STOCK MARKET FORECASTING USING TIME SERIES ANALYSIS

UNDER THE SUPERVISION OF
SRI. PHULEN MAHATO

SUBMITED BY

CHITRANKU SARKAR

ROLL NO: 6241107-07171

REG. NO: 1032111400447

SUBHODEEP SAHA

ROLL NO: 6241107-07174

REG. NO: 1032111600471

AVRODEEP PAL

ROLL NO: 6241107-07164

REG. NO: 1032111600442

INTRODUCTION

- THE STOCK MARKET IS A PLACE WHERE STOCKS REPRESENTING COMPANY OWNERSHIP ARE BOUGHT AND SOLD.
- IMPLEMENTING THE CONCEPT OF ALGORITHMIC TRADING WHICH USES AUTOMATED, PRE-PROGRAMMED TRADING STRATEGIES TO PREDICT STOCK PRICES.
- TIME SERIES FORECASTING (PREDICTING FUTURE VALUES BASED ON HISTORICAL VALUES) APPLIES WELL TO STOCK FORECASTING.
- MACHINE LEARNING MODELS LIKE SVR, ANN, AND RNN HAVE IMPROVED TIME SERIES FORECASTING.
- LSTM NETWORKS EXCEL IN STOCK PREDICTION BY STORING AND UTILIZING PAST INFORMATION EFFECTIVELY.

MOTIVATION

- THE UNPREDICTABLE NATURE OF THE STOCK MARKET DRIVES THE NEED FOR ADVANCED PREDICTIVE MODELS.
- ACCURATE STOCK PRICE PREDICTION, IMPROVED PREDICTIVE MODELS HELP INVESTORS MAKE INFORMED DECISIONS, REDUCING FINANCIAL RISKS.
- ADVANCEMENTS IN MACHINE LEARNING OFFER NEW OPPORTUNITIES TO ENHANCE TIME SERIES FORECASTING ACCURACY.
- THE SUCCESS OF LSTM NETWORKS IN OTHER DOMAINS INSPIRES THEIR APPLICATION TO STOCK MARKET PREDICTION.

OBJECTIVE

The main objective of our model is to predict the close price of a stock by analyzing complex historical data of the same, using Long Short-Term Memory (LSTM) networks by executing the following taks.

Load Stock Data (as a .CSV file) and Pre-process it

Prepare Dataset and Define the LSTM Model

Train the models for each component

Combine train and test predictions to get final prediction

Show (plot) predicted vs actual values for the test set

TECHNOLOGY USED

Python Language

Python is a rich language for Data Science and AI

Libraries Used

pandas, sklearn, numpy, tensorflow, etc.

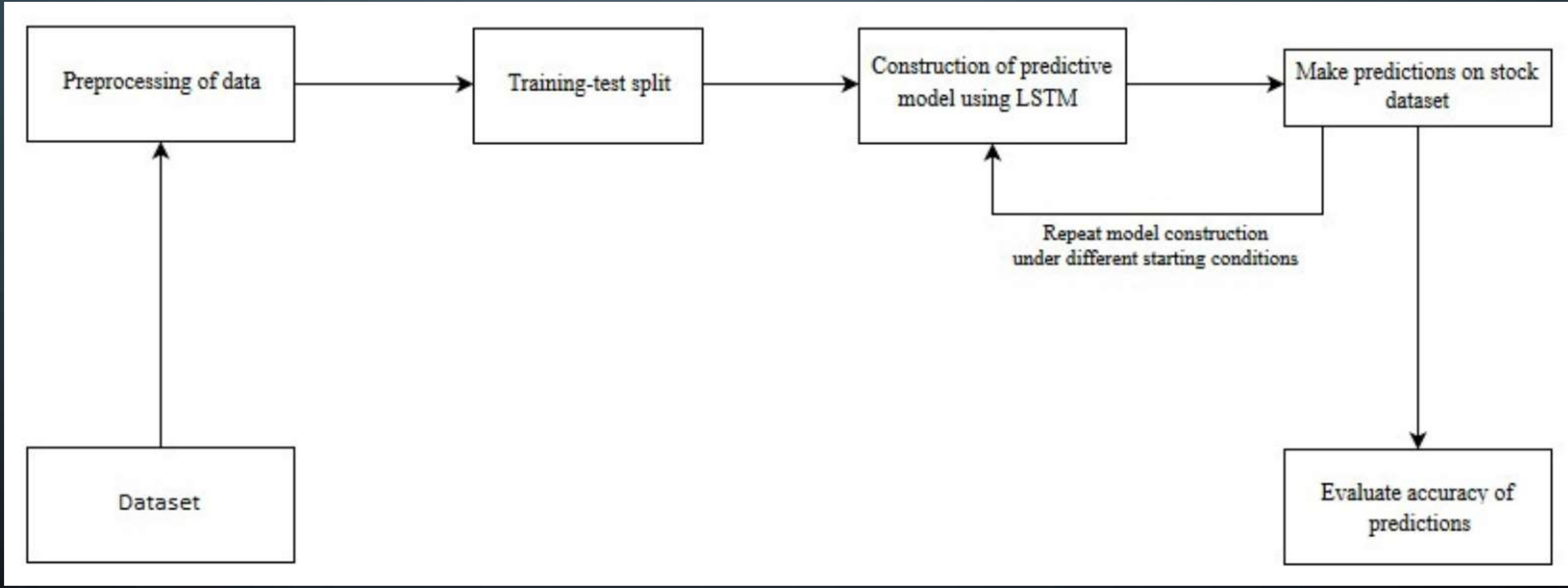
Algorithm

Long Short Term Memory (LSTM)

Environment

Google Colab is a cloud-based platform to execute Python codes.

GENERIC MODEL FOR STOCK MARKET FORECASTING



DATASET

The dataset used in this project is [TSLA] (Tesla) from June 29, 2010 to December 31, 2020. This is a series of data points indexed in time order or time series. Our goal was to predict the closing price for any given date after training.

Date	Open	High	Low	Close	Adj Close	Volume
29-06-2010	3.8	5	3.508	4.778	4.778	93831500
30-06-2010	5.158	6.084	4.66	4.766	4.766	85935500
01-07-2010	5	5.184	4.054	4.392	4.392	41094000
02-07-2010	4.6	4.62	3.742	3.84	3.84	25699000
06-07-2010	4	4	3.166	3.222	3.222	34334500
07-07-2010	3.28	3.326	2.996	3.16	3.16	34608500
08-07-2010	3.228	3.504	3.114	3.492	3.492	38557000
09-07-2010	3.516	3.58	3.31	3.48	3.48	20253000
12-07-2010	3.59	3.614	3.4	3.41	3.41	11012500

Table 1: TESLA Stock Data

DATA PREPROCESSING

The Steps of Preprocessing of Data is shown below:

Load Stock Data (as a .CSV file)	Read the CSV file. Convert the 'Date' column to datetime. Set 'Date' as the index.
Filter Data	Exclude data after 2020-12-31 due to too much noise (COVID-19 effect).
Resample Data	Resample to weekly frequency using the mean of daily adjusted close prices.
Seasonal Decomposition and drop NaN values	Perform seasonal decomposition with a multiplicative model and a period of 52 weeks. Extract trend, seasonal, and residual components. Drop Not a Number values.

DATA PREPROCESSING

Scale Data

Normalize the 'trend', 'seasonal', and 'residual' components to a range of 0 to 1, ensuring all values fall within this interval.
Fit and transform each component to achieve uniform scaling, improving model training and performance.

Split Data

Split the data into training and testing sets (last 24 months for testing).

Prepare Dataset

Define the number of previous weeks (n_input) to use for prediction.
Set the number of weeks to predict (n_output).

STOCK PRICE PREDICTOR MODEL

As the name suggests, the Stock Price Predictor Model is designed to forecast the close price of a stock by analyzing complex historical data. It is built using Long Short-Term Memory (LSTM) networks. Using LSTM networks for stock price prediction ensures the model can effectively capture long-term dependencies and patterns in financial time series data, thereby improving accuracy and reliability in forecasting future stock prices.

UNDERSTANDING THE WORKING OF LSTM

How it Works? (LSTM)

- The key to LSTM is the **Memory Cell state** which stores the information. It runs straight down the entire chain.
- LSTM has the ability to **remove or add information** to these cell state, regarded by structures called **gates.**
- **Gates** are composed of sigmoid neural net layer and a multiplication operation.
- **Sigmoid** layer outputs are **zero or one.**
- It consists of 3 gates: **Input Gate, Forget Gate, Output Gate.**

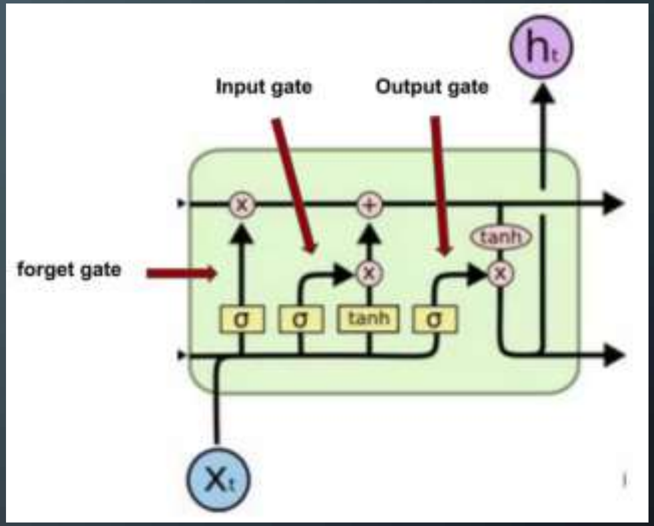


Fig 1: LSTM Architecture

UNDERSTANDING THE WORKING OF LSTM

First,

- Forget gate looks at h_{t-1} and x_t and outputs a number between 0 and 1.
- '1' represents **keep the information**
  '0' represents **remove the information.**

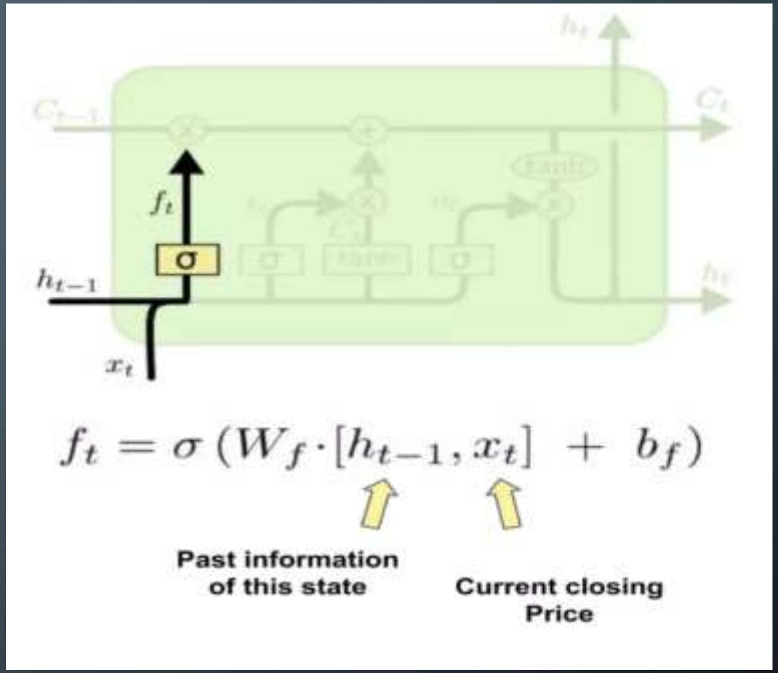


Fig 2: Working of Forget Gate

UNDERSTANDING THE WORKING OF LSTM

First,

- Forget gate looks at h_{t-1} and x_t and outputs a number between 0 and 1.
- '1' represents **keep the information**
  '0' represents **remove the information.**

Second,

- Input gate decides which values will be updated, in order to do that a tanh layer creates a vector of C_t (bar).
- Combining these two, create an update to the state.

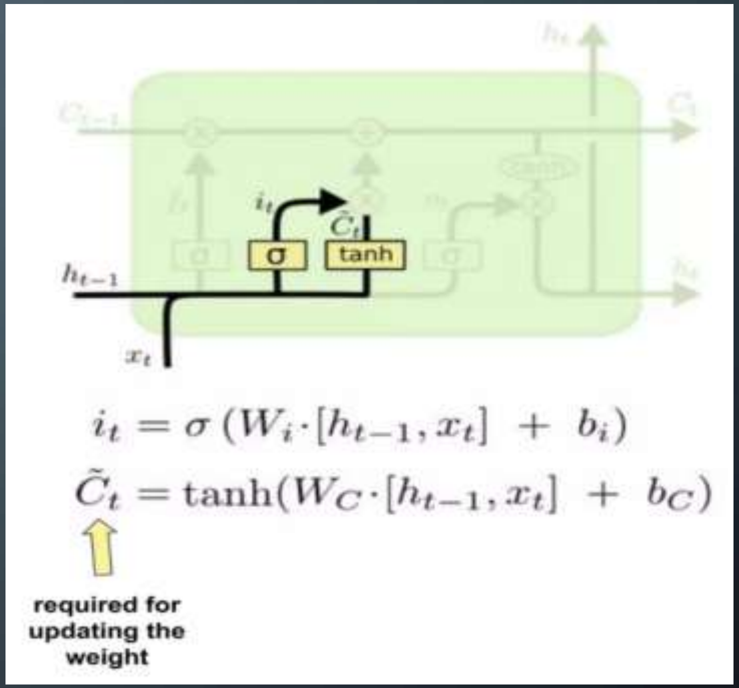


Fig 3: Working of Input Gate

UNDERSTANDING THE WORKING OF LSTM

First,

- Forget gate looks at h_{t-1} and x_t and outputs a number between 0 and 1.
- '1' represents **keep the information**
  '0' represents **remove the information.**

Second,

- Input gate decides which values
will be updated, in order to do that a tanh layer creates a vector of C_t (bar).
- Combining these two, create an update to
the state.

Third,

- It's time to update the old cell c_{t-1} to c c_t

Fourth,

- Output will be based on our **cell state.**
- A sigmoid layer will decide what parts of the cell state we are going to output.

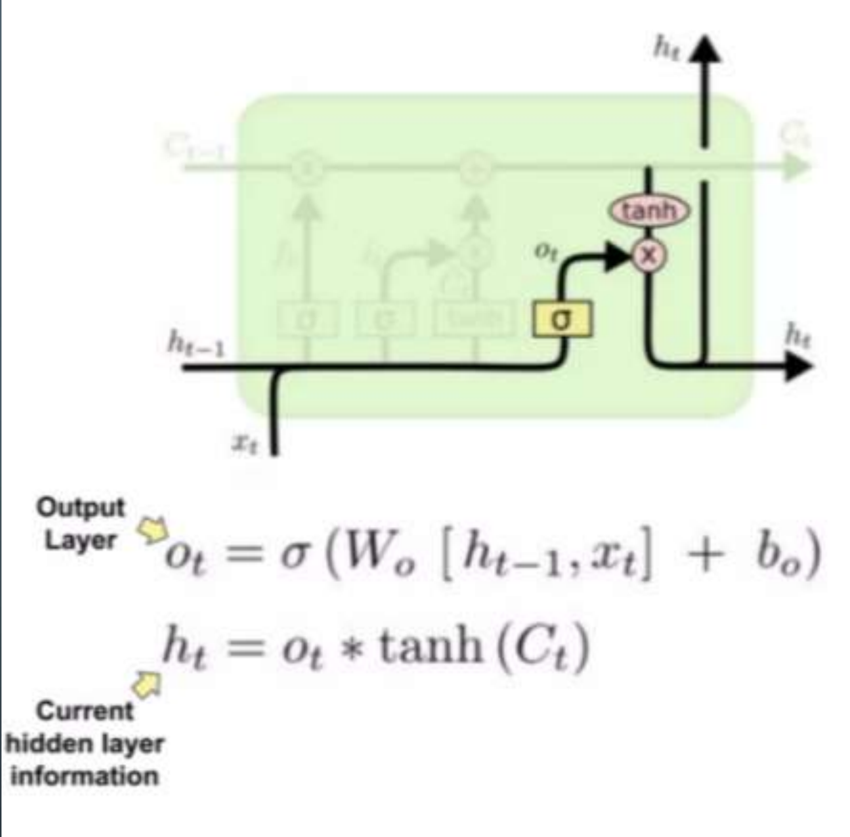


Fig 4: Working of Output Gate

VISUALIZING MULTIPLE ASPECTS OF DATA

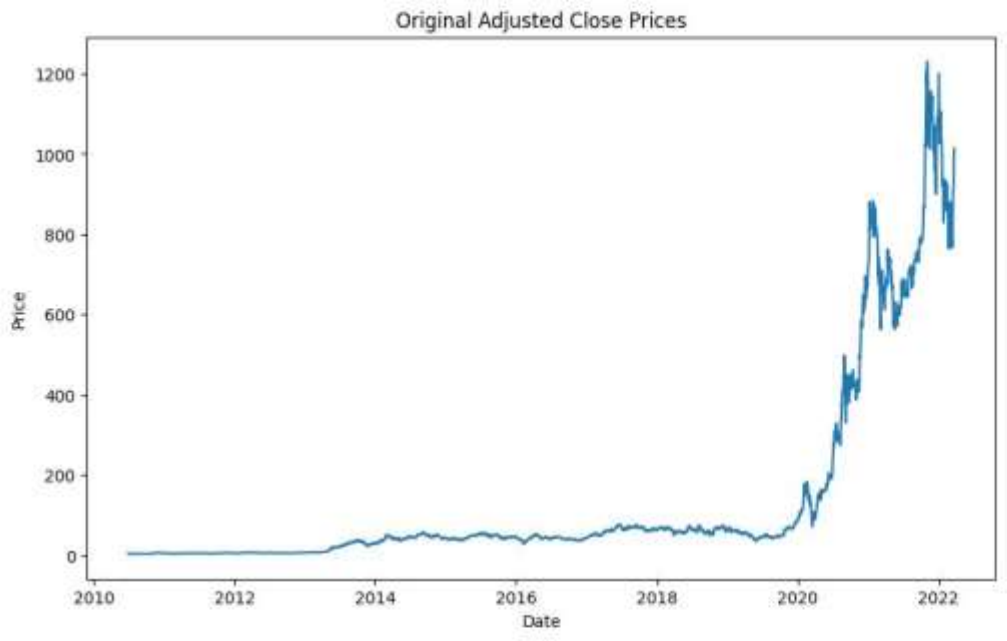


Fig 5: Original Closing Price Graph

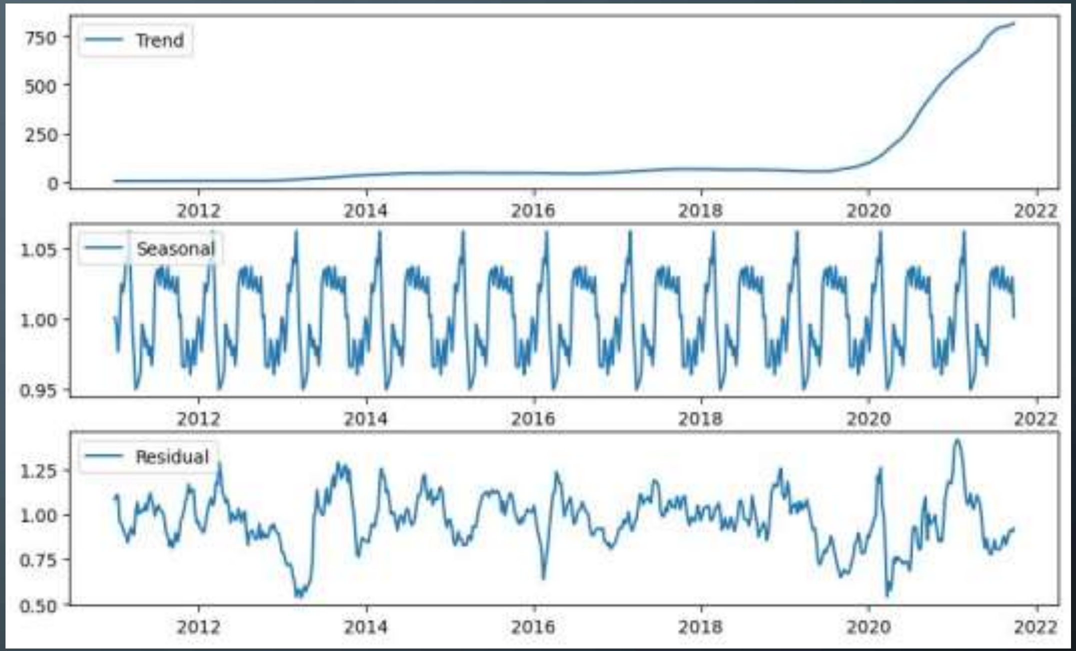


Fig 6: Seasonal Decomposition of the Data Graph

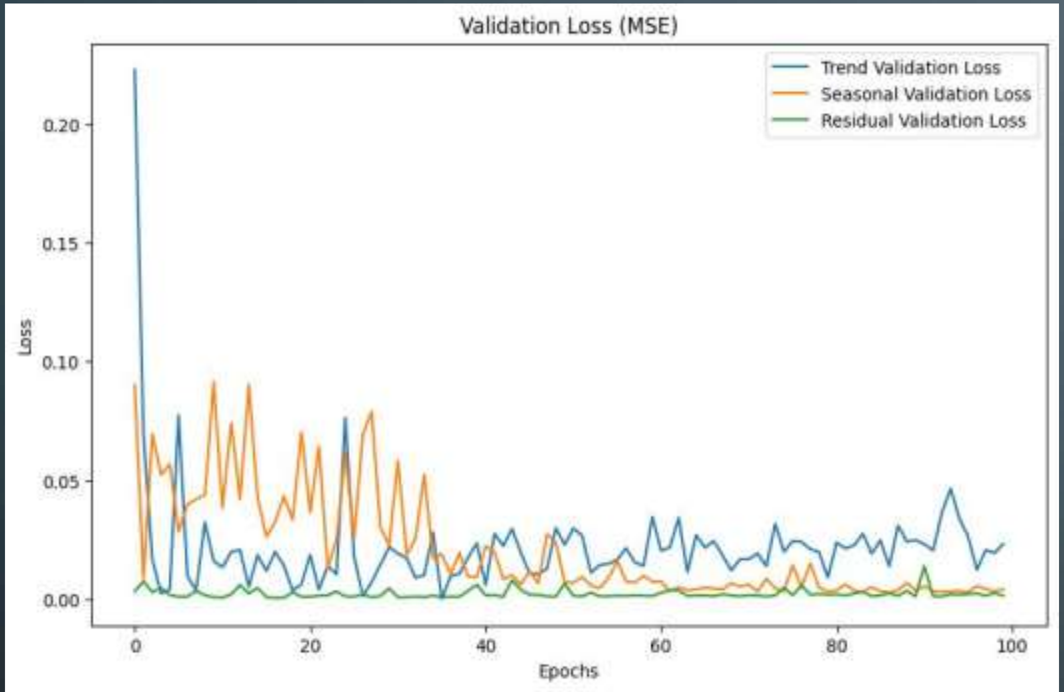


Fig 7: Validation Loss Graph

Epoch	Trend Validation Loss	Seasonal Validation Loss	Residual Validation Loss
1	0.222865	0.090186	0.003251
11	0.013341	0.038465	0.000555
21	0.018600	0.036452	0.000927
31	0.019068	0.058189	0.000704
41	0.005907	0.022225	0.001524
51	0.029744	0.006909	0.001336
61	0.020337	0.007389	0.002611
71	0.016771	0.006152	0.001600
81	0.023774	0.003589	0.001872
91	0.023108	0.005301	0.013828

Table 2: Validation Loss in Tabular Form

FINAL RESULT

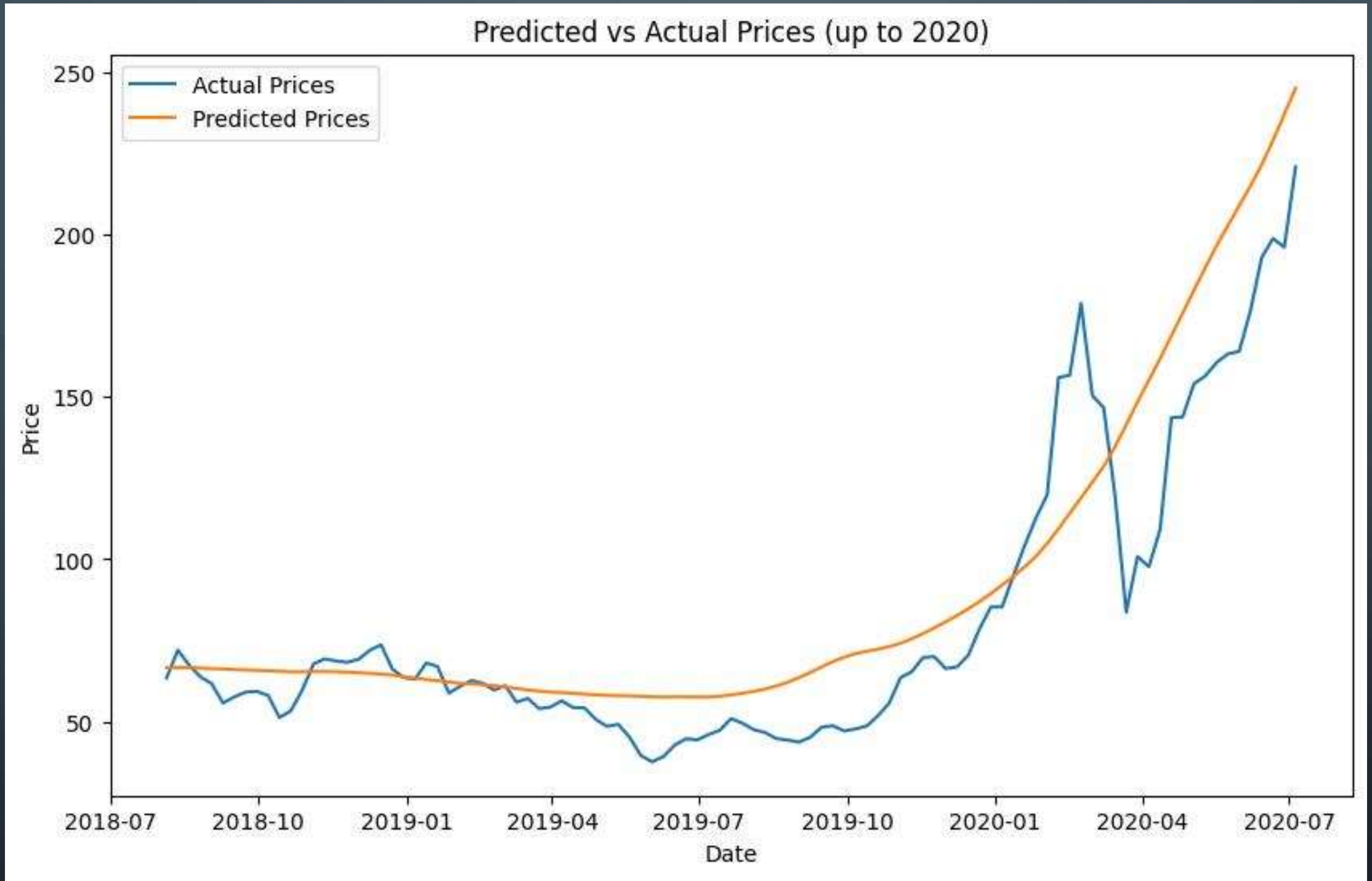


Fig 8: Actual vs Predicted Graph

CONCLUSION

- **Challenges in Stock Market Prediction:** No model can guarantee successful predictions due to the complexity and numerous influencing factors.
- **Importance of Preprocessing Techniques:** Effective preprocessing, like decomposing and differencing for stationarity, enhances model performance.
- **LSTM's Strengths:** LSTM models capture long-term dependencies, making them effective for sequential data in time series forecasting.
- **Component-wise Model Training:** Training LSTM on decomposed components (trend, seasonal, residual) allows for detailed and accurate forecasting.
- **Comparative Analysis for Better Decisions:** Comparing various time series models is crucial for making informed stock trading decisions to minimize losses and maximize profits.

FUTURE SCOPE

- **Focus on Public Sentiments:** Integrate sentiment analysis from social media and financial news to enhance stock price predictions.
- **Hybrid Model Development:** Develop a hybrid model combining historical data with sentiment analysis for more accurate forecasting.
- **Incorporate Environmental Factors:** Include environmental factors like floods and storms in prediction models to improve accuracy.
- **Expand to Cryptocurrency:** Extend the application to predict cryptocurrency trading using time series and sentiment analysis.
- **Enhance Predictive Models:** Continuously refine predictive time series models for higher accuracy in stock market forecasting.

THANK YOU!