European Economic Letters
ISSN 2323-5233
Vol 14, Issue 2 (2024)
http://eelet.org.uk

Forecasting Trends in Stock Prices Using Transformer Networks

Duggirala Uma
Research Scholar
Presidency University, Bengaluru

Dr. Sunil M Rashinkar
Associate Professor
School of Management, Presidency University, Bengaluru

Abstract:

Financial time series forecasting is undoubtedly the top choice of computational intelligence for finance researchers in both academia and the finance industry due to its broad implementation areas and substantial impact. The ever-changing tides of the stock market present a significant challenge for investors and financial institutions seeking to make informed decisions. Predicting future stock prices with accuracy remains a highly sought-after capability, offering the potential to optimize investment strategies and navigate market volatility. Traditional forecasting methods have limitations, often struggling to capture the complex dynamics and non-linear relationships inherent in stock price data. This project delves into the potential of deep learning techniques for stock price forecasting. We explore the application of a Transformer-based model, a powerful deep learning architecture known for its ability to learn long-term dependencies in sequential data. By leveraging this model's capabilities, we aim to develop a more robust and accurate approach to forecasting stock prices.

Keywords: Deep learning, financial time series, National stock Exchange, Stock price prediction, Transformers-based model

Introduction

In the financial sector, artificial intelligence (AI) is reshaping market dynamics through cutting edge machine learning and deep learning algorithms. These techniques are widely used for predicting financial instrument prices, analyzing market trends, optimizing portfolios and identifying investment opportunities. (Guo, Stock Price Prediction Using Machine Learning, 2022). The practical applications of machine learning in finance cover supervised and unsupervised algorithms, ensemble methods, time series analysis and deep learning methods. The machine learning models like Linear Regression, Logistic Regression, Support Vector Machine, Random Forest, K Nearest Neighbor and Deep learning models like Auto Encoders, Deep Belief Networks, Multilayer Perceptron, Convolution Neural Networks, Self-Organizing Maps are used in forecasting stock prices. (Wang, 2022) Each of these techniques have been increasingly used in forecasting stock prices due to their ability to capture complex patterns in data. ML and DL models for stock price forecasting often rely on a wide range of features beyond just historical price data. (A. Vaswani, 2017). These features can include technical indicators, fundamental data and macro-economic indicators which include Model Selection, Time Series Analysis, Ensemble Methods, Model evaluation, Data Preprocessing etc., ML and DL techniques have significantly advanced the field of stock price forecasting, enabling traders, investors, and financial institutions to make more informed decision based on data driven insights. (George E.P. Box, 2015) However, it is essential to recognize the inherent uncertainties and risks associated with predicting stock prices, and to use forecasting models as one of many tools in the decision-making process.

Problem Definition:

Forecasting stock prices remains a challenging task despite the plethora of traditional and computational techniques available. Some of the key challenges of forecasting stock prices are the complexity of financial markets, noisy data, non-stationarity, randomness and uncertainty. (Sarode, 2015). Apart from that the correlation between the influencing factors with the stock prices may not imply causation. Distinguishing between spurious correlations and meaningful causal relationships is a significant challenge. Addressing these challenges requires a combination of advanced modelling techniques, domain expertise, careful feature engineering and robust validation procedures. It is also important to recognize the limitations of any forecasting model and to incorporate human judgement and qualitative analysis into the decision-making process.

Objectives:

- To evaluate the performance of each model in terms of predictive accuracy, computational efficiency, and ability to capture temporal dependencies in stock data.
- To compare the model performance based on the interval and the technical indicator moving average of the data.
- To develop a custom transformer based neural network tailored to the specific requirements of stock price prediction.
- To design the architecture to handle sequential data effectively, incorporating features like attention mechanisms and positional encodings.
- To use appropriate evaluation metrics for assessing the predictive accuracy through Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and directional accuracy.

Review of literature:

Research within the field of finance has evolved to encompass the complex interplay between stock markets and their diverse array of data sources. This heterogeneity refers to the inclusion of data from various channels, spanning stock markets, foreign exchange markets, weather systems, and beyond. This data encompasses fundamental elements such as stock prices and trading volumes, as well as unstructured data like stock news and social network sentiments (Guo, Stock Price Prediction Using Machine Learning, 2022).

The rapid advancement of science and technology has facilitated the accumulation of vast volumes of financial data, providing a robust foundation for stock market analysis. However, accurately forecasting future stock prices remains a significant challenge. Traditional statistical models often struggle with the inherent complexities of time series data, particularly non-linear relationships and long-term dependencies (Saikat Mondal, 2020).

To address these limitations, researchers have explored the application of machine learning and deep learning algorithms**.** Machine learning models offer the advantage of learning non-linear patterns from historical data, leading to potentially improved predictions compared to traditional statistical models. Deep learning techniques, such as Artificial Neural Networks and Convolutional Neural Networks, have emerged as powerful alternatives, capable of providing even higher levels of accuracy (Zhao, 2023; Schmidhuber, 1997, 2015).

Depending on the prediction horizon, various input parameters are selected, ranging from high-frequency trading data to daily, weekly, or monthly stock prices. Studies often employ multiple deep learning models for performance comparison, utilizing raw price data or novel approaches like Deep and Wide Neural Networks (DWNN) (Adityanarayanan Radhakrishnan, 2023). Transformer-based models are gaining traction for stock price prediction, having been applied to forecast S&P volatility and analyze social media data for forecasting purposes. However, research combining news and technical data for stock price forecasting, particularly using Transformers, remains relatively unexplored.

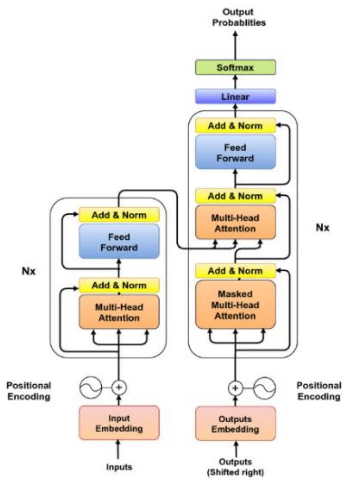
Several recent studies have explored the potential of Transformers for stock price forecasting. (J.Liu, 2019) and (Saidul Islam, 2023) demonstrated the effectiveness of Transformers for capturing information from investment texts and social media data, achieving accuracies of 64% and 96% respectively. Daiya and Lin proposed TRANSDICE, (Divy Daiya, 2021) a model combining Transformers with specialized convolutions to extract features from financial indicators and news, achieving a 3% improvement over existing methods.

Research has also explored Transformers for stock price direction prediction. (Qianggang Ding et al., 2020) proposed a Hierarchical Multi-Scale Gaussian Transformer achieving accuracies of 58% and 57% on Chinese and US stock data. (Eduardo Ramos Perez, 2021) investigated hybrid models combining Transformers with other algorithms, demonstrating effective volatility estimation on the S&P dataset. (Bryan Lim Sercan O Arik, 2021) introduced the Temporal Fusion Transformer for multi-horizon time series forecasting, achieving strong results on complex datasets. Finally, (Yawei Li, 2022) proposed the Transformer Encoder Attention architecture for predicting asset price movements based on historical data and sentiment analysis, achieving a 64% accuracy on two of their datasets. These studies showcase the growing interest and promising results of applying Transformers in various aspects of stock price forecasting.

Transformer and its architecture: Transformers are a powerful neural network architecture that revolutionized Natural Language Processing (NLP) tasks. Unlike LSTMs, which rely on recurrent connections, Transformers primarily use an attention mechanism to learn relationships between elements within a sequence. This allows them to excel at tasks like machine translation, text summarization, and question answering, often achieving state-of-the-art performance. Transformers typically follow an encoder-decoder structure:

- **Encoder:** This part of the network processes the input sequence (e.g., a sentence in machine translation). It breaks down the sequence into individual elements (words) and captures their relationships using the attention mechanism.

- **Decoder:** This part of the network utilizes the encoded representation from the encoder to generate the output sequence (e.g., the translated sentence). It also employs the attention mechanism to focus on relevant parts of the encoded representation during generation.



Transformer Architecture

Attention Mechanism:

The attention mechanism is the heart of a transformer. It allows the model to focus on important parts that is, instead of processing the entire sequence element by element, the attention mechanism lets the model focus on the most relevant parts of the input sequence for each element. Unlike LSTMs, which rely on sequential processing, the attention mechanism can capture relationships between any two elements in the sequence, regardless of their distance. This is crucial for tasks like machine translation, where word order can be significantly different between languages.

Transformer Layers:

Both the encoder and decoder in a transformer consist of multiple identical layers. Each layer typically includes:

- **Multi-head Attention:** This is a scaled version of the attention mechanism that allows the model to attend to different aspects of the input simultaneously.

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 2 (2024)
http://eelet.org.uk

- **Positional Encoding:** Since transformers lack recurrent connections, positional encoding is added to the input to convey the order of elements within the sequence.
- **Feed Forward Network:** This is a simple neural network layer that further processes the information from the attention mechanism.

Proposed methodology:

This article focuses on analyzing the Transformer model while predicting the stock price movements of a stock ASHOKLEY.NS from NSE downloaded from Yahoo Finance. This stock is selected as its beta value is 1.04 high, that is greater than 1. Specifically, considered six types of data sets with different intervals One day, 5 Days moving average for a period of 5 years, and 10 years and maximum duration. Single stock is selected so that the accuracy metrics can be compared easily. The data is collected from 2019 till date and has 1237 rows for 5 year data daily data set, 2014 till date for 10 year data and has 2459 rows and 2004 till date and has 5463 rows for max Year daily data The data is taken from yahoo finance web site.

The Transformer model for each of the duration and the interval, a total of 6 models were chosen for analysis. Each model was trained on 70% of the total data and was tested on 30% data. The Transformer model was built with multiple transformer encoder blocks, attention mechanisms, and a feed-forward network to capture complex relationships in stock price data. The model is then compiled with an optimizer, loss function, and metrics for training and evaluation. Dropout layers are used for regularization, and a final dense layer predicts the closing price were used. With head size 128, num_heads 4, ff_dim 2, num_trans_blocks as 4, mlp_units 256, dropout 0.10 and mlp_dropout as 0.10 and early stopping patience as 5 the model has been developed.

Model Validation: To validate the models, six performance metrics, Test loss, Root mean squared Error, Mean Absolute Error, Mean Absolute Percentage Error , R-Square and Directional Accuracy were used to test the training and testing data sets.

Results and discussions:

In this section we discuss the performance of Transformers on three data sets namely five year daily, 10 year daily, max daily and 5y 5 day moving average, 10 year 5 day Moving average, Max(duration) 5day Moving Average on ASHOKLEY.NS stock which was taken from yahoo finance. Six models were created based on the duration of the stock and the interval chosen for Transformer Network models. The data was divided into training and testing sets, which is a fundamental concept in machine learning, especially when training models for prediction tasks. This prevents overfitting. During training, the model learns patterns and relationships within the training data. The testing set, unseen by the model during training, is used to evaluate the model's generalizability. By evaluating unseen data (testing set), we can identify models that are at risk of overfitting and potentially take corrective measures for prediction. The figures show the actual value vs. trained and testing sets for Transformers and LSTM models.

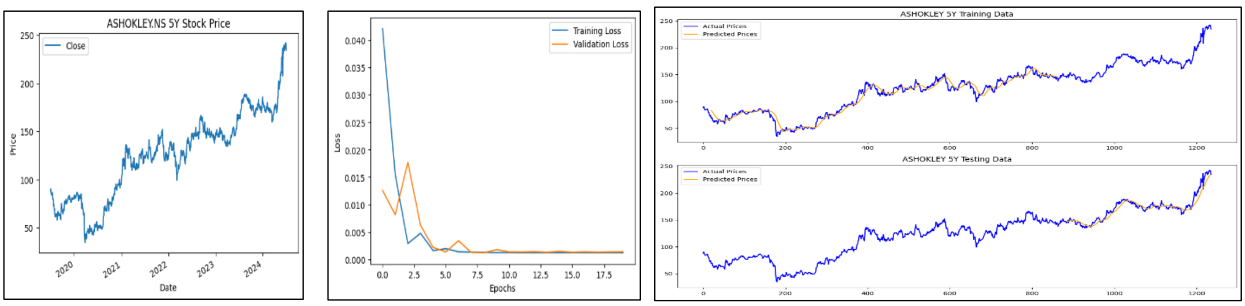


Fig. Transformer model 5 Y 1 day (Actual, Test loss, Training, Testing)

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 2 (2024)
http://eelet.org.uk

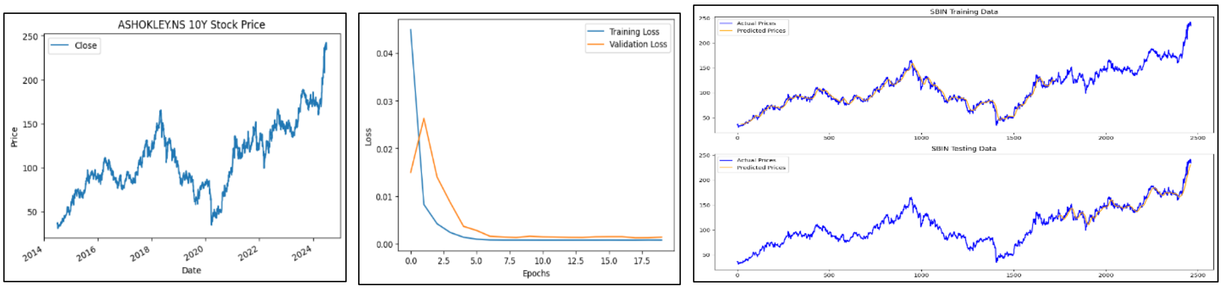


Fig. Transformer model 10 Y 1 day (Actual, Test loss, Training, Testing)

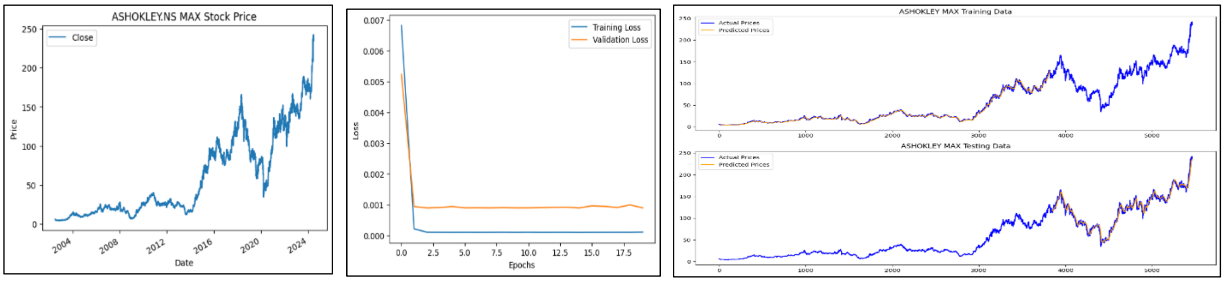


Fig. Transformer model Max Year 1 day (Actual, Test loss, Training, Testing)

Data and Interval	Test Loss	RMSE	MAE	MAPE	R SQUARE	DA
5Y 1D	0.00137638	0.037099588	0.027459583	16.91894382	0.89286798	52.72206304
10Y 1D	0.001395042	0.037350256	0.027771008	21.87099896	0.907385279	49.51185495
MAX 1D	0.000889878	0.030243576	0.022611197	45.89199316	0.967198749	49.19653894

Table. Data and interval (without Moving Averages) and Accuracy metrics

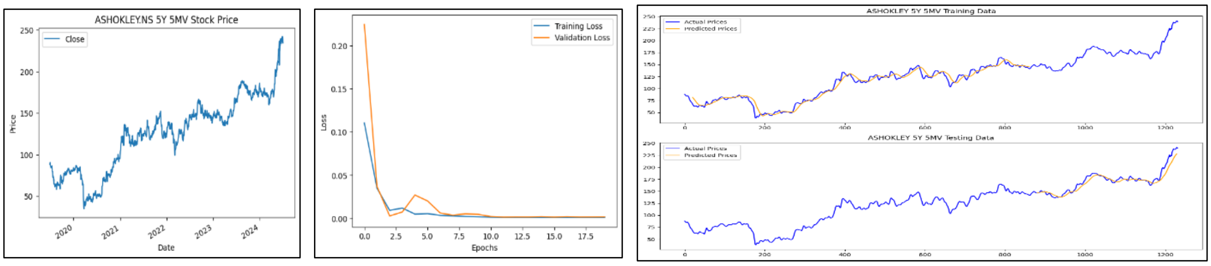


Fig. Transformer model 5 Year 5 day Moving Average(Actual, Test loss, Training, Testing)

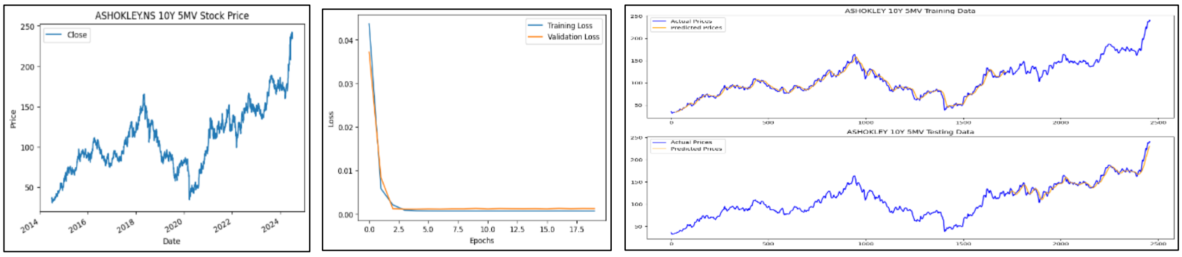


Fig. Transformer model 10 Year 5 day Moving Average(Actual, Test loss, Training, Testing)

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 2 (2024)
http://eelet.org.uk

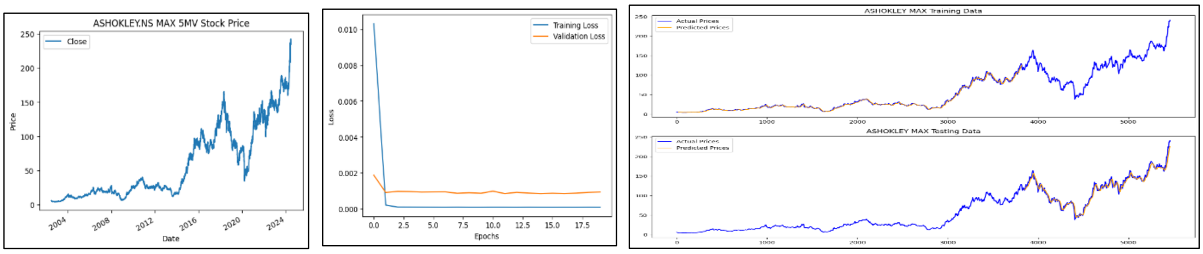
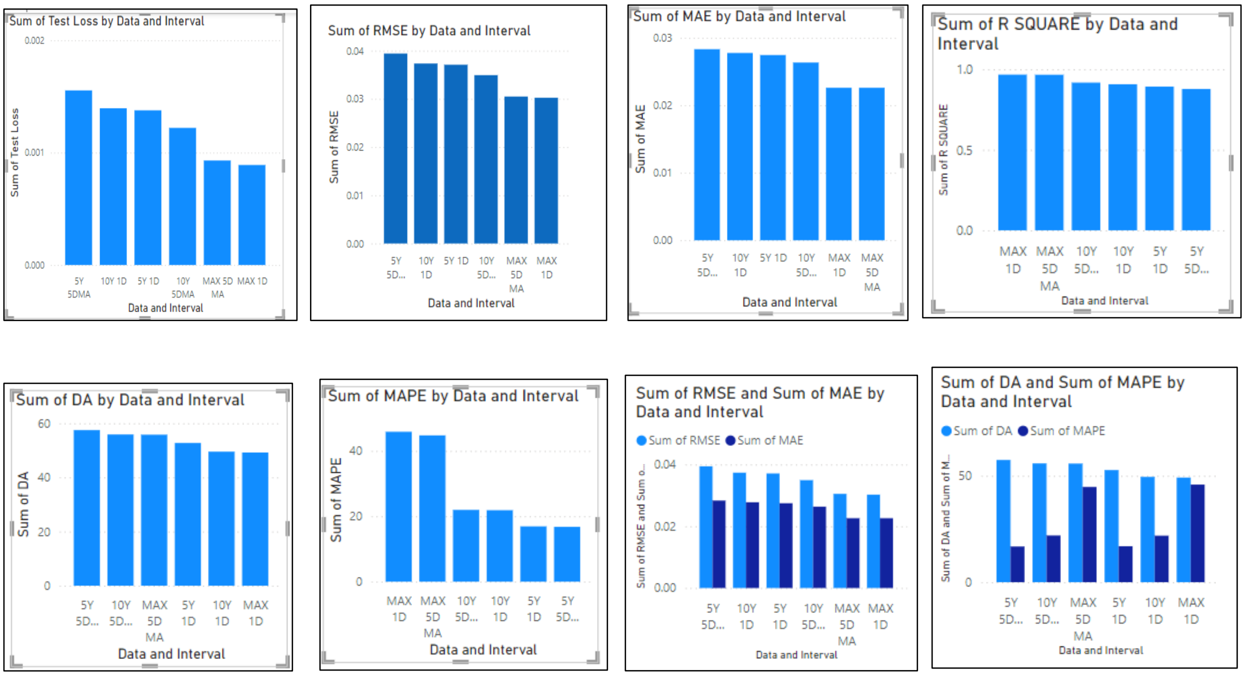


Fig. Transformer model Max Year 5 day Moving Average(Actual, Test loss, Training, Testing)

Data and Interval	Test Loss	RMSE	MAE	MAPE	R SQUARE	DA
5Y 5DMA	0.00155383	0.039418649	0.028325937	16.78203128	0.87809117	57.47126437
10Y 5DMA	0.001220133	0.034930411	0.026353533	21.9676804	0.917563766	55.86592179
MAX 5D MA	0.000930132	0.030498057	0.02261072	44.78399213	0.966453255	55.78231293

Table. Data and interval (with Moving Averages) and Accuracy metrics



Test Loss and Performance Metrics:

Test Loss: Lower test loss values generally indicate better model performance on unseen data. Here, all the test loss values are below 0.002, suggesting the models achieved an adequate fit to the data. Of all the observations Max 1 day test loss is minimum 0.000889878, indicating the model will perform better for longer durations.

RMSE: Similar to Test Loss, lower RMSE (Root Mean Square Error) measures the average difference between predicted and actual values. The lower RMSE signifies a better fit. Here, the values range from 0.030 (MAX 1D) to 0.039 (5Y 5DMA), indicating a relatively small average difference between the predicted and actual values.

MAE (Mean Absolute Error): The lower MAE translates to better performance. Here, the values range from 0.022 (MAX 1D and MAX 5D MA) to 0.028 (5Y 5DMA and 10Y 1D), suggesting a close match between predicted and actual values on average.

MAPE (Mean Absolute Percentage Error): Lower MAPE indicates better accuracy, particularly when dealing with percentages. Here, the values range from 16.78% (5Y 5DMA) to 45.89% (MAX 1D), with MAX 1D having a considerably higher percentage error. This could indicate that the model struggles more with predicting significant price movements for the MAX interval.

R-Squared (Coefficient of Determination): A higher R-squared value (closer to 1) suggests a stronger correlation between predicted and actual values. Here, all the R-squared values are above 0.87, indicating a good positive correlation between the model's predictions and the actual stock prices.

Direction Accuracy (DA): This metric specifically focuses on how well the model predicts the direction of the trend (up or down) rather than the exact value. A higher DA implies better directional forecasting accuracy. Here, the values range from 49.19% (MAX 5D MA) to 57.47% (5Y 5DMA), suggesting that the models were moderately successful in predicting the directional movement of the stock prices.

Key Observations from the Performance Metrics:

- Overall, the results seem to suggest that the transformer models achieved a reasonable fit to the data, with relatively low errors and a good positive correlation between predictions and actual values.
- MAX 1D appears to have the highest errors (RMSE, MAE) and the lowest MAPE, possibly indicating that the model performs well on capturing smaller movements but struggles with larger ones for the maximum daily interval.
- 5Y 5DMA has the highest Direction Accuracy (DA) of 57.47%, suggesting it might be better at predicting the overall trend direction for the 5-year window with 5-day moving averages.

Considerations and Further Exploration:

- It would be beneficial to compare the performance of these transformer models with a baseline model, such as ARIMA or LSTM, to assess if the transformers offer a significant improvement.
- Since you've mentioned limitations, it would be helpful to understand what challenges you faced while using transformers for this task.
- Experimenting with different hyperparameters for the transformer model might lead to further improvements.
- You could explore including additional features or data sources that might be relevant for stock price prediction, such as news sentiment or economic indicators.

Overall, the analysis suggests that the Transformer model has the potential to forecast stock prices for ASHOKLEY.NS data. However, its performance seems to be better for shorter forecasting horizons (1-day) and might require overfitting mitigation techniques.

Suggestions:

To improve the Transformer model's performance for stock price forecasting on ASHOKLEY.NS data, considering the observations about interval, duration, and overfitting are (i) While the model performs well for 1-day forecasts, consider incorporating additional high-frequency data points (e.g., intraday prices) to potentially capture even more granular patterns that might influence short-term price movements (ii) Explore techniques specifically designed to handle long-range dependencies in time series data. This could involve: (a) Dilated convolutions: These convolutions allow the model to "look back" further in the sequence, potentially capturing long-term relationships in the data (b) Recurrent Neural Networks

(RNNs) with attention mechanisms: LSTMs or GRUs with attention can focus on relevant parts of the historical data for longer forecasts.

Conclusions:

The transformer model achieved the best overall performance in predicting Ashokley.NS stock prices when trained on maximum duration of daily data. This is evident from the lowest Test Loss, RMSE and MAE values observed for this configuration. Including a 5-day moving average generally led to improvement in the model's performance across data durations, as evidenced by lower error metrics and high Rsquare and Direction Accuracy.

This research investigated the use of Transformer networks for forecasting trends of stock prices on the ASHOKLEY.NS data. The model demonstrated potential for accurate predictions, particularly for shorter forecasting horizons (1-day) as measured by 1-minus-RMSE. However, the performance seems to improve with moving averages and for longer horizons. Additionally, some degree of overfitting was observed, requiring techniques like dropout regularization for improvement. While these findings are promising, limitations include the use of a single stock and a specific data duration. Future research could explore incorporating additional features or comparing the performance with other forecasting models to gain a more comprehensive understanding of Transformer networks' efficacy for stock price prediction.

Future Work:

- Experimenting with hyperparameter tuning to see if the model's performance can be further improved for this task.
- Exploring incorporating additional features (e.g., economic indicators) that might be relevant for stock price prediction.
- Comparing the performance of the Transformer model with other forecasting techniques like LSTMs or Prophet to see which method yields the best results for your specific data and forecasting needs.

References

A. Murkute and T. Sarode, "Forecasting Market Price of Stock using Artificial Neural Network," International Journal of Computer Applications, vol. 124, no. 12, pp. 11–15, Aug. 2015.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.

1. Adityanarayanan Radhakrishnan, Mikhail Belkin and Caroline Uhler, (2023) "Wide and Deep Neural Network Achieve Consistency for Classification", PNAS, Vol.120, N0. 14, https://doi.org/10.1073/pnas.2208779120
2. Aman Sehgal,(2021), Forecasting Stock Price Movements for Intra-Day Trading using Transformers and LSTM, International Journal of Computing and Artificial Intelligence, https://doi.org/10.33545/27076571.2021.v2.i1a.25
3. Bryan Lim, Sercan O Arik, Nicolas Loeff, Tomas Pfister,(2021), Temporal Fusion Transformers for Interpretable Multi- Horizon Time Series Forecasting, International Journal of Forecasting, Vol 37, Issue 4, Pages 1748-1764, https://doi.org/10.1016/j.ijforecast.2021.03.012
4. Chaojie Wang et al (2022), "Stock Market Index Prediction Using Deep Transformer Model, Expert systems with applications, 208.118128/10.1016/j.eswa.2022.118128
5. Chaojie Wang et al (2022), "Stock Market Index Prediction using Deep Transformer Model", Expert Systems with Applications 208:118128, http://dx.doi.org/10.1016/j.eswa.2022.118128
6. Divy Daiya, Che Lin, (2021), "Stock Movement Prediction and Portfolio Management Via Multimodal Learning with Transformer", IEEE International Conference on Acoustics, Speech and Signal Processing, http://dx.doi.org/10.1109/ICASSP39728.2021.9414893
7. Eduardo Ramos Perez, Pablo J. Alonso-Gonzalez, Jose Javier Nunez-Velazquez, (2021),"Multi-Transformer: A New Neural Network-Based Archietecture for forecasting S& P Volatility, https://doi.org/10.3390/math9151794

8. Emma Yann Zhang et al (2023), “From Turing to Transformers: A Comprehensive Review and Tutorial on the Evolution and Applications of Generative Transformer Models”, MDPI, Sci 2023, 5, 46, https://doi.org/10.3390/sci5040046
9. Fama, E. F., & Blume, M. F. (1966). Filter rules and stock market trading. Journal of Business, 39, 226–241.
10. Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. Journal of Financial Economics, 22, 3–25.
11. G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. Time series analysis: forecasting and control. John Wiley & Sons, 2015.
12. Gaurang Sonkavde et al (2023), “Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A systematic Review, Performance Analysis and Discussion of Implications”, MDPI, International Journal of Financial Studies, 11:94, https://doi.org/10.3390/ijfs11030094
13. https://heidloff.net/assets/img/2023/02/transformers.png, https://heidloff.net/article/foundation-models-transformers-bert-and-gpt/, Niklas Heidloff, (2023) Foundation Models, Transformers, BERT and GPT
14. J. Liu, H. Lin, X. Liu, B. Xu, Y. Ren, Y. Diao, and L. Yang. Transformer-based capsule network for stock movement prediction. In Proceedings of the First Workshop on Financial Technology and Natural Language Processing, pages 66–73, 2019.
15. J. Schmidhuber, “Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, Jan. 2015.
16. J. Yen et al. (2023), “Stock Price Forecast: Comparison of LSTM, HMM and Transformer”, ICBIS, AHCS 14, PP. 126-136, https://doi.org/10.2991/978-94-6463-198-2_15
17. Jintao Liu et al., (2019), “Transformer based Capsule Network for Stock Movements Prediction”, Dalian University of Technology, Dailian, China.
18. Nadeem Malibari et al , (2021), “Predicting Stock Closing Prices in Emerging Markets with Transformer Neural Networks: The Saudi Stock Exchange Case,Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia
19. Niklas Hiedloff, (2023), Foundation models, Transformers, BERT, GPT, hiedloff.net, https://heidloff.net/article/foundation-models-transformers-bert-and-gpt/
20. Peng, Xiaoqi, (2023), “Stock Forecasting Using Transformers”, Master’s Thesis, Nanyang Technological University, Singapore, https://hdl.handle.net/10356/169344
21. Polamuri, Subba Rao, Kudipudi Srinivas, and A. Krishna Mohan. 2019. Stock Market Prices Prediction using Random Forest and Extra Tree Regression. International Journal of Recent Technology and Engineering 8: 1224–28. [Google Scholar] [CrossRef]
22. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
23. Saidul Islam, Hanae Elmekki, Ahmed Elsebai, Jamal Bentahar, (2023), “A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks”, Research Gate, Electrical Power Engineering, http://dx.doi.org/10.48550/arXiv.2306.07303
24. Tariq Saeed Mian , (2023), Evaluation of Stock Closing Prices Using Transformer Learning, Eng. Technol. Appl. Sci. Res., vol 13, No. 5 PP 11635-11642, https://doi.org/10.48084/etasr.6017
Tashreef Muhammad et al (2022), “Transformer Based Deep Learning Model for Stock Price Prediction: A Case Study on Bangladesh Stock Market,
25. Yawei Li, Shuqi Lv, Xinghua Liu, (2022) Incorporating Transformers and Attention Networks for Stock Movement Prediction, Shandong University of Financial Economics, http://dx.doi.org/10.1155/2022/7739087
26. Yixin Guo (2022), “Stock Price Prediction Using Machine Learning”, Economics Spring, Stockholm School of Social Science Master 9 Dissertation)