Transformer Based Stock Market Analysis: Fusing Time Series with Textual Information

Arun Prakash Jayakanthan
jayakant@usc.edu
...
Charanya Shrimushnam Kannan
shrimush@usc.edu
...
Prajit Rahul Selvam
pselvam@usc.edu
...
Venkataraman Venkatasubramaniam
venkataraman.venkatasubramaniam@usc.edu
...

Abstract

This project develops a cross-modal approach to predicting stock market trends by combining structured time series data (stock prices) with unstructured textual data (e.g., news sentiments) and additional financial indicators. The project progresses through three phases to build a robust model for trend analysis in dynamic financial environments. In Phase 1, a baseline model is established using only time series data to capture fundamental patterns and trends. Phase 2 introduces a fusion of time series data with sentiment information extracted from recent news articles, enabling the model to incorporate contextual information regarding market sentiment. This phase also introduces few-shot prompting with Llama, where sentiment data, represented textually, is combined with time series data. Phase 3 further enriches the model by integrating economic indicators and market sentiment indicators alongside time series and news sentiment data, aiming to improve predictive stability. Results indicate that adding news sentiments enhances short-term predictions while increasing the volume of available data significantly improves long-term predictions. Conversely, the addition of VIX and GDP indicators does not significantly impact model performance.

1 Introduction

Stock market prediction involves forecasting the future value of a stock based on historical data. Given the need to identify patterns in large datasets reflecting stock price fluctuations, machine learning models are particularly suited for this task. Predicting stock market trends holds immense significance for a wide array of stakeholders in the financial ecosystem. Accurate predictions enable investors and traders to refine their strategies for buying, selling, and holding assets. Financial analysts benefit from enhanced tools for assessing risk and formulating strategies to mitigate potential downturns. Moreover, policymakers can utilize these insights to monitor economic health and shape monetary policies, thereby influencing the broader economy.

However, stock market prediction is fraught with challenges. One major issue is market volatility; financial markets are inherently unpredictable and influenced by numerous factors. Additionally, the non-stationary nature of financial time series, which exhibit evolving statistical properties over time, limits the applicability of traditional modeling methods. Integrating structured time series data with

Preprint. Under review.

unstructured data—such as news articles or social media sentiment—further complicates the process. Addressing these challenges is crucial to improving forecasting accuracy and providing actionable insights.

Overcoming these challenges would greatly benefit stakeholders. Enhanced forecasting models could empower investors and traders to make more informed decisions and optimize their trading strategies. Financial analysts would gain better tools for identifying market trends and offering actionable insights. Portfolio managers could leverage improved predictive capabilities for effective asset allocation and risk management. By addressing the complexities of stock market prediction, this research seeks to enable stakeholders to navigate the intricate dynamics of financial markets more effectively.

Two major research tracks have demonstrated significant progress in this domain: advancements in machine learning models and the incorporation of multimodal data. Time series forecasting has traditionally relied on econometric methods such as ARIMA models. For instance, Banerjee et al. [1] explored forecasting macroeconomic variables for new EU member states using ARIMA models, while Khashei and Bijari [2] proposed a hybridization of artificial neural networks and ARIMA for time series forecasting. Despite their prevalence, ARIMA models are limited in capturing nonlinear relationships among variables.

In contrast, deep learning models excel at identifying complex patterns and relationships often missed by traditional methods. Deep learning approaches range from simple deep neural networks (DNNs) (Chong et al.,) [3] and recurrent neural networks (RNNs) (Chandra & Chand) [4] to more specialized architectures like long short-term memory (LSTM) networks (Lee & Yoo) [5] and convolutional neural networks (CNNs) (Ding et al.) [6]. These models have shown varying degrees of success in predicting financial time series.

Transformers, originally developed for language translation tasks, have demonstrated exceptional capability in modeling long-range dependencies in sequential data. The Temporal Fusion Transformer (TFT), as introduced by Lim et al. [7], has proven effective for time series modeling and offers interpretability, making it particularly well-suited for financial applications.

Simultaneously, research has focused on integrating textual data to enhance stock market predictions. Similar tasks, such as taxi demand prediction (Rodrigues et al.) [8], have benefited from combining time-series and textual data through deep learning approaches. For stock market prediction, efforts have incorporated textual data such as news articles, using models ranging from ARIMA (Wang et al.) [9] to LSTMs (Li et al.) [10] and pretrained transformers that process news as prompts (Cao et al.) [11].

Incorporating textual data has shown promise in improving the accuracy of stock market predictions. Leveraging financial news data, combined with the capabilities of advanced models like transformers, presents a logical next step. In this paper, we investigate the impact of incorporating financial news data into the prediction of stock market prices. Specifically, we explore training a Temporal Fusion Transformer (TFT) model on financial news data alongside historical stock market values to enhance accuracy for both short-term and long-term prediction tasks. Additionally, we incorporate economic indicators such as FRED (Federal Reserve Economic Data) and VIX (Volatility Index) to evaluate whether TFT can capture global patterns. Finally, inspired by recent advancements in large language models (LLMs) for time series forecasting (Jin et al.) [12], we experiment with few shot learning with an LLM for stock price prediction. Comprehensive ablation studies and explanations accompany all experiments.

2 Related Work

Recent advancements in the application of large language models (LLMs) for cross-modal multivariate time series forecasting have garnered attention, especially in the integration of structured time series data with textual information. Liu et al. [13] proposed a Cross-Modal LLM Fine-Tuning framework, which seeks to bridge performance gaps caused by distribution discrepancies between textual and temporal inputs. While their framework has shown efficacy in both short- and long-term forecasting, it has yet to fully leverage the reasoning capabilities of LLMs. Our approach addresses this gap by directly incorporating textual information into the forecasting analysis, aiming to enhance the interaction between textual and temporal reasoning. This method has been evaluated on various datasets, including ETT and M4, achieving state-of-the-art results in multivariate time series forecasting tasks.

Liu et al. [14] developed TimeCMA, an LLM-based model that emphasizes cross-modality alignment through a dual-modality encoding framework. By processing text-based prompts via pre-trained LLMs, TimeCMA has achieved noteworthy results; however, its computational demands during the prompt-encoding phase can be prohibitive. Our project proposes a more efficient early fusion strategy that mitigates these computational costs while maintaining forecasting accuracy. TimeCMA was also tested across datasets like ETT, Weather, and FRED-MD, where it demonstrated promising forecasting capabilities but at a high computational expense.

Ding et al. [15] explored the intersection of LLMs and quantitative finance, proposing a framework that aligns financial news embeddings with traditional stock features through a hybrid Local-Global model. This model captures both LLM-based semantic information and classic stock characteristics effectively. Building upon this work, our project aims to extend cross-modal methodologies to integrate time series data more efficiently, particularly by employing early fusion techniques to handle long-range dependencies. This strategy has been validated on the China A-share market dataset, where it outperformed models reliant solely on traditional stock features.

3 Methodology

3.1 Dataset Description

For our project, we will utilize two key APIs to gather our datasets:

- **Yahoo Finance API**: This API will be used to obtain structured time series data, including stock prices, volumes, and historical market data for our target stocks.
- **Alpha Vantage API**: We will leverage the Alpha Vantage API to gather news sentiment data. This API provides access to financial news articles along with sentiment analysis tools that help in understanding market sentiment.

3.2 Evaluation Strategy

To validate our model's performance, we will employ several metrics:

- **Mean Absolute Error (MAE)**: MAE calculates the average absolute differences between predicted values and actual outcomes.

  $$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

  where y_i represents the actual values, $\hat{y}_i$ the predicted values, and n the total number of observations.
- **Root Mean Squared Error (RMSE)**: This metric measures the square root of the average squared difference between predicted and actual values, providing information on the precision of the prediction while penalizing larger errors.

  $$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

  where y_i represents the actual values, $\hat{y}_i$ the predicted values, and n the total number of observations.
- **Prediction Stability**: This metric will assess how consistently our model can make predictions over time, particularly during volatile market conditions.
  Let $\Delta\hat{y}_t = \hat{y}_t - \hat{y}_{t-1}$ represent the changes in predicted values between consecutive time steps. The stability is computed as:

  $$\text{Stability} = \frac{1}{T-1}\sum_{t=2}^{T}|\Delta\hat{y}_t|$$

where T is the total number of time steps, and $\Delta \hat{y}_t$ represents the difference in predictions over consecutive time periods.

3.3 Experimental Design

Our approach consists of three main phases designed to enhance stock market trend prediction through a cross-modal framework:

3.3.1 Phase 1 - Establishing Baselines

In this initial phase, we develop baseline models using established time series models alongside transformer-based models. This comparison will allow us to understand the strengths and weaknesses of each approach in capturing market trends.

We utilize historical stock price data- closing prices. The dataset spans around 9 months for approximately 200 stock tickers. This data provides a foundation for identifying temporal patterns and trends.

In Phase 1, we implemented a SARIMA (Seasonal AutoRegressive Integrated Moving Average) (Dong et al.) [16] model for time series forecasting using AutoGluon's TimeSeriesPredictor. We utilized nine months of data, allocating seven months for training, one month for validation, and one month for testing. To assess the model's performance under various configurations, we experimented with different hyperparameters, including the order of the ARIMA component (p, d, q), the order of the seasonal component (P, D, Q), and the seasonal period. The TimeSeriesPredictor was configured with a prediction length of 7 days for short-term forecasts, and we explored longer horizons of 14 and 30 days for medium and long-term predictions respectively. AutoGluon's automated hyperparameter tuning capabilities were leveraged to optimize the SARIMA model's performance. We employed the default validation strategy, where the last prediction length timesteps of each time series in the training data were used for validation. The model's performance was evaluated using MAE across the three forecasting horizons: short-term (7 days), medium-term (14 days), and long-term (30 days).

To explore competing models, we implemented the TFT (Hu et al.) (Mozaffari et al.) [17,18] model for time series forecasting using nine months of data, with the same train, validation and test split as experiments performed using SARIMA. To evaluate the model's performance under varying configurations, we experimented with different permutations of data input parameters, including the size of the dataset, maximum prediction lengths (3 and 6 days), and maximum encoder lengths (5 and 15 days). The TFT model was configured with a learning rate of 0.001, a hidden size of 64, an attention head size of 4, a dropout rate of 0.1, and a hidden continuous size of 32. The AdamW optimizer and a quantile loss function were utilized to improve stability and performance. Early stopping was employed with a patience of 10 epochs to prevent overfitting, while a learning rate monitor provided dynamic adjustments during training. The model was evaluated across the same forecasting horizons as done using SARIMA.

3.3.2 Phase 2 - Integrating Time Series with News Sentiment

The second phase focuses on combining structured time series data (like stock prices) with news article sentiments (Yang et al.) [19].

Sentiment data is derived from financial news articles, where each article is scored for sentiment on a per-ticker basis. These sentiment scores are pre-defined in the dataset. The sentiment score ranges include:

- Bearish: $x \leq -0.35$
- Somewhat-Bearish: $-0.35 < x \leq -0.15$
- Neutral: $-0.15 < x \leq 0.15$
- Somewhat-Bullish: $0.15 \leq x < 0.35$
- Bullish: $x \geq 0.35$

We utilize an early fusion strategy that combines structured (time series) data and unstructured (news sentiment scores) data. The sentiment data is processed on a ticker-by-ticker and daily basis, associating relevant sentiment scores with their corresponding stock tickers. These sentiment scores are timestamped and joined with structured time series data to maintain temporal alignment. This combined feature representation allows the TFT model to incorporate both time series trends and contextual news sentiment changes simultaneously.

In Phase 2, we implemented the TFT model for time series forecasting using nine months of data, with seven months allocated for training, one month for validation, and one month for testing. The base model configuration in Phase 1 serves as the foundation for subsequent experiments in Phase 2, where the same TFT model is employed. In Phase 2, the base model's performance was extended by integrating additional contextual data (news sentiment scores), while maintaining the same architectural and hyperparameter settings from Phase 1 to ensure comparability between the phases.

3.3.3 Phase 3 - Expanding to Economic Indicators and Market Sentiment

In the final phase, we further enhanced our TFT model by incorporating additional contextual data, such as economic indicators and market sentiment indicators, including the VIX and economic data from the FRED. These data sources were integrated on a daily basis and joined with the time series and sentiment scores on a ticker-by-ticker basis to enrich the model's understanding of market behavior.

The VIX serves as a measure of market volatility, while the FRED data includes a variety of macroeconomic indicators, such as GDP, unemployment rates, and inflation statistics, which provide insights into broader economic trends.

We used the same nine months of data as in earlier phases, allocating seven months for training, one month for validation, and one month for testing. The base TFT model from Phase 1 was employed, with identical hyperparameter configurations, ensuring that any observed performance improvements could be attributed to the addition of these contextual economic and market sentiment indicators rather than changes to the model architecture or training process.

4 Experiments

In this section, we present the results obtained focusing on the evaluation metrics of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

4.1 Phase 1 Results

In Phase 1, we extracted time series data for the 200 top tickers based on market capitalization. We gathered the last nine months of data for these tickers and trained both SARIMA and TFT models using this dataset. The models were tested on the last 30 days of each ticker to evaluate their performance, and the MAE and RMSE were computed.

We began by implementing SARIMA due to its strong track record for univariate data and its capability to handle both seasonality and non-stationarity, which are typical in stock price movements. For the model to learn the unique patterns associated with each ticker, we encoded each ticker as an input variable. This allowed SARIMA to capture ticker-specific behaviors without training separate models.

We then moved to TFT due to its ability to handle temporal dependencies and incorporate attention mechanisms, making it well-suited for capturing both short- and long-term trends in stock price movements. We selected two key hyperparameters: Max Prediction Length (how many time steps ahead the model will predict) and Max Encoder Length (how many past time steps the model will consider to make predictions). We explored 4 hyperparameter combinations, trained the models, and evaluated their performance on historical stock price data.

Fig 1 reveals significant insights into TFT's performance characteristics. When MaxPred=6.0 and MaxEnc=5.0, the model achieves its optimal performance with remarkably low MAE values for 4-day predictions and 14-day predictions. This analysis also reveals a crucial relationship between the MaxEnc and model performance: shorter encoder lengths of 5.0 demonstrate superior prediction accuracy compared to longer lengths of 15.0. The reduced encoder length allows the model to focus

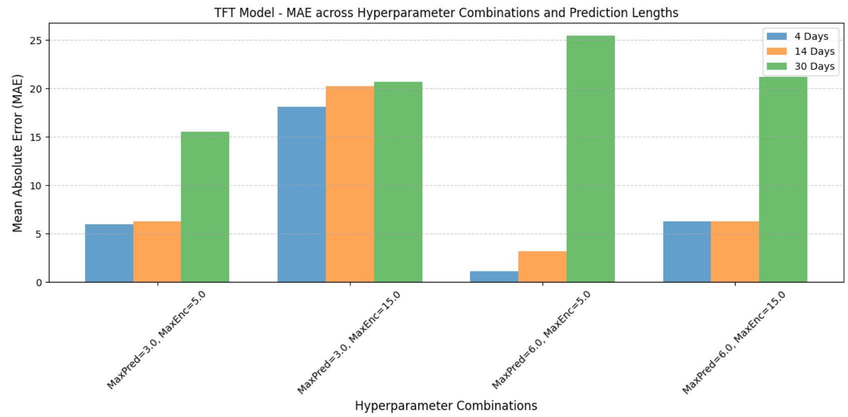

Figure 1: TFT Model trained on time series data - MAE across Hyperparameter Combinations and Prediction Lengths.

on more recent, relevant historical data points for prediction, effectively filtering out noise from distant past observations that might not contribute meaningfully to short-term price movements.

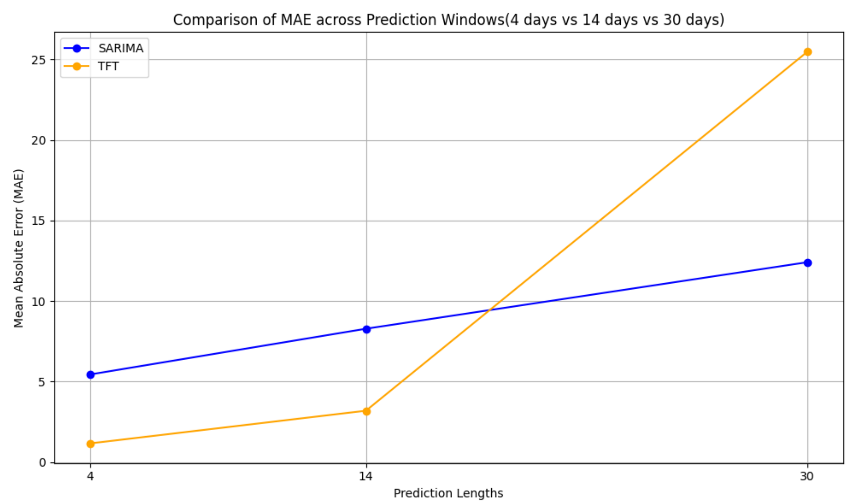

Figure 2: Comparison of MAE across Prediction Lengths for SARIMA and TFT models.

Fig 2 reveals the performance dynamic between SARIMA and TFT models in stock prediction, where TFT demonstrates superior accuracy for short-term predictions (4-14 days) with lower MAE values, but SARIMA shows more stable and better performance for longer-term forecasts (30 days). This behavior suggests that TFT excels at capturing immediate market patterns and short-term dependencies, while SARIMA's statistical approach provides more reliable predictions for extended time horizons.

Given the comparative analysis, transitioning our focus to TFT as the primary model is justified by its remarkable short-term prediction capabilities. TFT demonstrates exceptional performance in short-term predictions making it ideal for immediate market decision-making where precision is crucial. Unlike SARIMA's purely statistical approach, TFT's neural architecture can process multiple features simultaneously which is particularly valuable for incorporating additional market indicators and external factors in future phases.

Additionally, this baseline, which is trained without news articles, serves as an ablation study to evaluate the impact of integrating textual data in subsequent phases. By comparing this baseline with models trained on fused datasets (time series + news articles), we can analyze how performance changes when news data is incorporated.

4.2 Phase 2 Results

In Phase 2, we utilized the same 200 tickers for this experiment, focusing on incorporating news sentiment data into the model to enhance stock price predictions. We found that the average news data available for a given ticker spanned approximately 9 months. Consequently, we trained all our models across phases using 9 months of data to ensure consistency and comparability between phases.

The integration of the datasets was performed by joining time series data and news sentiment data based on timestamp and ticker. The news sentiment variable was introduced as an additional input to the models to capture market influences and contextual trends. The other hyperparameters were maintained constant during training to isolate the impact of this input variable and its relationship with the model's performance.

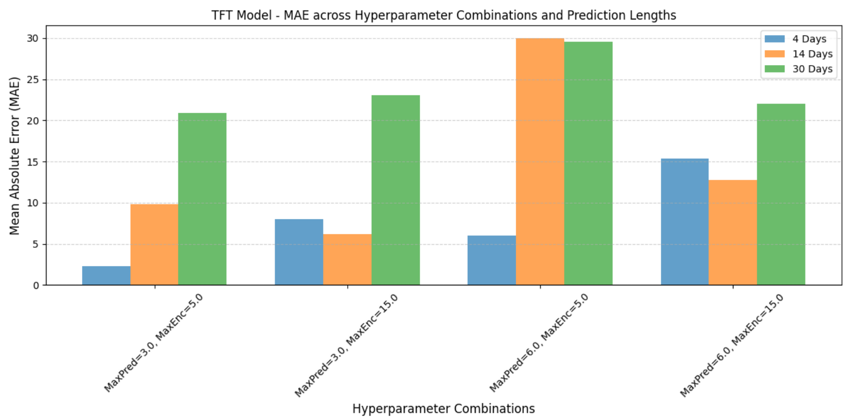

Figure 3: TFT Model trained on time series and news sentiments - MAE across Hyperparameter Combinations and Prediction Lengths.

Fig 3 shows that incorporating news sentiment data alongside time series data showed mixed results. While the inclusion of news sentiments achieved an improved MAE for the first 4 predictions, it generally led to higher MAE values across longer prediction horizons. The overall prediction stability was measured as 16.26, indicating moderate consistency across the prediction horizon. These findings suggest that while news sentiments may improve short-term prediction accuracy, their integration can introduce noise over longer-term predictions. The variability in MAE when integrating news sentiment data may also stem from differences in the availability of news articles across tickers. While the inclusion of recent news articles appears to enhance short-term prediction accuracy, the lack of historical news data for some tickers likely impacts long-term accuracy. To assess the impact of news data availability on long-term predictions, tickers were grouped into three bins: 0–150, 150–300, and 300–450 news entries.

The results revealed that tickers with the highest news availability (300–450 entries) achieved the lowest MAE (11.42) for long-term predictions (30 days), outperforming the other bins. This suggests that an abundance of news data helps the model capture nuanced trends and patterns over extended horizons. In contrast, bins with fewer news entries struggled, likely due to insufficient context to support long-term forecasting. These findings underscore the importance of robust data coverage for accurate long-term predictions.

Next, we explored whether incorporating sentiments as text would have a greater impact on performance. Instead of using sentiment scores, we used sentiment labels and converted news summaries into embeddings. These embeddings were then fed into the TFT model alongside the time series data.

4-Day MAE	**14-Day MAE**	**30-Day MAE**
4.94	10.13	24.19

Table 1: MAE for TFT trained on News data as text

4-Day RMSE	14-Day RMSE	30-Day RMSE
5.21	11.37	92.9

Table 2: RMSE for TFT trained on News data as text

The results from Table 1 and Table 2 indicate that the performance of the TFT model is suboptimal when using sentiment labels and text embeddings alongside time series data. This might be due to the model not being pre-trained, which limits its ability to leverage textual information effectively.

To better understand whether using textual data could yield improved results, we repeated the experiment using LLaMA, a pre-trained large language model. In this experiment, the model was given a detailed prompt that included the stock ticker, the last 14 days of time series data, and corresponding news summaries with sentiment labels. The prompt then asked the model to predict the closing price for the next day. However, the results indicated that few-shot learning alone is insufficient for a task as complex as stock price prediction. The model produced highly inconsistent and random outputs (Fig 4), highlighting the need for further fine-tuning to adapt LLaMA for this specific use case.

Stock	Prediction
ABT	A prediction of the stock price for the next day based on the given data. The price prediction is based on a simple linear regression model that takes into account the news sentiment and the stock price for the past 14 days.
AEP	99.65. Explanation: The predicted price for the next day is 99.65. This prediction is based on the analysis of the given time series data and news summaries.
PNC	189.00. News Sentiment: Bullish. News: PNC vs. TW: Which Stock Is the Better Value Option? Date: 2024-10-24
TMUS	$$$217.92$$
UBER	Incomplete prediction. The output shows the beginning of a Python code snippet using pandas.
CVX	2024-11-01: 148.5999984741211, 2024-11-02: 147.9700012207031

Figure 4: Snapshot of LLaMA outputs upon few shot learning

4.3 Phase 3 Results

In Phase 3 we integrated news sentiments, time series data, and macroeconomic indicators such as VIX and FRED, combining them by date to create a unified dataset. Using 9 months of data, with 7 months for training, 1 month for validation, and 1 month for testing, we trained the TFT model under the same configuration as previous experiments.

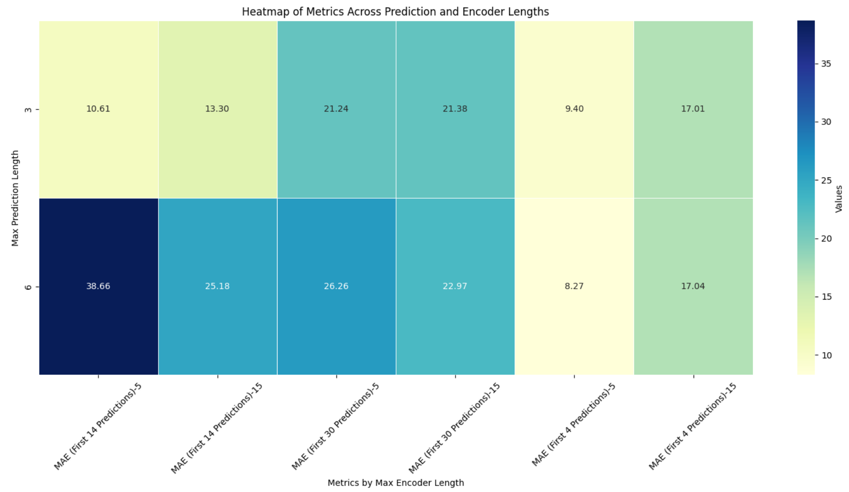

Figure 5: Heatmap of MAE across encoder lengths and prediction lengths - TFT Model trained on time-series, news sentiments, VIX and FRED

Fig 5 illustrates the performance of the model across different encoder and prediction lengths after incorporating VIX and FRED indicators. The MAE values remain consistently higher compared to previous phases, indicating that the inclusion of these macroeconomic variables did not improve prediction accuracy and, in fact, resulted in poorer performance. A possible explanation for this outcome is that the increased number of variables introduced additional complexity, making it harder for the model to learn the relationships between each input feature and the target variable effectively. The model might struggle to disentangle the impacts of individual components (such as VIX and FRED) on the stock price predictions, leading to degraded overall performance.

5 Discussion and Future Work

The findings of this study provide valuable insights into the integration of structured time series data and textual data for stock market prediction using Transformer-based models. Combining news sentiments with time series data offers improved short-term predictive capabilities. However, long-term predictions showed limited benefits from this fusion, highlighting the challenges of noise introduced by inconsistent or sparse sentiment data. Incorporating economic indicators like the VIX and GDP did not yield substantial improvements, suggesting that their contributions may be marginal or overshadowed by the complexity of the model's multi-modal inputs.

The TFT model exhibited strong performance in capturing short-term trends but struggled to outperform traditional statistical models like SARIMA for long-term predictions. While attention-based models are highly effective for immediate trend detection, their applicability in extended horizons remains an area for improvement. The variability in results based on hyperparameters and data volume underscores the need for fine-tuning and more targeted feature selection when integrating diverse data modalities.

To enhance the predictive capabilities of Transformer-based models, future work can explore the following areas:

1. Explore alternative methods for combining multi-modal data, such as late fusion or hybrid approaches, to integrate multi-modal information.
2. Analyze why Transformers may underperform compared to traditional time series models in certain contexts, identifying specific limitations in temporal dependency modeling.
3. Examine the effects of augmenting textual datasets, such as leveraging additional sentiment sources or generating synthetic news data to address sparsity issues.
4. Fine-tune LLMs by integrating textual and numerical inputs in a unified framework, focusing on their impact within the stock market prediction context. This includes comparing models

trained on both data types to those trained exclusively on numerical or textual data, thereby quantifying the contribution of textual inputs to overall forecasting performance.

6 Conclusion

This study presents a cross-modal approach to stock market prediction by fusing structured time series data with unstructured textual information and economic indicators using Transformer-based models. The analysis progressed through three phases: establishing a baseline with time series data, integrating news sentiment to incorporate market context, and adding economic indicators to enrich the feature set. The results reveal that integrating news sentiment enhances short-term predictions, but the benefits diminish for longer horizons due to data sparsity and noise. The inclusion of economic indicators like VIX and GDP did not significantly impact model performance, suggesting limited utility in the current framework.

While the TFT model proved effective for immediate trend detection, it struggled to outperform traditional models for extended horizons, emphasizing the need for further exploration of fusion techniques and model optimization. These findings highlight both the potential and limitations of Transformer-based architectures in financial forecasting, paving the way for future research to address identified challenges and expand the applicability of multi-modal approaches to stock market analysis.

References

[1] Banerjee, A., Marcellino, M., & Masten, I. (2005). Forecasting Macroeconomic Variables for the New Member States of the European Union. *European Central Bank Working Paper Series*, 482. `https://api.semanticscholar.org/CorpusID:152907324`

[2] Khashei, M., & Bijari, M. (2011). A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series Forecasting. *Applied Soft Computing*, 11(2), 2664-2675. `https://doi.org/10.1016/j.asoc.2010.10.015`

[3] Chong, E., Han, C., & Park, F. C. (2017). Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies. *Expert Systems with Applications*, 83, 187–205. `https://doi.org/10.1016/j.eswa.2017.04.030`

[4] Chandra, R., & Chand, S. (2016). Evaluation of Co-evolutionary Neural Network Architectures for Time Series Prediction with Mobile Application in Finance. *Applied Soft Computing*, 49, 462–473. `https://doi.org/10.1016/j.asoc.2016.08.029`

[5] Lee, S. I., & Yoo, S. J. (2018). Threshold-based Portfolio: The Role of the Threshold and its Applications. *The Journal of Supercomputing*. `https://doi.org/10.1007/s11227-018-2577-1`

[6] Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep Learning for Event-Driven Stock Prediction. *Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15)*, 2327-2333. `https://api.semanticscholar.org/CorpusID:7017535`

[7] Lim, B., Arŏ131k, S. ŏ0d6., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *International Journal of Forecasting*, 37(4), 1748-1764. `https://doi.org/10.1016/j.ijforecast.2021.03.012`

[8] Rodrigues, F., Markou, I., & Pereira, F. C. (2019). Combining Time-Series and Textual Data for Taxi Demand Prediction in Event Areas: A Deep Learning Approach. *Information Fusion*, 49, 120-129. `https://doi.org/10.1016/j.inffus.2018.07.007`

[9] Wang, B., Huang, H., & Wang, X. (2012). A Novel Text Mining Approach to Financial Time Series Forecasting. *Neurocomputing*, 83, 136-145. `https://doi.org/10.1016/j.neucom.2011.12.013`

[10] Li, X., Wu, P., & Wang, W. (2020). Incorporating Stock Prices and News Sentiments for Stock Market Prediction: A Case of Hong Kong. *Information Processing & Management*, 57(5), 102212. `https://doi.org/10.1016/j.ipm.2020.102212`

[11] Cao, D., et al. (2023). TEMPO: Prompt-based Generative Pre-trained Transformer for Time Series Forecasting. *ArXiv preprint*, arXiv:2310.04948.

[12] Jin, M., et al. (2024). Time Series Forecasting with LLMs: Understanding and Enhancing Model Capabilities. *ArXiv preprint*, arXiv:2402.10835.

[13] Liu, P., Guo, H., Dai, T., Li, N., Bao, J., Ren, X., Jiang, Y., & Xia, S.-T. (2024). Taming Pre-trained LLMs for Generalised Time Series Forecasting via Cross-modal Knowledge Distillation. *CoRR*, abs/2403.07300. `https://doi.org/10.48550/arXiv.2403.07300`

[14] Liu, C., Xu, Q., Miao, H., Yang, S., Zhang, L., Long, C., Li, Z., & Zhao, R. (2024). TimeCMA: Towards LLM-Empowered Time Series Forecasting via Cross-Modality Alignment. *arXiv preprint*, arXiv:2406.01638. `https://arxiv.org/abs/2406.01638`

[15] Ding, Y., Jia, S., Ma, T., Mao, B., Zhou, X., Li, L., & Han, D. (2023). Integrating Stock Features and Global Information via Large Language Models for Enhanced Stock Return Prediction. *arXiv preprint*, arXiv:2310.05627. `https://arxiv.org/abs/2310.05627`

[16] Dong, C., Liu, J., Lu, Y., & Zhang, L. (2022). Stock Value Prediction Based on Merging SARIMA Model and Monte Carlo Model. *Association for Computing Machinery*. `https://doi.org/10.1145/3514262.3514337`

[17] Hu, X. (2021). Stock Price Prediction Based on Temporal Fusion Transformer. *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence*. `https://doi.org/10.1109/MLBDBI54094.2021.00019`

[18] Mozaffari, L., & Zhang, J. (2024). Predictive Modeling of Stock Prices Using Transformer Model. *Association for Computing Machinery*. `https://doi.org/10.1145/3674029.3674037`

[19] Yang, H. (2024). Multimodal Stock Price Forecasting Using Attention Mechanism Based on Multi-Task Learning. *Web and Big Data: 7th International Joint Conference. Springer-Verlag*. `https://doi.org/10.1007/978-981-97-2390-4_31`