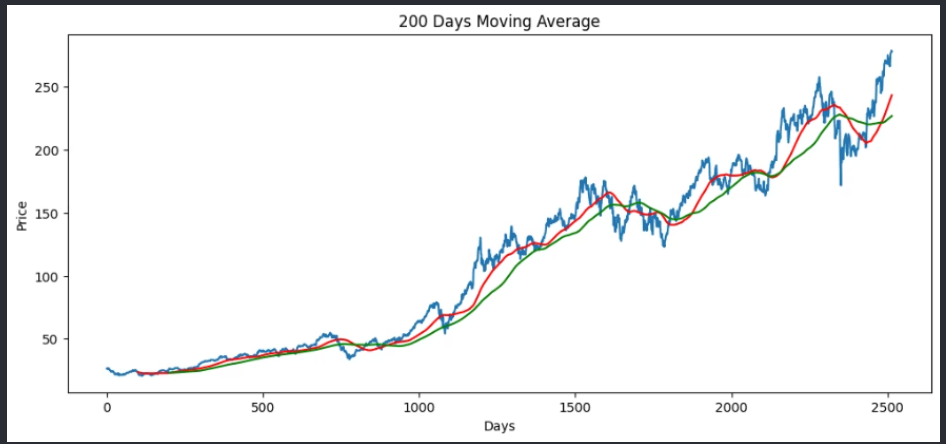
Data Sourcing & Core Columns

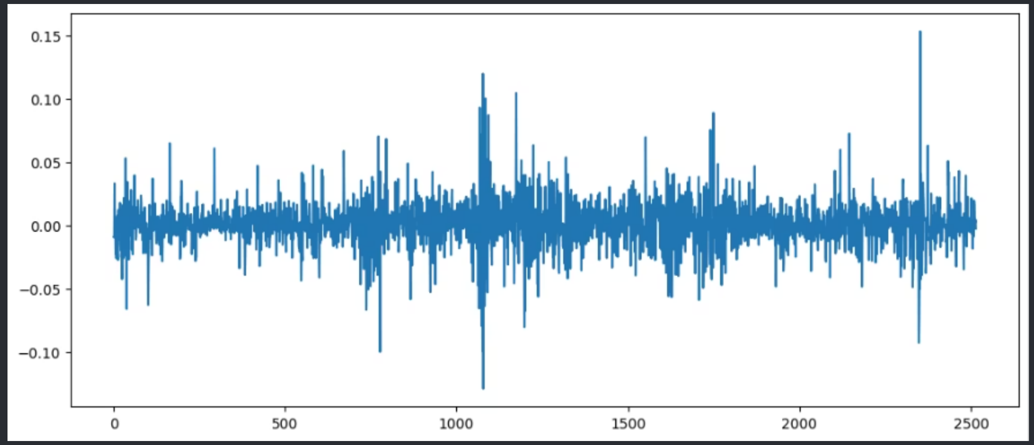
Data is dynamically sourced from **Yahoo Finance** using the robust yfinance library. Core columns include: Open, High, Low, Close, and Volume. We are using Closing Price of the stock for our Model training and prediction.

Additional data visualisation:

- 100-day Moving Average (MA) to identify longer-term trends.
- 200-day Moving Average (MA) for broader market direction insights.
- Daily % Change to capture volatility and momentum.







Preprocessing Steps

Feature Selection

Selected Closing Price as the main feature for prediction, providing a focused and reliable target for the LSTM model.

Missing Values

Thorough verification and handling of any missing values to ensure data quality and model robustness.

Data Splitting

Implemented a chronological train-test split, preserving time-series integrity by avoiding data shuffling.

Normalization

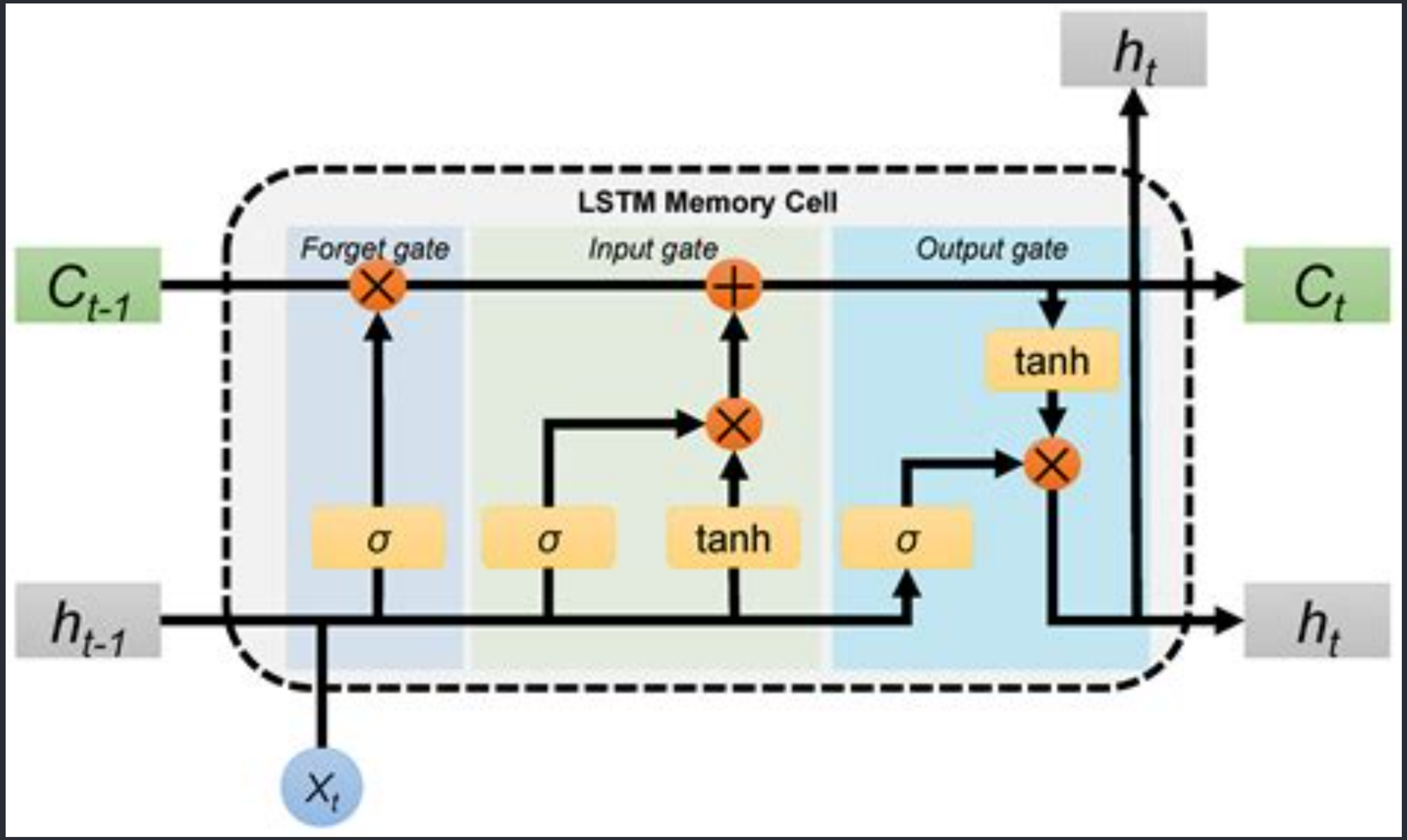
Applied **MinMaxScaler (0–1)** across all features, ensuring optimal neural network convergence.

Sequence Creation

Utilised a sliding window of **100 past days** to generate input sequences for the LSTM model.

LSTM Architecture

The LSTM cell processes past hidden states and current inputs through forget, input, and output gates, allowing the model to retain or discard information and learn long-range dependencies.



LSTM Architectures Tested

Several LSTM model configurations were rigorously experimented with to identify the optimal architecture for stock price prediction. Each iteration aimed to reduce the Mean Squared Error (MSE), a key metric for regression tasks.

1

Model 1

LSTM(32) → Dense(1)

MSE: 0.001755475

2

Model 2

LSTM(50) → Dense(1)

MSE: 0.0012578

3

Model 3

LSTM(50, RS=True) → LSTM(50) → Dense(1)

MSE: 0.0010533223

4

Model 4

LSTM(100, RS=True) → LSTM(80) → Dense(1)

MSE: 0.000493770

5

Model 5 (Final)

LSTM(128, RS=True) → LSTM(64) → Dense(25) → Dense(1)

MSE: 0.000492830

Chosen Model Architecture

The final model architecture was selected based on its superior performance and ability to capture complex temporal dependencies in stock price data, achieving the lowest MSE.

- **Loss Function:** Mean Squared Error (MSE), ideal for regression tasks.
- **Optimizer:** Adam, chosen for its adaptive learning rate capabilities.
- **Epochs:** 100, allowing the model sufficient iterations to converge and learn patterns.

Model Training Setup

100 epochs

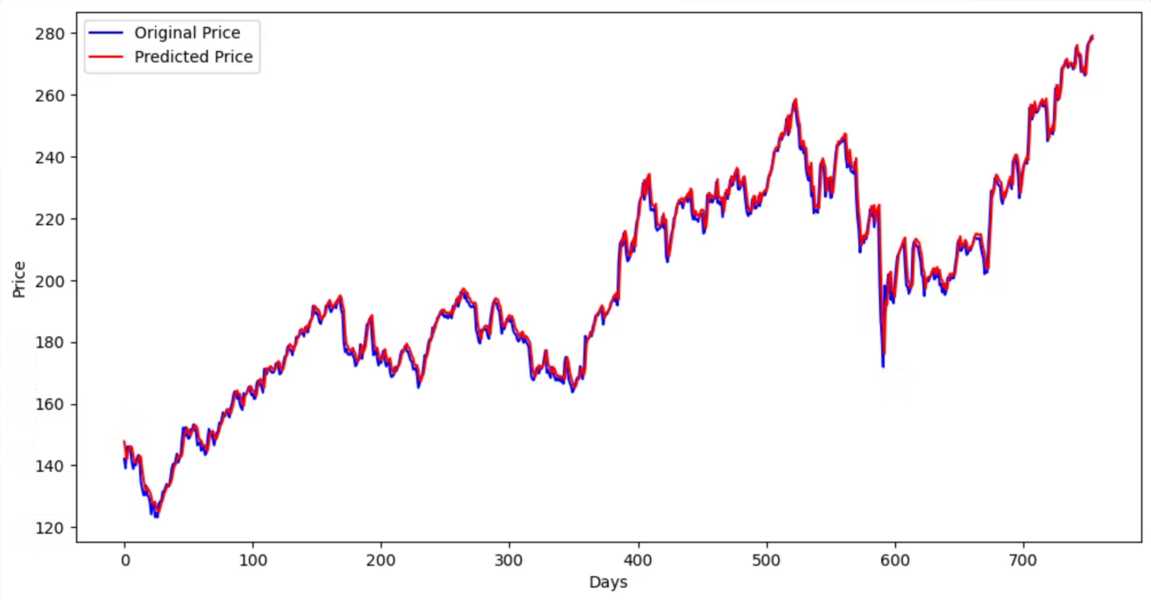
The model was trained over **100 epochs**, ensuring thorough learning from the historical data.

Steady Loss Decrease

The training process demonstrated a **steady decrease in loss** over epochs, indicating effective learning and convergence.

Historical Data

Training exclusively used real historical chronological data, maintaining the temporal sequence crucial for time-series forecasting.



Evaluation Metrics

The following metrics represent the final evaluation of the model's performance on the test dataset.

11.945	3.456	0.9896
MSE	RMSE	R-Squared
The Mean Squared Error on the original price scale, offering real-world interpretability.	The Root Mean Squared Error, providing the error in the same units as the stock price.	A measure of how well the predictions approximate the real data, with values close to 1 indicating a good fit.

Additional Metrics: Train MSE: 4.020, Train RMSE: 2.005, Train R^2: 0.998

Conclusion

Successfully developed a full end-to-end stock price forecasting pipeline, from data acquisition to prediction.

The chosen LSTM-based model demonstrated superior performance, achieving the lowest test loss among all experimented architectures.

Our StockSight system exhibits strong predictive ability, making it a valuable tool for time-series forecasting in financial markets.