

**Tribhuvan University**

**Faculty of Management**

Project on

**Machine Learning**

**Stock Price Prediction**

**using LSTM**

**Submitted To**

**Mr.** Pawan Niroula

Department of Information Technology

Orchid International College

**Submitted By**

Aadarsha Upadhyaya

11632/20

**BIM** Seventh Semester

# **STUDENT DECLARATION**

I hereby declare that I have completed the machine learning project titled “**Stock Price Prediction**” using an **LSTM** model under the guidance of **Mr. Pawan Niroula**, a supportive, kind, and talented instructor. This project fulfills the requirements for the 3-month **Data Science and Machine Learning** course at Orchid International College. This work is entirely my own and has not been submitted elsewhere.

**Date:** August 31, 2024

**Name:** Aadarsha Upadhyaya

**Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

# **ACKNOWLEDGEMENTS**

I would like to extend my heartfelt gratitude to **Orchid International College** for providing the opportunity to undertake this project and for the practical and valuable learning experiences it has offered.

I acknowledge that the successful completion of this project is the result of collective effort. I am deeply grateful to all those who contributed their support and assistance throughout this endeavor.

My sincere appreciation goes to **Mr. Pawan Niroula**, our project supervisor, for his invaluable guidance, constructive suggestions, and active encouragement. He helped me understand one of the most esoteric subjects with ease by simplifying intricate concepts related to mathematics, data visualizations, machine learning libraries. Through his mentorship, I learned not only how to conduct thorough research but also how to maintain a strong curiosity and a passion for learning. He provided me with a fresh perspective on the fields of machine learning and data science. His expertise and motivation have been instrumental in the completion of this project for which I am heavily indebted towards him.

Thank you everyone who helped and supported me throughout this endeavor that led to the completion of this project.

Sincerely,

**Name:** Aadarsha Upadhyaya

**TU Symbol no.:** 11632/20

# **CERTIFICATE FROM THE SUPERVISOR**

This is to certify that the machine learning project titled “**Stock Price Prediction**” has been conducted by **Aadarsha Upadhyaya** as an additional academic endeavor under the guidance of **Mr. Pawan Niroula** at **Orchid International College**. This project, carried out over the designated period, has been supervised and supported by me, ensuring adherence to academic standards and research integrity.

To the best of my knowledge, the work presented in this project is original and has not been submitted or published elsewhere. The project reflects a substantial effort and demonstrates a high level of understanding and application of machine learning techniques.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of the Supervisor

**Name:** Pawan Niroula

**Designation:** Project Supervisor

**Date:** August 31, 2024

**LIST OF ABBREVIATIONS**

| **Abbreviation** | **Definition** |
| --- | --- |
| NEPSE | Nepal Stock Exchange |
| LSTM | Long Short-Term Memory |
| MA | Moving Average |
| EMA | Exponential Moving Average |
| RSI | Relative Strength Index |
| GRU | Gated Recurrent Unit |
| RNN | Recurrent Neural Network |
| ReLU | Rectified Linear Unit |

**Table of Contents**

[STUDENT DECLARATION 2](#_dl4vl5dmabqs)

[ACKNOWLEDGEMENTS 3](#_g0z9aa16d2e5)

[CERTIFICATE FROM THE SUPERVISOR 4](#_ull1xx33n0xz)

[ABSTRACT 7](#_yto6jdt5qeao)

[INTRODUCTION 8](#_1atozr4tgv5h)

[DATASET 10](#_9due0uj6rh7j)

[MODELING 13](#_34e1lz73ty)

[RESULTS 16](#_99gdoafucfz7)

[CONCLUSION 20](#_w5e9hycjelut)

[REFERENCES 22](#_2e4ch2pyl5p7)

# **ABSTRACT**

This project focuses on developing a predictive model for stock prices on the Nepal Stock Exchange (**NEPSE**) using historical price data and technical indicators through a Long Short-Term Memory (**LSTM**) neural network. The primary goal is to offer a tool that enhances investment decisions by providing accurate stock price forecasts. Data was collected from the ShareSansar website, incorporating various technical indicators such as Relative Strength Index (**RSI**) and Exponential Moving Average (**EMA**).

The NEPSE, being the sole secondary market for Nepal's publicly listed securities, plays a crucial role in the nation's financial system. Predicting stock prices in such a volatile and dynamic market presents significant challenges due to factors like market instability, political uncertainty, and regulatory changes. Traditional prediction methods often fall short in capturing the complex, nonlinear relationships inherent in stock price movements. However, advancements in machine learning, particularly deep learning models like LSTM, have shown promise in addressing these complexities.

This project employs LSTM models, known for their capability to remember past observations and make predictions based on sequential data, to analyze NEPSE data. The integration of technical indicators with LSTM aims to improve prediction accuracy and offer valuable insights for investors. Additionally, the study highlights the potential benefits of combining traditional financial data with emerging technologies to enhance stock market forecasting in Nepal, an area still in its early stages of research. The findings suggest that LSTM models can effectively model the intricacies of stock price behavior and contribute to more informed investment decisions in the Nepali stock market.

# **INTRODUCTION**

The primary objective of this project is to develop a predictive model for stock prices using historical prices along with the technical indicators of the stock (e.g, **RSI**, **EMA**, etc.). For that, the data was scraped from the website **ShareSansar** up to a specified range. By leveraging machine learning, specifically the algorithm Long Short-Term Memory (**LSTM**) neural network, this project aims to provide decently accurate stock price predictions that could potentially assist investors while deciding on financial investments.

**Overview:**

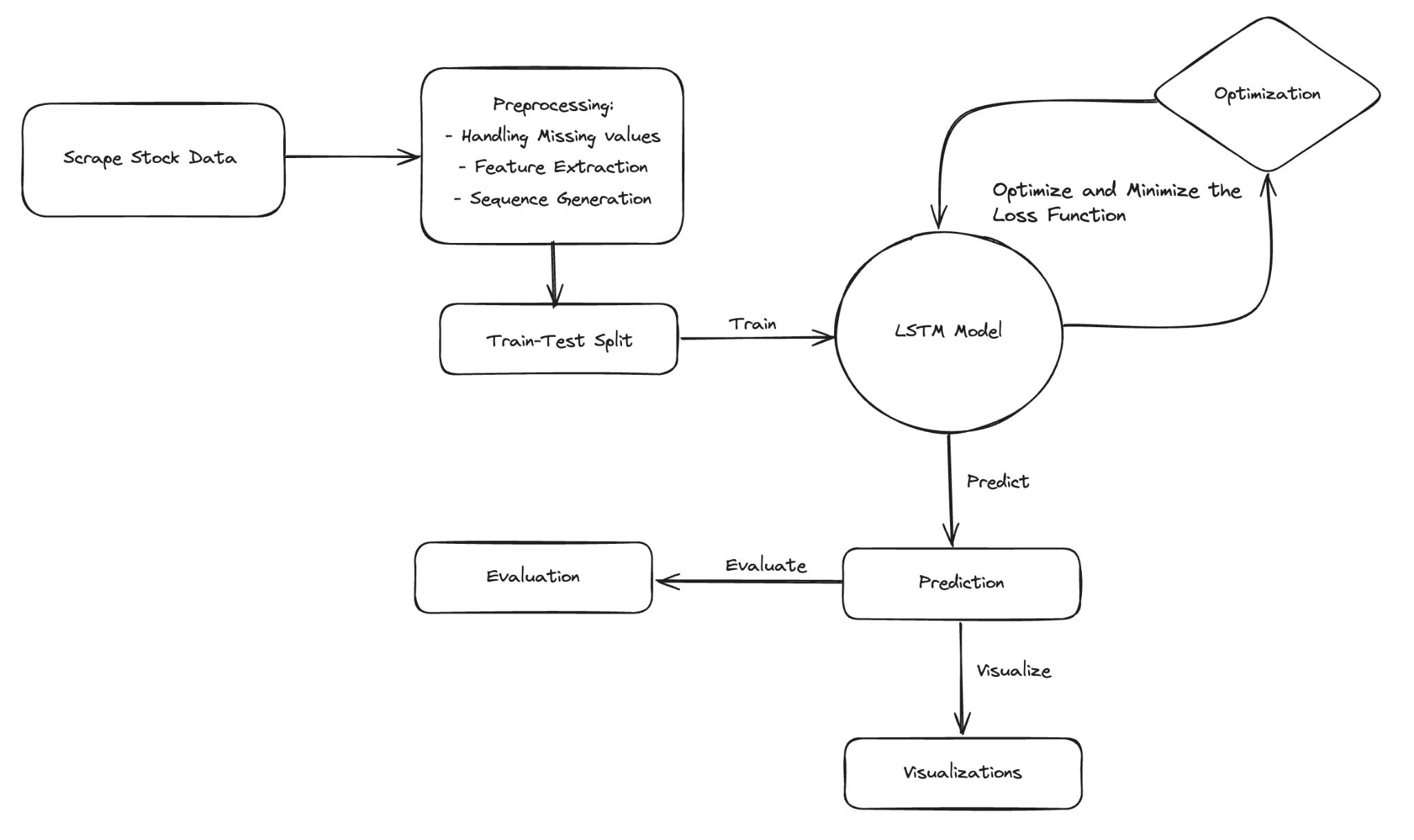
The Nepal Stock Exchange (**NEPSE**) is the primary stock market of Nepal, where investors trade and invest in publicly listed companies. It commenced its trading floor on 13th January 1994, after being established under **Securities Exchange Act, 1983**. Currently, it is the sole secondary market for the listed securities of Nepal. It plays a critical role in the financial ecosystem of Nepal, with its indices reflecting the economic health of the country. Predicting stock prices in such a market is of immense value to investors and policymakers.

It is to be noted that due to the market's inherent volatility, unstable political environment, and frequent changes in market policies and regulations along with multitude of other factors influencing price movements, accurate prediction of stock prices is a big challenge for scholars and researchers focusing on these topics. Traditional methods have been deemed to fall short in capturing the complex and nonlinear relationship in the data. However, several advancements in machine learning technologies, particularly deep learning models, have shown great promise in modeling these complexities.

Moreover, the integration of sentiment analysis from social media and news platforms has proven to enhance the predictive power of stock price models. For instance, Coyne, Madiraju, and Coelhodemonstrated in their paper that sentiment analysis from StockTwits could predict stock price movements with an accuracy of around 65%, particularly when focusing on posts from influential users. This highlights the potential for combining financial data with sentiment analysis to improve prediction accuracy **(Coyne, Madiraju, & Coelho, 2017)**.

In Nepal particularly, stock market prediction remains a relatively nascent area of research, with limited studies focusing on the application of advanced deep learning techniques. The volatility and unique market dynamics of the Nepali stock exchange, coupled with inconsistent financial news and media sentiment, present significant challenges. Recent efforts have started to explore the integration of financial news sentiment with traditional stock data to enhance the prediction accuracy, leveraging models like LSTM and GRU. These attempts highlight the growing interest in improving stock market forecasting in Nepal, with a focus on adapting global techniques to local contexts **(Shahi et al., 2020)**

LSTMmodels are particularly well-suited for time series forecasting, as they can retain memory of the past observations, which is crucial for predicting the future trends in stock prices. So this project is an attempt to leverage the LSTM architecture to analyze historical NEPSE data along with technical indicators such as moving averages and relative strength index (RSI). The model’s ability to learn from sequential data makes it an ideal choice for this task.

**Workflow:**

***Figure 1:*** *Workflow of Stock Prediction System*

# **DATASET**

The dataset used for this project was scraped from a website named ShareSansar, which is a local financial portal of Nepal, and is trusted by quite a lot of investors for its high-quality data and information related to stock market and companies. The scraped data consisted of the historical price data for a particular company listed in Nepse up to a specified range. The features include latest trading prices, volume traded, open price, close price and other relevant financial data. This comprehensive dataset forms the basis for training and evaluating the stock price prediction model.

**Features:**

The features selected for this study are essential to capture the dynamics of stock movements. The key features are listed below:

* **Date:** The specific trading day.
* **Closing Price:** The final price at which the stock was traded on a particular day.
* **Opening Price:** The price at which the stock opened on a particular day.
* **High Price:** The highest price reached during the trading day.
* **Low Price:** The lowest price reached during the trading day.
* **Volume:** The number of shares traded during the day.

**Preprocessing;**

Proper preprocessing of the dataset is crucial for ensuring that the model can effectively learn and predict the future prices of various stocks. In this phase, I prepared the dataset for modeling by transforming the raw NEPSE price data into a structured format that the LSTM model can effectively learn from. The following steps outline the preprocessing workflow:

1. **Feature Extraction:**

To enhance the predictive power of the model, several technical indicators were computed and added to the dataset. Firstly, the target variable was calculated by subtracting the opening price from the closing price for each day representing the daily price movement. To predict future movements, the **Target** feature was shifted up by one row, meaning each row in this column now shows the target value for the next day.

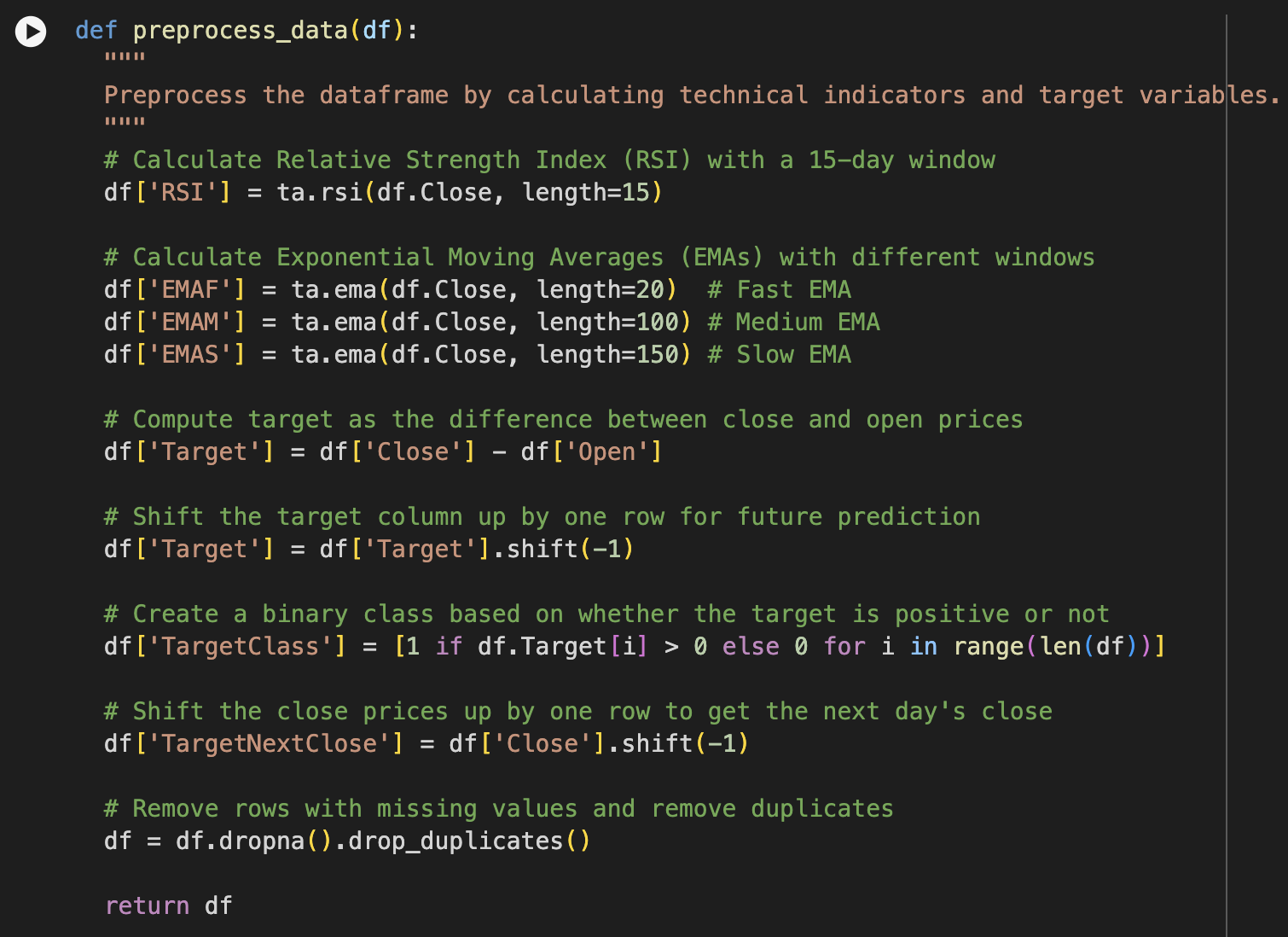
Similarly, a new binary column called **TargetClass** was created to indicate whether the stock price increased or decreased on the following day. If the **Target** value was greater than 0 (indicating a price increase), the **TargetClass** was set to 1; otherwise, it was set to 0. Finally, technical indicators like Relative Strength Index (**RSI**) and Exponential Moving Averages (**EMA**) were computed and added to the dataset. RSI indicates the momentum, signaling potential buy or sell opportunities based on recent price changes. EMAs give more weight to recent prices, making it more responsive to new information.

1. **Scaling and Sequence Generation:**

To ensure that the model’s training process is not biased by the scale of the features, all numerical (continuous) variables were scaled using Min-Max scaler. This normalization process transforms the features to a consistent range, typically between 0 and 1, which helps in improving the convergence speed and stability of the model during training. Finally, the sequence of inputs required by the LSTM model are generated using a function.

1. **Handling Missing Data:**

Missing values in the dataset were addressed using appropriate imputation techniques. For instance, missing price values were filled using forward-fill or backward-fill methods, depending on the nature of the missing data. Missing volume data, if any, was handled similarly to maintain the continuity of the dataset. It is a very crucial phase in the machine learning workflow to impute the missing values as it can introduce various discrepancies in the model.

***Figure 2****: Pre-processing Function*

# **MODELING**

**Model Architecture**

For this project, I decided to choose the Long Short-Term Memory (LSTM) model due to its effectiveness in capturing the temporal dependencies inherent in stock market data. LSTMs are particularly well-suited for time-series prediction as they can learn and remember long-term dependencies, which is crucial for predicting stock prices based on historical data.

The architecture of the LSTM model is as follows:

* **Input Layer:** The model receives input in the form of sequences, with each sequence consisting of 30 days’ worth of features. These features include historical price data and technical indicators, and the input shape is specified as (sequence\_length, 9), where sequence\_length is 30, and 9 represents the number of features.
* **LSTM Layer:** The model consists of 3 LSTM layers, where the first layer contains 256 units and is configured with return\_sequences=True, allowing the model to retain and process the full sequence of data. The second layer has 128 units and is also configured with return\_sequences=True to continue learning from the sequences.

Lastly, the third layer comprises 64 units and is configured with return\_sequences=False, providing the final sequence output for further processing.

* **Dense Layers:** Following the LSTM layer, the model includes Dense layers that refine the learned features and output the final prediction of the next day’s stock price. There are two dense layers and the first layer consists of 128 units with a **ReLU** activation function, which helps in refining the learned features from the LSTM layers. The subsequent Dense layer has 64 units, also using a **ReLU** activation function to further process the features and prepare them for the final prediction.
* **Output Layer:** The final Dense layer outputs a single value, which represents the predicted stock price for the next day.
* **Activation Layer:** An additional activation layer with tanh activation function is applied to the output to ensure that the final predictions are appropriately scaled.

**Hyperparameters**

The hyperparameters of the model were chosen based on experimentation and prior research and from a couple of references from the youtube videos. They are:

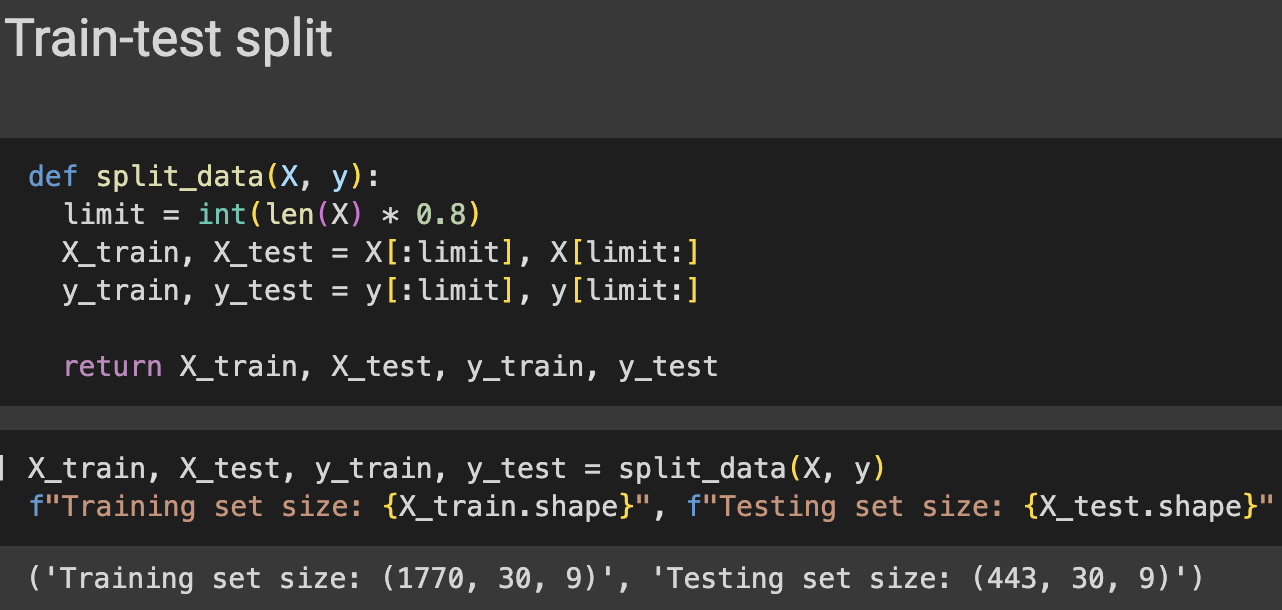
* **Sequence Length:** The model looks back at 30 days of historical data to predict the next day’s price.
* **Number of Epochs:** The model was trained for 50 epochs, using early stopping to prevent overfitting and ensure optimal performance.
* **Batch Size:** A batch size of 20 was used to balance computational efficiency and model performance.
* **Learning Rate:** A learning rate of 0.001 was used for optimizing the algorithm that determines the step size at each iteration while moving toward a minimum of a loss function.

**Training Process**

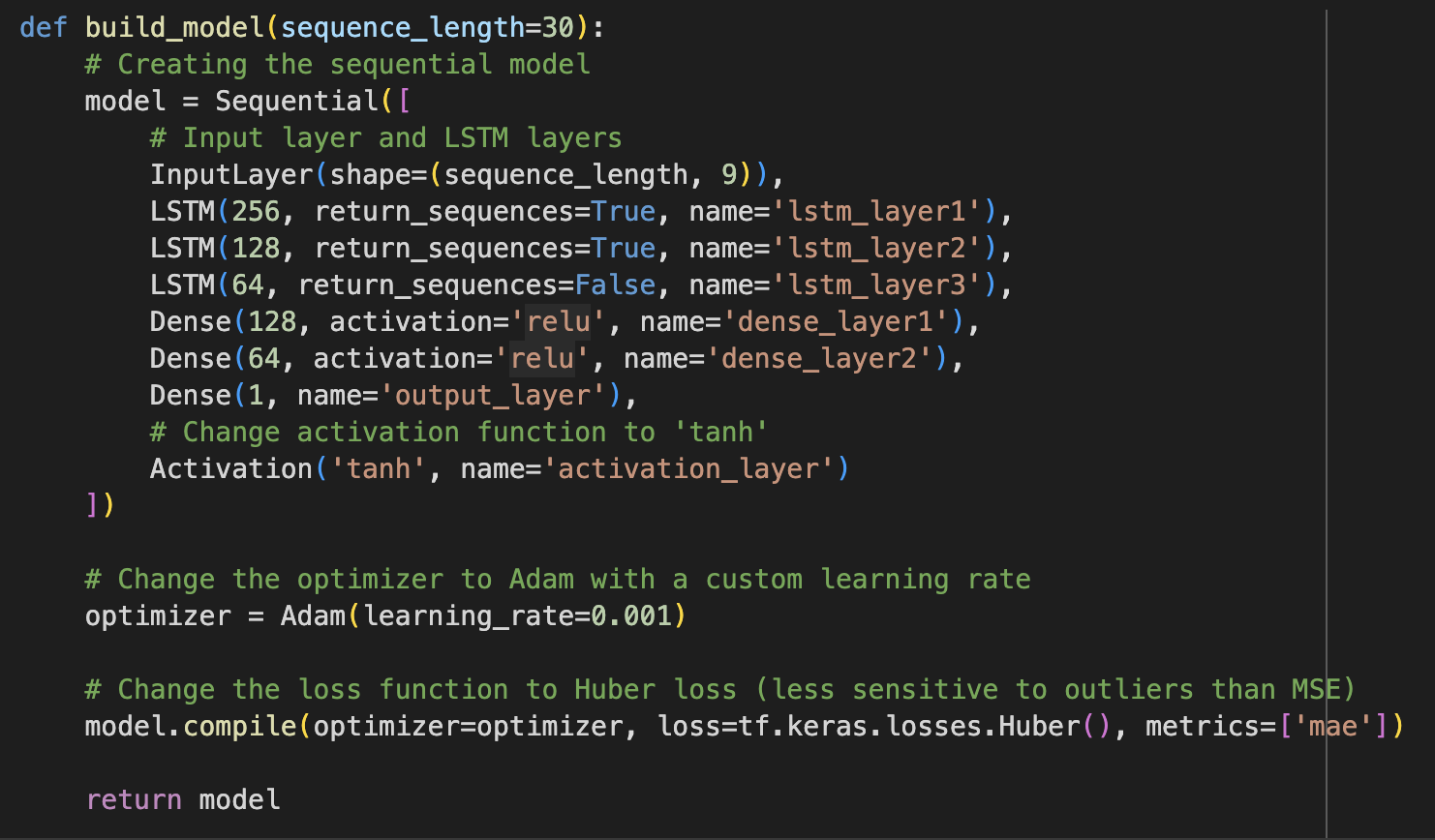
Before training the model on the processed dataset, we need to perform several steps, they are mentioned below:

* **Train-Test Split:** The overall dataset was split into two different parts of which the training dataset consisted of the 80% of the data and the rest 20% in the testing set. An additional validation set was also preserved from the 10% of the total training set for validating the model's learning.
* **Loss Function:** The model was trained using the **Huber** loss function from the tensorflow library that seemed to perform better for regression tasks like this when compared to other functions like **mean squared error** and **mean absolute error**.
* **Optimization Technique:** The **Adam** optimizer was employed to minimize the loss function. It is a popular choice due to its adaptive learning rate and ability to handle sparse gradients, making it well-suited for training deep learning models like LSTM.

During the training time, the model iteratively learned to minimize the prediction error on the training set, with the ultimate goal of generalizing well to unseen data.



***Figure 3****: Train-Test Split*



***Figure 4****: LSTM Model Architecture*

# **RESULTS**

**Model Performance:**

The performance of the LSTM model was evaluated by comparing the predicted stock prices with the actual prices on a test dataset. Below are some of the key results:

***Table 1:*** *Actual vs. Predicted Comparison Table*

| **Actual Price** | **Predicted Price (Model)** |
| --- | --- |
| 295.0 | 316.98 |
| 302.0 | 335.95 |
| 297.0 | 313.79 |
| 297.3 | 294.58 |
| 298.0 | 288.08 |
| 303.0 | 287.82 |
| 300.0 | 289.8 |
| 298.8 | 293.48 |
| 311.9 | 295.96 |
| 216.5 | 295.95 |

Moreover, the model achieved the following performance metrics:

* **Mean Absolute Error (MAE):** 0.00895
* **Mean Squared Error (MSE):** 0.00014
* **Root Mean Squared Error (RMSE):** 0.01190
* **R2 Score:** 0.9727

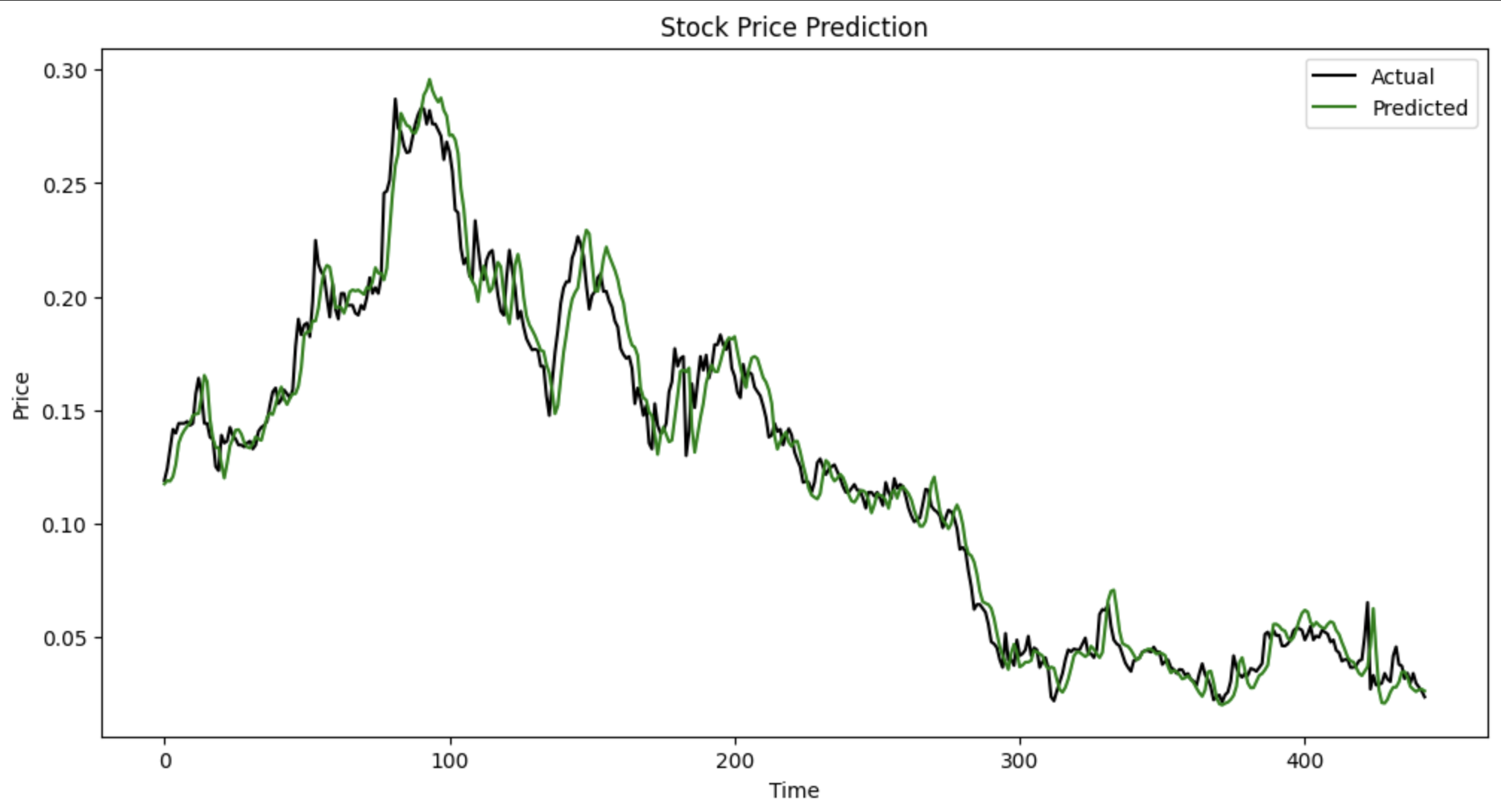
These metrics indicate a decent performance, with the model showing good accuracy in predicting the stock prices, as reflected in the low error values and a high R2 Score.

**Visualizations**

To better understand the model’s performance, several visualizations were plotted to compare the predicted values and the actual prices.

1. **Predicted vs. Actual Prices:**

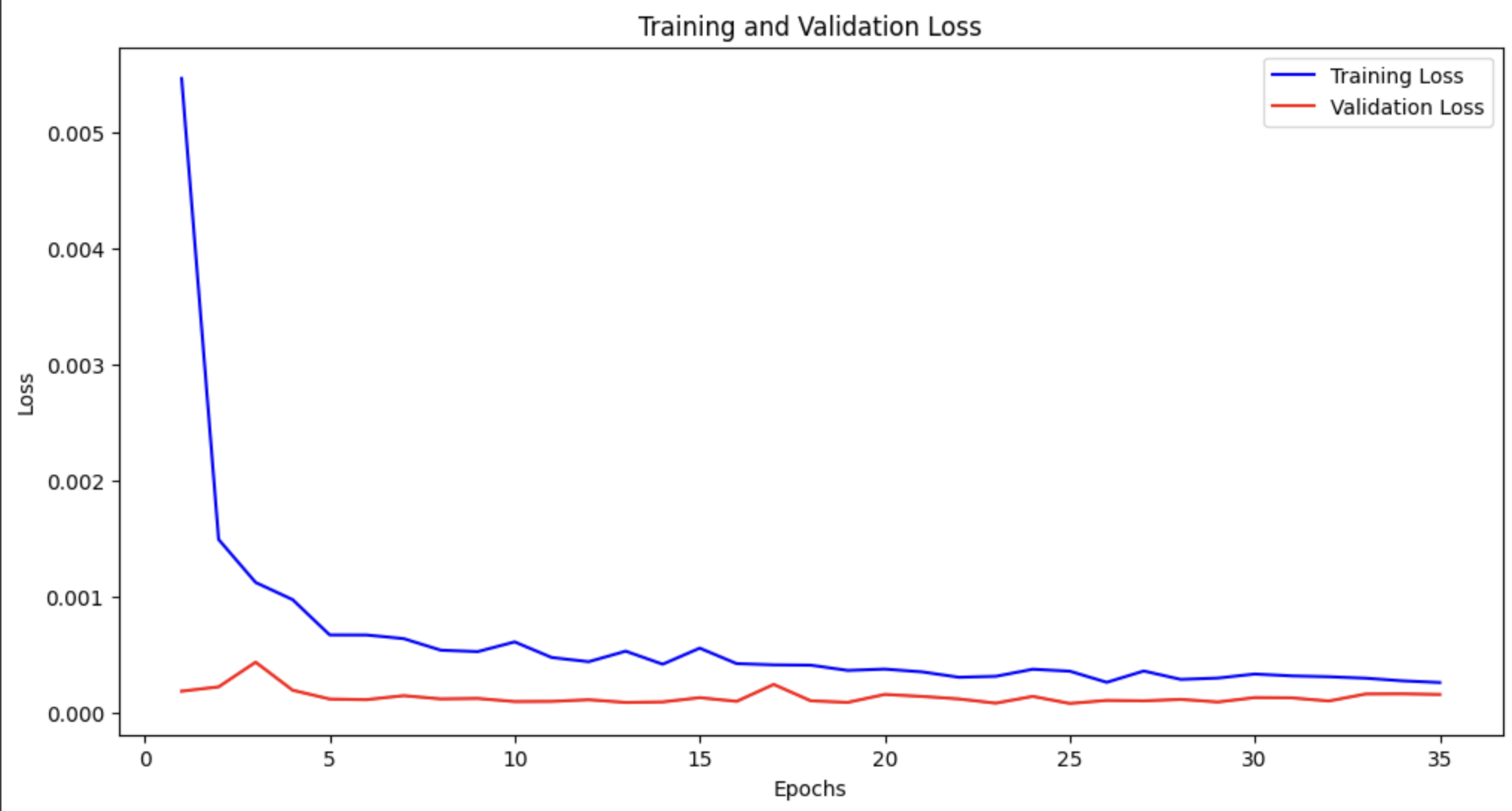
A time-series plot comparing the predicted stock prices to the actual prices over time was constructed which helps in highlighting the model’s ability to closely follow the actual price trends. It can be observed in the below visualization that the model is closely able to identify the trends and price movements for a particular stock with some small discrepancies due to inherent factors that affect the market's stability.



***Figure 5****: Predicted vs. Actual Price Plot*

1. **Loss Curve:**

A plot of the training and validation loss over epochs was also visualized to gain insights into the model’s learning process and identify any signs of overfitting or underfitting. There can be observed some spikes in the validation loss while training which could be due to several factors that the model is not being able to effectively learn.



***Figure 6:*** *Training and Validation Loss Curves*

These visualizations better our understanding by offering a clear picture of its predictive capabilities and areas for improvements.

**Discussion**

The result demonstrates that the LSTM model was effective in capturing the temporal dependencies in the stock market data. The low Mean Absolute Error (**MAE**) and Root Mean Squared Error (**RMSE**) values indicate that the model was able to predict the stock prices with a higher degree of accuracy. The high **R2** score of 0.9727 further supports that the model explained a significant portion of the variance in the stock prices.

However, there were some instances where the predicted values deviated significantly from the actual prices. This could be due to several factors, such as the inherent volatility of the stock market, limitations in the dataset, or the need for further hyperparameter tuning. Despite these outliers, the overall performance was strong, suggesting that the model generalized well to unseen data.

There was no significant evidence of overfitting or underfitting, as the model performed consistently well on both the training and validation datasets.

# **CONCLUSION**

**Summary of Findings:**

The LSTM model demonstrated a robust ability to predict stock prices with considerable accuracy. The performance metrics, including a Mean Absolute Error (**MAE**) of 0.00895, Mean Squared Error (**MSE**) of 0.00014, and Root Mean Squared Error (**RMSE**) of 0.01190, highlight the model's precision in forecasting stock price movements. Additionally, the high **R2** score of 0.9727 indicates that the model successfully captured and explained a substantial portion of the variance in the stock market data. The results suggest that the LSTM model effectively learned the temporal dependencies within the data, translating into reliable predictions.

Despite these promising outcomes, some deviations between predicted and actual prices were observed. These discrepancies could be attributed to factors such as stock market volatility, dataset limitations, or the need for further hyperparameter optimization. Nonetheless, the overall performance was consistent and robust, with no significant evidence of overfitting or underfitting, as the model performed well across both training and validation datasets.

**Future Recommendations:**

1. **Data Enhancements:**

We can enhance the model by incorporating more features, such as trading volume, market sentiment indicators, or macroeconomic factors, could provide a more comprehensive view and potentially improve prediction accuracy. We can expand the dataset with more diverse and extensive historical data that could help the model generalize better and handle anomalies or rare events more effectively.

1. **Hyperparameter Tuning:**

We can further tune and experiment with different architectures, activation functions, or learning rates (hyperparameters) that could refine the model’s performance.

1. **Handling Volatility:**

Developing methods to better account for stock market volatility could enhance the model's robustness in predicting extreme or unexpected market movements.

1. **Model Comparison:**

Comparing the LSTM model with other advanced models, such as Transformer-based architectures or hybrid models, could provide insights into alternative approaches that may yield improved performance.

By addressing these areas, future work could further advance the predictive capabilities of the model and provide more accurate and reliable stock price forecasts.

# **REFERENCES**

Coyne, S., Madiraju, P., & Coelho, J. (2017). Forecasting Stock Prices using Social Media Analysis. In *Proceedings of the 2017 IEEE International Conference on Big Data* (pp. 169-174).

DOI: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.169

Borty, P. C., Sani, U., Rahman, R., & Majumdar, A. (2017). *Tweets Sentiment Analysis Using Machine Learning Techniques*. BRAC University.

Shahi, T. B., Shrestha, A., Neupane, A., & Guo, W. (2020). Stock price forecasting with deep learning: A comparative study. *Mathematics, 8*(9), 1441.<https://doi.org/10.3390/math8091441>