Stock Price Prediction using Time Series Forecasting Transformer

1st Arshiya Kishore, Cheshta Biala, Mannast Sadana, Yatharth Gautam
Department of Computer Science
Thapar Institute of Engineering and Technology
Patiala, India
akishore_be21@thapar.edu,cbiala_be21@thapar.edu,msadana_be21@thapar.edu,ygautam_be21@thapar.edu

***Abstract*—This study focuses on time-series forecasting of monthly stock prices for companies listed on the National Stock Exchange (NSE) using historical data from 1990 to 2021. A Transformer-based architecture, tailored for sequential data, is employed to address the challenges of market volatility and non-linear price movements. The methodology includes robust data preprocessing, feature engineering, and optimization techniques. Key evaluation metrics, including Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R^2, demonstrate the model's superior performance in capturing long-term dependencies and trends. The findings highlight the efficacy of Transformer models in financial forecasting, offering a scalable and interpretable solution for real-world applications such as portfolio optimization and risk management.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Predicting stock prices is a crucial yet complex task in the financial sector, as stock prices are influenced by a multitude of factors, including market sentiment, macroeconomic indicators, and geopolitical events. The non-linear and volatile nature of stock price movements makes accurate forecasting challenging. This research focuses on predicting the next month's stock prices for companies listed on the National Stock Exchange (NSE) using historical data from 1990 to 2021. To address these challenges, we leverage deep learning models, specifically Transformer-based architectures, known for their ability to handle sequential data and capture long-term dependencies. The objective is to develop a robust and scalable time-series forecasting model that can provide reliable predictions, enabling investors and analysts to make informed, data-driven decisions in a dynamic market environment.

II. Related Work

Hu et al. (2021) [1] explored the Temporal Fusion Transformer (TFT) for stock price prediction, demonstrating its superiority over models like SVR and LSTM in multi-horizon forecasting. The TFT's attention-based architecture effectively handled both temporal and static data, achieving lower prediction errors. However, the model's reliance on large datasets and exclusion of external factors like market news limited its applicability in data-scarce scenarios. Malibari et al. (2021) [2] applied Transformer models to the Saudi Stock Exchange, achieving over 90% accuracy in stock price predictions. Their approach outperformed traditional models such as ARIMA and LSTM, but it focused on short-term predictions without incorporating external factors like economic indicators, which could improve its accuracy. Wang et al. (2022) [3] applied Transformer architecture to predict stock market indices, showing superior performance over CNNs and RNNs. The model's ability to capture long-range dependencies made it more effective in forecasting. However, it was limited to short-term predictions and did not integrate macroeconomic factors, which could enhance its forecasting capability. Wang (2023) [4] proposed a hybrid BiLSTM-MTRAN-TCN model, combining BiLSTM, Transformer, and TCN for improved accuracy and stability in stock price forecasting. While the model showed superior performance in terms of RMSE reduction and R^2 improvement, it depended on specific time window lengths and did not consider external data sources like market news, which could further enhance predictions. Wang (2023) [4] proposed a hybrid model, BiLSTM-MTRAN-TCN, which combines BiLSTM, modified Transformer (MTRAN), and Temporal Convolutional Networks (TCN) to improve prediction accuracy and stability in stock price forecasting. The model demonstrated superior performance compared to other models like LSTM and CNN-BiLSTM, with significant reductions in RMSE and improvements in R^2. Despite its success, the study pointed out that the model's performance is dependent on specific time window lengths, which may limit its adaptability. Moreover, the study did not explore the integration of external data sources, such as market news or economic indicators, which could provide additional context and improve the accuracy of predictions.

III. Proposed Architecture and Methodology Adopted.

A. Proposed Architecture

Each stage is designed to ensure the creation of a robust time-series forecasting model. 1. Data Collection: The dataset consists of historical stock price data from the National Stock Exchange (NSE) spanning 1990–2021. Key features include Open, High, Low, Close, and Volume. 2. Preprocessing: Ensures the dataset is clean and normalized, suitable for feeding into the model. 3. Feature Engineering: Derives

Identify applicable funding agency here. If none, delete this.

meaningful features to improve predictive performance. 4. Model Selection: Deep learning models such as LSTM and Transformer are implemented for sequential data analysis. 5. Training: Optimizes model weights using the Adam optimizer and fine-tunes hyperparameters. 6. Prediction and Evaluation: Compares predicted stock prices with actual values using metrics like MSE and MAPE to validate performance.

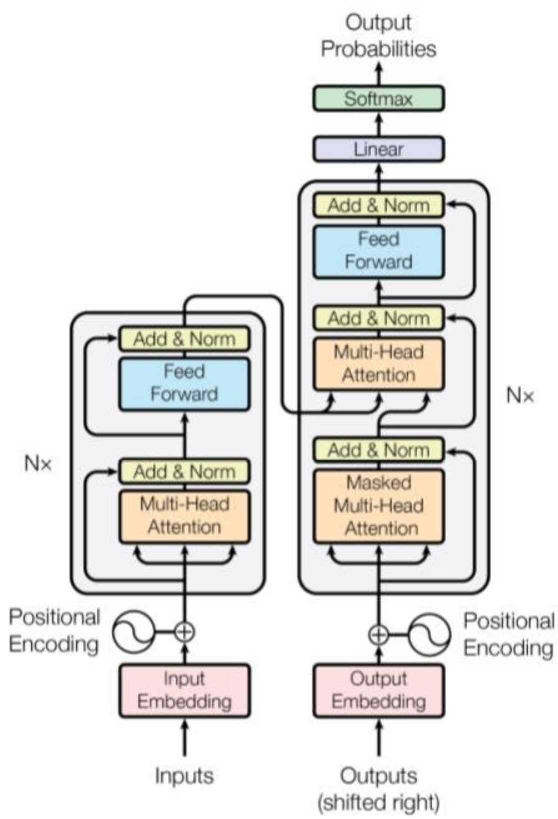

Fig. 1. Model Architecture

B. *Technical Aspects*

1. Preprocessing

- Normalization: The raw data is normalized using the Min-Max Scaler to scale features between 0 and 1, ensuring uniformity and preventing large feature values from dominating the model's learning process.
- Handling Missing Values: Missing entries are interpolated linearly using Pandas functions, ensuring temporal continuity in the data. This step minimizes the loss of valuable information due to gaps in the dataset.
- Feature Engineering: New features, such as moving averages and volatility measures, are calculated to capture trends and price fluctuations over time. These enhance the dataset's ability to represent underlying patterns critical for accurate predictions.

3. Model Architecture

- LSTM Model: Designed to handle sequential data and long-term dependencies. It uses memory cells and gating mechanisms to learn time-series patterns effectively.
- Layers: Sequential LSTM layers followed by dense layers for regression output. o Activation Function: Rectified Linear Unit (ReLU) in intermediate layers and linear activation for the output layer.
- Time Series Forecasting Transformer Model: Employs self-attention mechanisms to capture long-range dependencies and handle multivariate time-series data efficiently. This model can identify intricate patterns and dependencies across time steps.
- Layers: Encoder layers with multi-head attention and feed-forward networks, concluding with a dense output layer.

4. Optimization

- Optimizer: Adam optimizer is used due to its adaptive learning rate capabilities.
- Loss Function: Mean Squared Error (MSE) minimizes the error between predicted and actual values, focusing on accurate price forecasting.

IV. Results and Experiments

A. *Dataset details, Train and test data split*

- Source: Historical stock price data from NSE (1990–2021).
- Features: Open, High, Low, Close, and Volume.
- Preprocessing: Missing values were handled using linear interpolation, and features were scaled to a 0–1 range using Min-Max Scaler.
- Train-Test Split: An 80:20 ratio was used to divide the data into training and testing sets.

B. *System configuration used for training computing results*

- Platform: Google Collab with GPU-enabled support for faster training.
- Software: Python 3.9 with libraries such as TensorFlow, NumPy, Pandas, and Matplotlib. • Hardware: A system with 12 GB GPU RAM.

C. *Training Details :*

Architecture:

- LSTM: Sequential LSTM layers for handling sequential data, followed by dense layers for output.
- Time Series Forecasting Transformer: Multi-head attention mechanisms with encoder layers, coupled with feed-forward networks for sequential prediction.

• Hyperparameters:

- Batch Size: 64.
- Epochs: 100, with early stopping to prevent overfitting.
- Optimizer: Adam optimizer for efficient convergence.

• Metrics:

- Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values.

- Mean Absolute Percentage Error (MAPE): Quantifies prediction accuracy as a percentage.
- R2 (Coefficient of Determination): Evaluates the proportion of variance in the dependent variable explained by the model. An R2 value closer to 1 indicates a better fit.

• Training Curve Analysis:

- Loss values stabilized within 50 epochs, indicating effective learning without overfitting.

Here we have worked and trained model on csv file of scrips of Micron company and here are the results:-

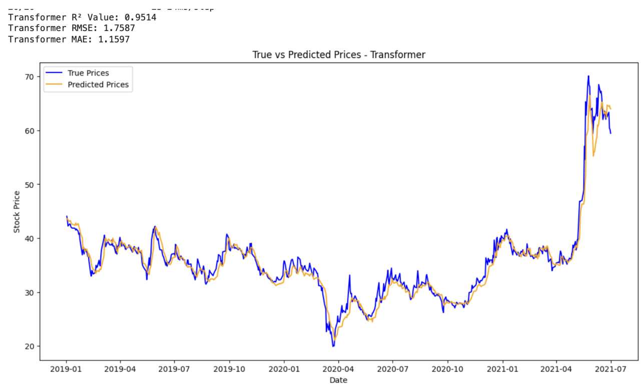

Fig. 2. Actual vs Predcited for Micron Company

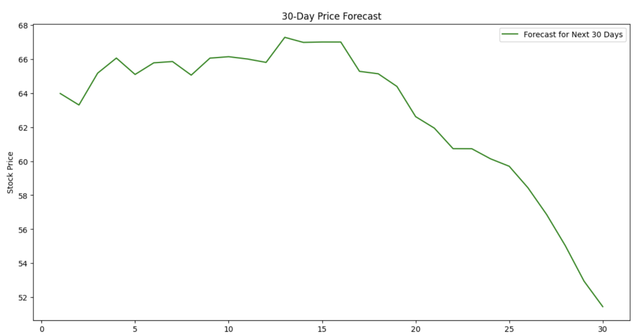

Fig. 3. Future Predicition for Micron Company

The performance of TIME SERIES FORECASTING transformer is:- Transformer R2 Value: 0.9514 Transformer RMSE: 1.7587 Transformer MAE: 1.1597

V. Conclusion and Future Scope

A. Conclusion

This research successfully demonstrated the application of deep learning techniques, including LSTM and Transformer architectures, for stock price prediction over a one-month horizon. The methodology employed advanced preprocessing steps, feature engineering, and state-of-the-art models to effectively capture both short-term fluctuations and long-term dependencies in historical stock data. Key findings include:

- The LSTM model effectively captured sequential dependencies, while the Transformer architecture leveraged attention mechanisms for enhanced pattern recognition.
- The models exhibited low RMSE and MAPE values, showcasing their predictive accuracy and generalization capability.
- The use of the R2 metric confirmed the models' ability to explain a substantial proportion of the variance in stock price trends, validating their reliability and interpretability.
- Feature engineering, such as the inclusion of moving averages and volatility measures, significantly improved the dataset's representational power, leading to better model performance.

This work establishes a solid foundation for leveraging deep learning in financial forecasting and provides a robust solution for short-term stock price predictions. Accuracy and performance of TIME SERIES FORECASTING TRANSFORMER on the model is:-

- Transformer R2 Value: 0.9514
- Transformer RMSE: 1.7587
- Transformer MAE: 1.1597

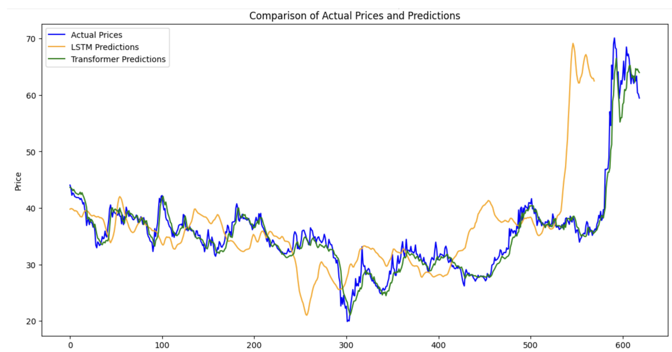

Fig. 4. Overall Performance of Model

B. Future Scope

While the results are promising, there are several opportunities to expand and enhance this research:

- Incorporating External Factors:Integrate external data sources such as market news, sentiment analysis, and macroeconomic indicators to improve prediction reliability.Leverage alternative data, like social media sentiment or economic reports, for a more comprehensive model.
- Longer Forecasting Horizons:Extend the prediction timeline to quarterly or annual forecasts, requiring adaptations in model design and training methodologies.
- Hybrid Models:Explore hybrid architectures that combine the strengths of LSTM, Transformers, and other advanced models like Temporal Convolutional Networks (TCN).Implement ensemble learning techniques to combine predictions from multiple models for higher accuracy.
- Real-time Deployment:Develop a real-time prediction system for live stock price forecasting, with continuous data ingestion and on-the-fly predictions.Integrate the

model into trading platforms to assist investors with portfolio optimization and risk management.

- Explainability and Interpretability: Incorporate techniques to explain model predictions, making them more interpretable for financial analysts.Utilize SHAP (SHapley Additive exPlanations) or attention-based visualizations to provide insights into feature importance.

VI. REFERENCES

REFERENCES

[1] Q. Ding, S. Wu, H. Sun, J. Guo, and J. Guo, "Hierarchical multi-scale Gaussian transformer for stock movement prediction," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 4640-4646.

[2] X. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, Marina Del Rey, CA, USA, Feb. 2018, pp. 261-269.

[3] N. Malibari, I. Katib, and R. Mehmood, "Predicting stock closing prices in emerging markets with transformer neural networks: The Saudi Stock Exchange case," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 876-885, 2021, doi: 10.14569/IJACSA.2021.0121277.

[4] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, "Stock market index prediction using deep transformer model," *Expert Systems with Applications*, vol. 208, Dec. 2022, Art. no. 118128, doi: 10.1016/j.eswa.2022.118128.

[5] S. Wang, "A Stock Price Prediction Method Based on BiLSTM and Improved Transformer," *IEEE Access*, vol. 11, pp. 104211-104223, 2023, doi: 10.1109/ACCESS.2023.3296308.