Stock Market Prediction Using LSTM Deep Learning Model: A Case Study on Shopify Inc.

Harnessing Deep Learning for Financial Forecasting

Seyi Falope

Shopify Inc.: Revolutionizing E-Commerce

Foundation and Growth: Established in 2006, originally as an online store, Shopify rapidly transformed into a leading e-commerce platform. This evolution highlights the company's adaptability and deep understanding of the e-commerce landscape.

Business Model: Shopify's business model revolves around providing a comprehensive suite of tools for entrepreneurs and businesses. These tools facilitate various aspects of online retail, including website creation, payment processing, marketing, and shipping. This model has been pivotal in democratizing e-commerce, allowing even small retailers to compete effectively in the global market.

Market Impact
Positioning: A major player in the e-commerce sector, Shopify rivals giants like Amazon and eBay. It is celebrated for its user-friendly design and extensive customization capabilities.

Growth Trajectory: Notable for its rapidly expanding merchant base, especially post-COVID-19, Shopify's revenue growth and diversification into new markets signify its strong industry standing.

Stock Performance: Shopify's stock has witnessed fluctuations, mirroring its adaptability to market shifts, technological advancements, and strategic choices. The COVID-19 era marked a significant growth phase for the company.

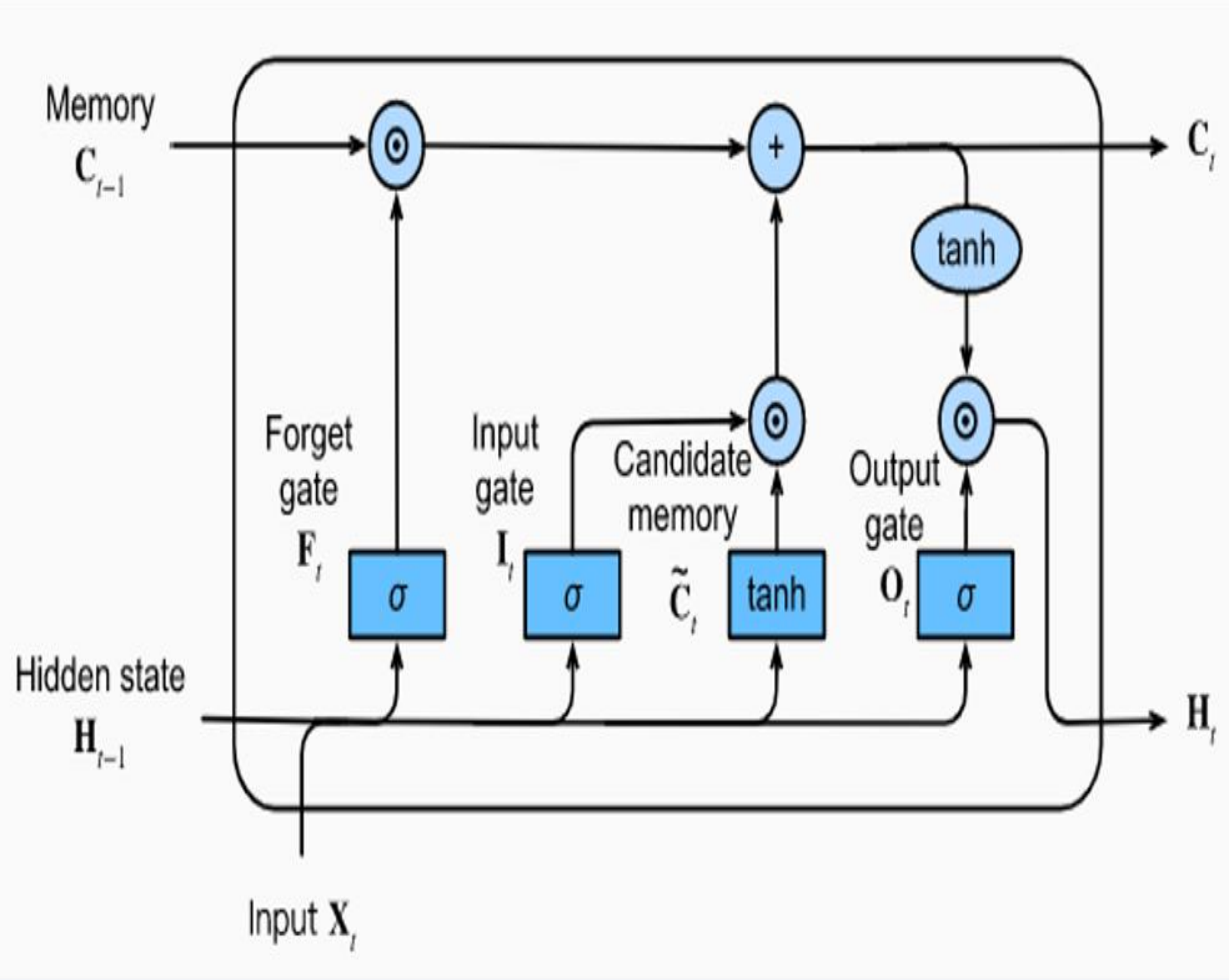
Investment Appeal: Shopify's dynamic stock, responsive to market trends and consumer behaviors, offers a valuable case for predictive analysis in e-commerce.

LSTM and Suitability for Stock Data

Long Short-Term Memory (LSTM) networks, a variant of Recurrent Neural Network (RNN), excel in processing time series data such as stock prices. Their ability to remember long-term dependencies and patterns in data renders them exceptionally suitable for stock market forecasting.

In the realm of financial predictions, deep learning, with its capacity to handle large and intricate datasets, provides more refined and precise insights than conventional models. The unique structure of LSTM networks allows them to capture the dynamic and often non-linear patterns in stock market data, making them a powerful tool for predicting stock trends and movements.

LSTM NETWORK



Data Acquisition, Preprocessing and Feature Engineering

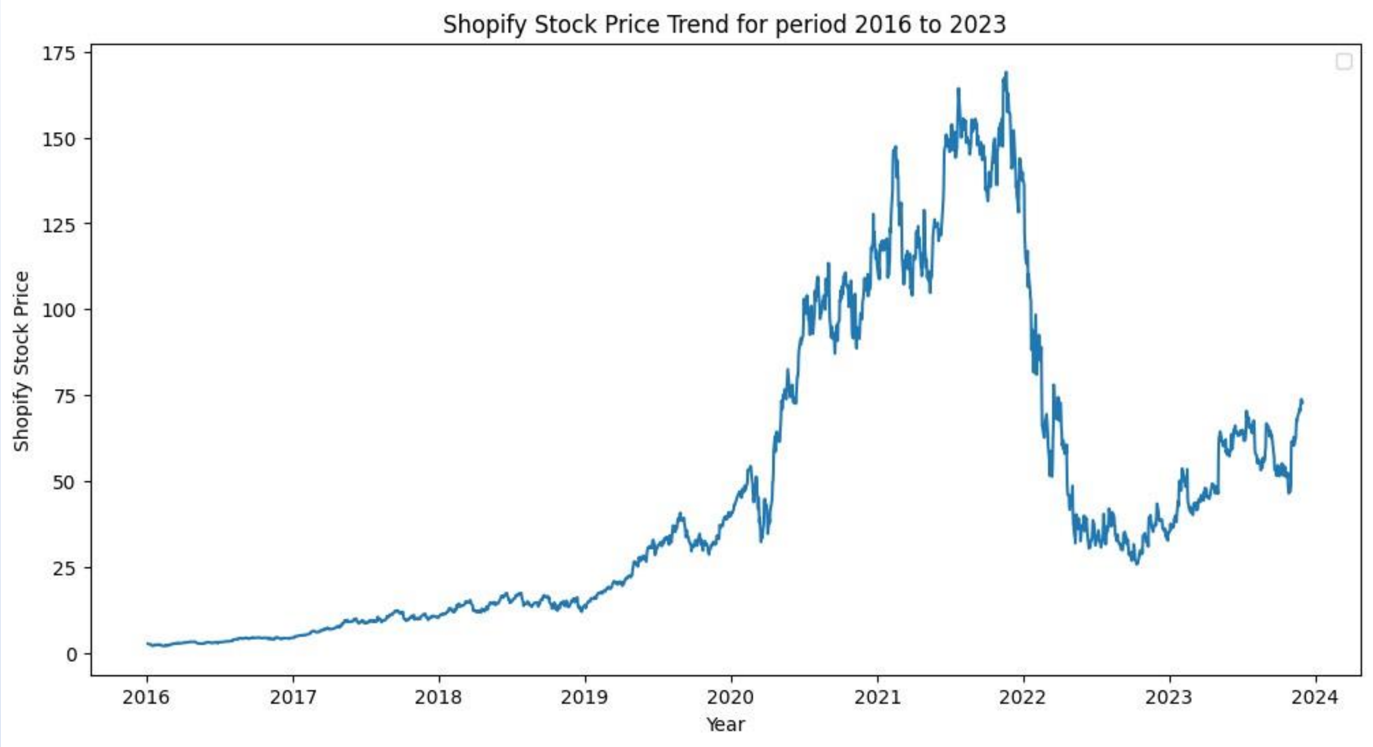
The data for Shopify's stock was sourced from Yahoo! Finance from the yfinance API.

Min Max Scaler: In the data preprocessing phase, a rigorous set of procedures was executed to ensure the data was ready for LSTM model training. One critical step was the normalization of feature values to guarantee uniformity. To accomplish this, the MinMax Scaler was employed, effectively transforming the features into a standardized range between 0 and 1. This normalization is crucial for LSTM models, as it helps prevent certain features from dominating others and ensures the model can learn effectively from all input variables.

Feature Engineering: Furthermore, feature engineering techniques were strategically applied to enhance the dataset's richness and provide valuable insights for subsequent analysis. Specifically, two new features, 'year' and 'month', were ingeniously crafted from the existing data. These new features extract temporal information, such as the year and month, from timestamp or date-related fields. Incorporating 'year' and 'month' into the dataset enables deeper exploration by allowing for the detection of yearly and monthly patterns or seasonality. This groundwork not only ensures the model's compatibility with the data but also sets the stage for more insightful exploratory data analysis and ultimately enhances the LSTM model's capability to uncover and exploit temporal patterns within the

Exploratory Data Analysis (EDA)

Before diving into predictive modelling, an exploratory data analysis was conducted. This step is crucial to understand underlying patterns, anomalies, or trends in the stock data. Various visualizations were utilized, such as line plots to observe the stock price trends over time, and box plots to examine the distribution and identify any outliers.



Line Trend for Shopify Stock Price:

This graph on the right depicts the stock price trend of Shopify from 2016 to 2023. It shows a significant increase in stock price up until around 2021, where it reaches a peak. After 2021, there's a sharp decline followed by a period of volatility and a modest upward trend towards the end of the period.

	Close
count	1992.000000
mean	46.480091
std	44.088219
min	1.933000
25%	11.471750
50%	32.701000
75%	65.805002
max	169.059998

Descriptive Statistics:

The image shows statistical summary information for a dataset of closing stock prices. It indicates there are 1,992 data points, with an average (mean) closing price of approximately 46.48. The data varies with a standard deviation of about 44.09, with the lowest (min) close at around 1.93 and the highest (max) close at approximately 169.06.

Exploratory Data Analysis (EDA) Cont'd

The histogram illustrates the distribution of Shopify's closing stock prices in dollars from 2016 to 2023. Here's a summary of the key points:

- **Most Common Price Range:** Shopify's stock most frequently closed between $0 to $25, indicating this was a common price range during the period.

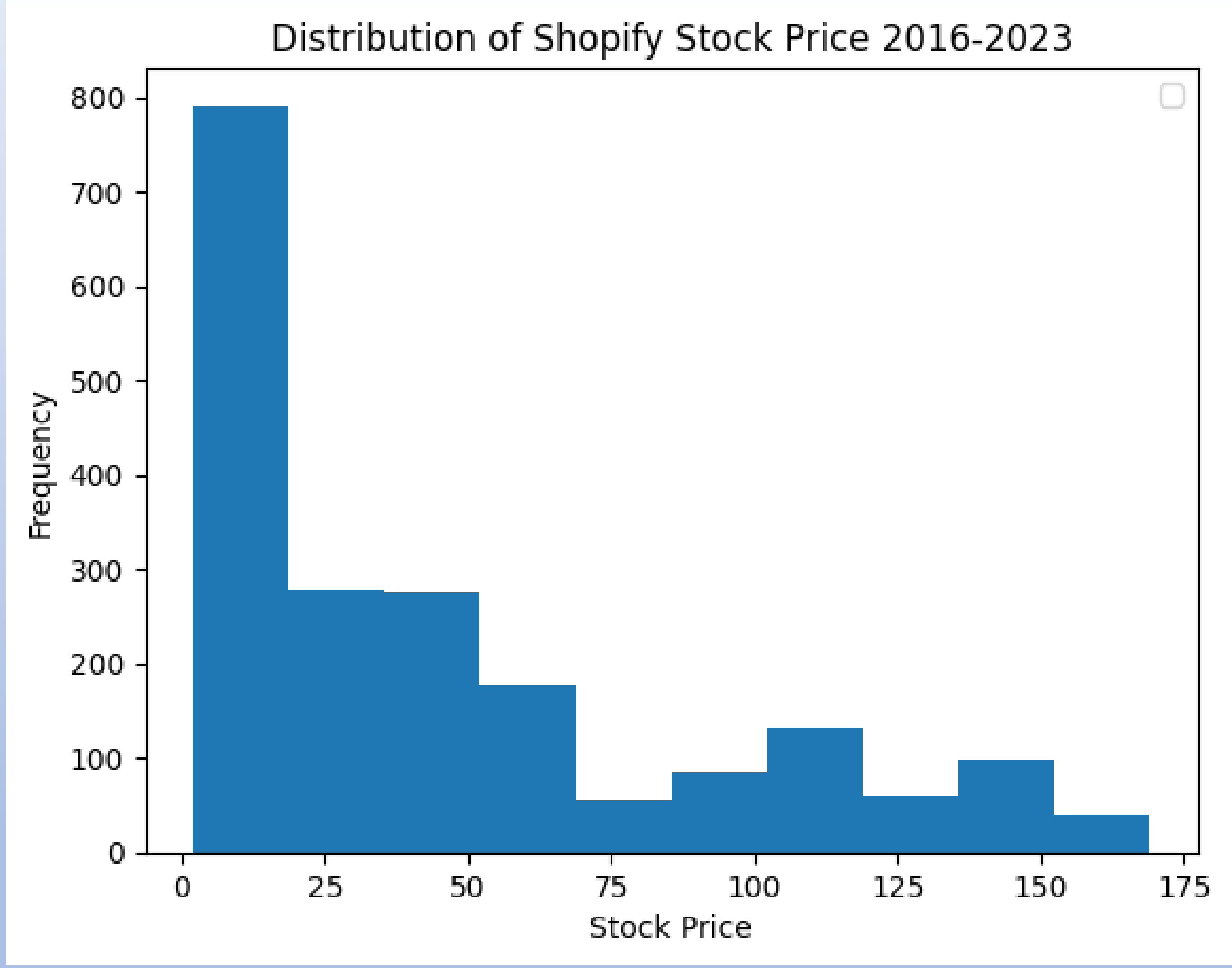
- **Right-Skewed Distribution:** The histogram shows a right-skewed distribution, meaning that while the stock commonly closed at lower prices, there were occasions when it closed significantly higher.

- **Decrease in Higher Price Frequency**: As the stock price increases, the frequency of these closing prices decreases, suggesting that higher closing prices were less common.

- **Occasional High Closing Prices:** There are a few instances where the stock closed at much higher prices, possibly reflecting periods of strong company performance or favorable market conditions.

- **Volatility and Growth:** The spread of closing prices from $0 to $175, along with the changes in frequency, suggests that Shopify's stock has experienced both volatility and growth over the seven-year period.

This pattern could reflect the company's growth trajectory, with initial lower closing prices in its earlier days and a general trend toward higher prices as the company matured, punctuated by periods of significant highs and lows.



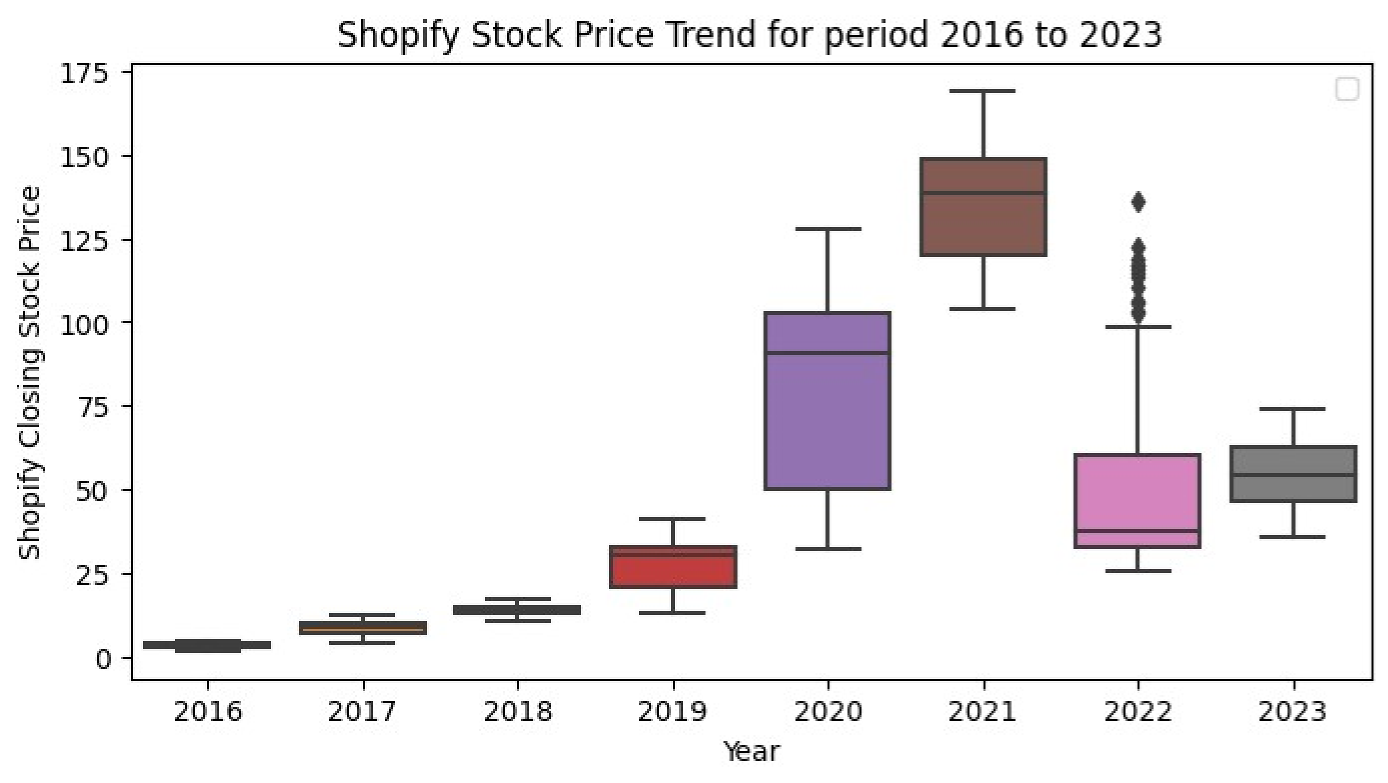
Exploratory Data Analysis (EDA) Cont'd

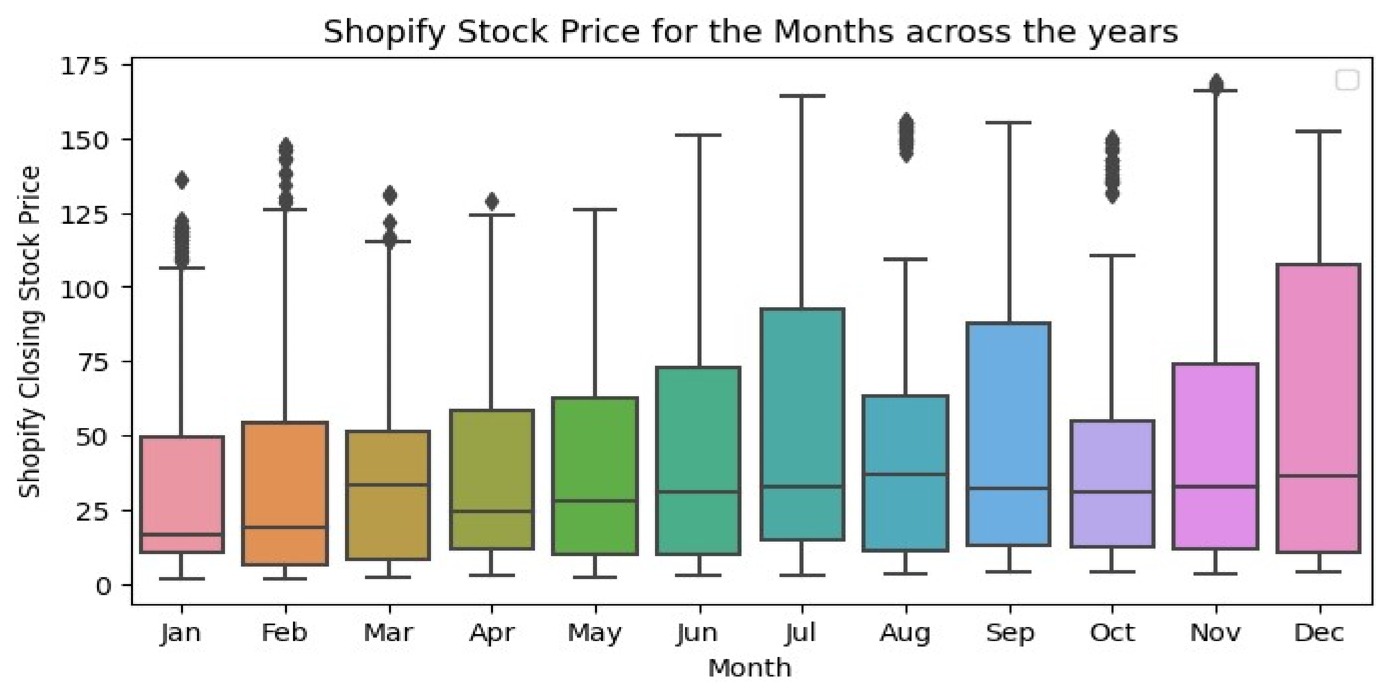
- **<u>Box Plot for Shopify Stock Price across the years:</u>**

The plot on the right is a box plot showing the annual Shopify's closing stock prices from 2016 to 2023. There's a noticeable increase in median stock price and variability in 2020 and 2021, potentially due to more people shopping online through Shopify during the COVID-19 pandemic. The stock price demonstrates some volatility, with a particularly wide range in 2021, and appears to stabilize somewhat in 2023.

- **Box Plot for Shopify Stock Price across the month across the years:**

The box plot illustrates Shopify's monthly closing stock prices, with each box representing the middle 50% of values (interquartile range) for that month over multiple years. Averagely the month of March and August has the highest stock price. The month of January has the lowest average stock price trend.





LSTM Model Architecture in Keras

The model employed is a sequential neural network consisting of three LSTM (Long Short-Term Memory) layers followed by a dense layer. The first LSTM layer has an output shape of 30 units, the second LSTM layer expands this to 60 units, and the third LSTM layer consolidates to 60 units without any additional dimension.

Return sequence was set for the first 2 LSTM layers and the Adams optimizer was employed .The dense layer, with an output shape of 1 unit, was used as this is regression task . The loss function of Mean Squared Error was employed and the Mean Absolute Error metric was used to provide a clear measure of predictive accuracy .The model has a total of 54,781 parameters, all of which are trainable.

Model Architecture

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 30, 30)            3840

 lstm_1 (LSTM)               (None, 30, 60)            21840

 lstm_2 (LSTM)               (None, 60)                29040

 dense (Dense)               (None, 1)                 61

=================================================================
Total params: 54781 (213.99 KB)
Trainable params: 54781 (213.99 KB)
Non-trainable params: 0 (0.00 Byte)
```

Model Training and Parameter Tuning

1 Data Splitting

The stock market dataset was divided into training and testing sets. Typically, a larger portion is used for training (80%) and a smaller portion for testing (20%). This split ensures the model is trained on a substantial amount of data while reserving enough unseen data for evaluation.

2 Normalization

Given the wide range of values in financial data, Min Max normalization was applied to scale the data, typically ensuring all values lie between 0 and 1. This step helps in accelerating the learning process and improving convergence.

3 Sequence Creation

For LSTM models, the data needs to be structured into sequences. Each input sequence contains data points from previous days (the past 30 days), and the model predicts the stock price for the next day.

4 Layer and Neuron Selection

The model's architecture, including the number of layers and neurons in each layer, was iteratively adjusted. Ultimately, three deep layers were used, and each of the layers had 30, 60, 60 respectively. Finding the right balance is crucial to avoid overfitting (too complex) or underfitting (too simple).

5 Optimization

Optimization technique like Adam was used to optimize the weight and parameter of the model , the batch size of 20 was utilized and 75 epochs were used.

Model Prediction and Evaluation

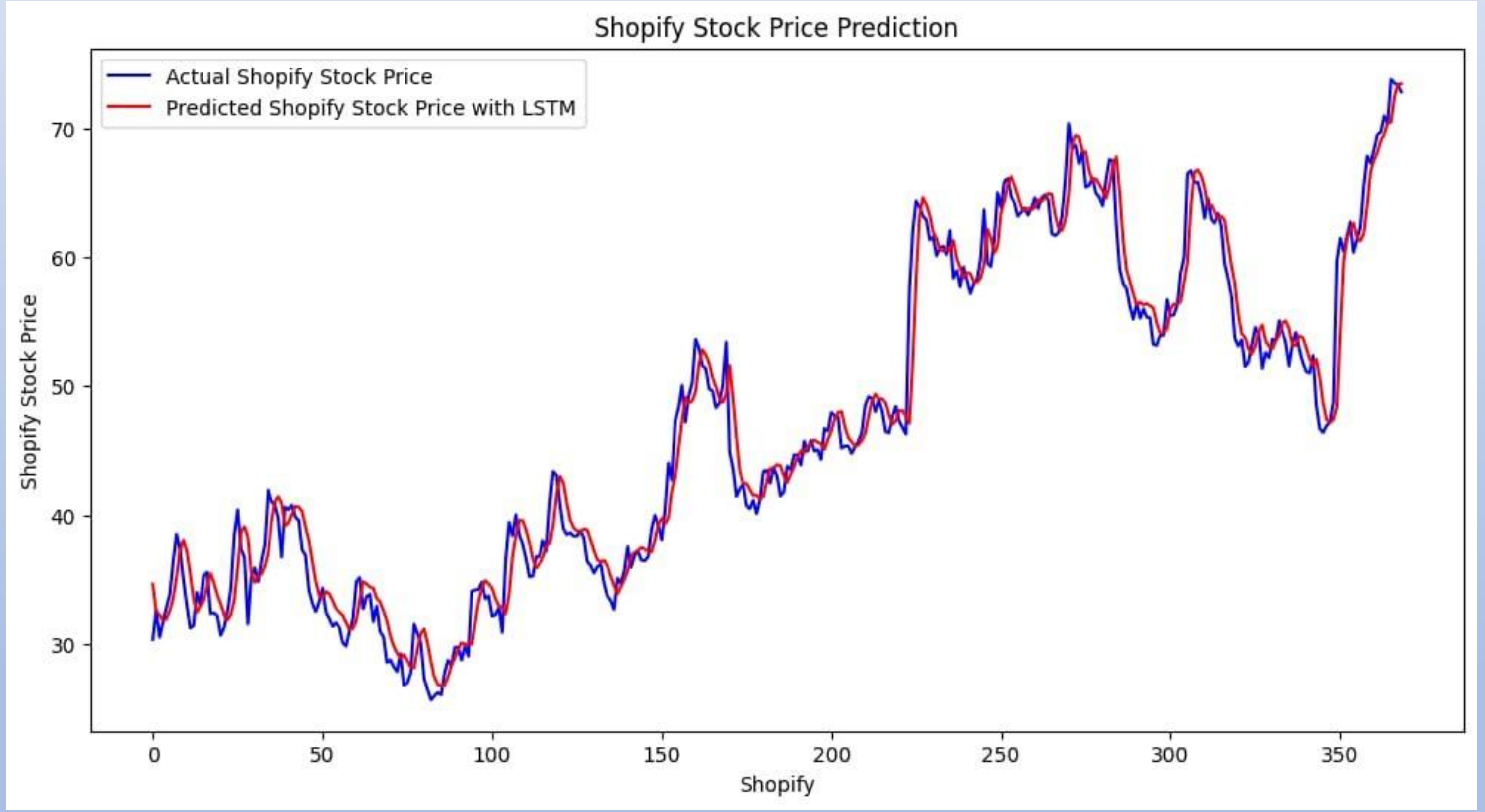
Evaluation Metrics

1 Evaluation Measures

The evaluation of the model's predictions was assessed by comparing the forecasted stock prices to the actual market prices. To quantify this, measurement criteria such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared were employed.

2 Performance on Test Data

The model's effectiveness is gauged by its test set results, indicating how well it can adapt to new, unseen data The visualization compares and shows the actual price trend the and the predicted price trend by the model . The Root Mean Square Error stands at 2.28, the Mean Absolute Error at 1.63, shows the average difference between the actual and predicted stock value and the R-squared value is 0.97, reflecting its predictive accuracy and how well the independence variable explained the dependent variable.



Forecasting Shopify's Stock Price for the next 60days

Upon completing the prediction and evaluation phases, we employed our model to project Shopify's stock price for the forthcoming 60-day period.

The outcome of this forecast, as depicted in the provided visual, indicates a modest upward trend for the stock value.

Conclusion

In conclusion, the LSTM model developed for predicting Shopify stock prices shows promise with a low Mean Absolute Error of 1.63, a low Root Mean Squared Error of 2.28, a high R-squared value of 0.97 and a detailed architecture. However, it faces limitations such as potential overfitting due to its complexity and the risk of not capturing long-term trends with a 30-day input sequence. Furthermore, reliance on Min Max normalization may not be ideal for financial data's volatility.

For improvement, regularization techniques such as Drop-out and hyperparameter optimization such as varying batch size and learning rate could be implemented to mitigate overfitting and potentially boost the model’s predictive accuracy. Exploring dynamic normalization and extending the sequence length might enhance the model's ability to generalize and capture more complex patterns. Finally, including additional performance metrics and data sources could provide a more nuanced evaluation and potentially boost the model's predictive accuracy.

Reference

Wei, Z., Yang, S., Yang, N., et al. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). https://ieeexplore.ieee.org/document/8078738

Weerasinghe, J., Mohamed, F., et al. (2019). Stock Price Prediction Using Deep Learning and XGBoost. 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE). https://ieeexplore.ieee.org/document/8941944

Nieto, M., Bonadiman, D., et al. (2017). Deep learning for stock prediction using numerical and textual information. Expert Systems with Applications, 69, 2017. https://www.sciencedirect.com/science/article/pii/S0957417416307172

Raja, M. A., AlZoubi, I., et al. (2015). A hybrid stock trading framework integrating technical analysis with machine learning techniques. Expert Systems with Applications, 42(2), 2015. https://www.sciencedirect.com/science/article/pii/S0957417414004455

Agrawal, R. K., Singh, V. P., et al. (2016). A comprehensive review on stock market prediction using artificial neural networks. Expert Systems with Applications, 55, 2016. https://www.sciencedirect.com/science/article/pii/S0957417416301909


THANK YOU!