Improving Cryptocurrency Forecasting with Media Sentiment and Dual-Stream Temporal Fusion Transformer Model

Min-Seok Choi, Jea-Ryong Lee, Hae-Jong Joo, Woong Choi
Kangnam University, Republic of Korea
minseok011017@kangnam.ac.kr, leejearyong@gmail.com, hjjoo@kangnam.ac.kr, wchoi@kangnam.ac.kr

Abstract. The cryptocurrency market poses significant challenges for accurate price prediction due to its inherent volatility. Prior studies have typically relied on technical indicators or employed simplistic fusion of sentiment data, resulting in limited improvements in both predictive accuracy and model interpretability. To overcome these limitations, this study proposes a Dual-Stream Temporal Fusion Transformer (TFT) that separately processes technical price indicators and sentiment data from social media and news sources, followed by a dynamic gating mechanism to integrate the two streams. Experimental results show that while the validation Mean Absolute Error (MAE) of the Dual-Stream model was 0.47% higher than that of the single-stream model, it achieved superior performance on the test dataset, reducing the MAE by 2.31%. This demonstrates the model's effectiveness in enhancing generalization through dynamic fusion of heterogeneous data sources. Furthermore, attention-based analysis revealed clearly differentiated temporal patterns across the price and sentiment streams, thereby improving model interpretability. Overall, this study provides a foundational framework for future research on advanced data fusion strategies and offers practical insights for real-world applications in cryptocurrency market prediction.

1 Introduction

The cryptocurrency market, characterized by its high volatility and inherent unpredictability, presents a major challenge for investors. Consequently, enhancing the accuracy and generalization performance of AI-based price prediction models has emerged as a critical issue. In particular, Bitcoin(BTC) exhibits behavioral patterns distinct from traditional financial assets, necessitating novel modeling approaches for effective forecasting. Within this context, recent studies have increasingly focused on leveraging diverse data sources and incorporating Transformer-based architectures into time series prediction tasks.

Among such models, the Temporal Fusion Transformer(TFT) has demonstrated promising results by dynamically suppressing irrelevant features and emphasizing informative variables through variable selection networks and gating mechanisms. In addition, the attention weights of TFT can be visualized, thereby offering both interpretability and predictive strength [1].

Meanwhile, collective investor sentiment—captured through social media and news—has been recognized as a significant factor influencing cryptocurrency prices. Sentiment data extracted from these sources reflects the underlying market psychology and can be used to improve predictive accuracy when integrated into forecasting models. Prior studies have attempted to incorporate such sentiment data alongside price indicators using a single-stream architecture. However, such approaches fail to fully capture the distinct characteristics of heterogeneous data types, resulting in limitations in both predictive performance and model interpretability.

To overcome these shortcomings, this study proposes a Dual-Stream TFT model that independently processes technical price indicators and sentiment data from social media and news. By learning each data stream separately and subsequently combining them via a fusion mechanism, the model aims to achieve more nuanced and accurate forecasts of BTC prices. Furthermore, the use of attention mechanisms allows for temporal importance analysis within each stream, thereby enhancing the interpretability of the predictive model.

2 Related Work and the Proposed Dual-Stream TFT

The cryptocurrency market is highly sensitive to investor sentiment and external emotional stimuli, prompting extensive research into sentiment-based price prediction. Prior studies have demonstrated that investor emotions expressed on Twitter have a statistically significant impact on Bitcoin prices [2]. Additionally, other approaches have utilized sentiment data derived from user-generated content (UGC) to enhance cryptocurrency price forecasting [3].

However, most of these studies adopt a single-stream architecture that merges price and sentiment indicators into a unified input. This approach tends to dilute the distinct characteristics of each data modality and limits the model's ability to interpret cross-modal interactions independently. To address this structural limitation, we propose a Dual-Stream TFT architecture that independently encodes technical price indicators and sentiment signals using separate TFT encoders. The encoded representations are then dynamically fused through a gating network.

This architecture allows each stream to learn modality-specific temporal representations, improving the model's generalization capability. Moreover, the gating mechanism enables interpretability by allowing the model to visualize the relative contribution of each stream at different time steps. In doing so, the proposed Dual-Stream TFT aims to simultaneously enhance both prediction performance and model interpretability.

3 Methodology and Model Architecture

3.1 Data Collection and Preprocessing

In this study, we constructed an integrated dataset using multiple data sources at hourly intervals. Bitcoin (BTC) OHLCV time series data were collected from the Binance API. Social sentiment scores (sentiment_balance_total) and their temporal rate of change (soc_sr) were obtained from the Santiment API, while news sentiment scores (ticker_sentiment_score) and their corresponding rate of change (news_sr) were gathered from the Alpha Vantage API. All data were aligned and merged based on Coordinated Universal Time (UTC).

To enhance the feature set, we computed technical indicators such as the moving average (MA) and relative strength index (RSI) based on closing prices. A half-life exponential decay accumulation was applied to the news sentiment scores to model sentiment persistence, using a four-hour half-life. For periods where both observed and decayed sentiment values were zero, the values were retained as zero to explicitly represent sentiment-absent intervals.

To address differences in scale and reduce noise across data sources, we calculated the first-order rate of change (sr) for both social and news sentiment scores. Any missing values resulting from the calculation of MA, RSI, or rate of change were removed. The final dataset was split into training, validation, and test sets in a 70:15:15 ratio for model development and evaluation.

3.2 Model Overview

The proposed Dual-Stream TFT architecture consists of two independent TFT encoders that separately process the price stream (price_feats) and sentiment stream (sent_feats). The outputs from these two encoders are dynamically fused through a gating network to generate the final prediction.

a) **Price stream**: Includes historical BTC price data and technical indicators such as MA and RSI, fed into a TFT encoder.

b) **Sentiment stream**: Includes social sentiment scores, their rate of change, news sentiment scores, and their rate of change, also fed into a separate TFT encoder.

c) **Gated fusion**: The outputs of the price and sentiment encoders (denoted as yp and ys, respectively) are concatenated and passed through a gating module (either a linear layer or a multi-layer perceptron) to compute a fusion weight:

$$\alpha=\sigma(\mathrm{gate}([yp,ys]))$$

The final prediction is computed as a weighted sum:

$$\hat{y}=\alpha\cdot yp+(1-\alpha)\cdot ys$$

This architecture enables modality-specific representation learning while offering interpretability via the learned gate weights, which indicate the contribution of each stream at different time steps.

3.3 Experimental Scenarios

To evaluate the performance of the model, we defined the following two experimental scenarios. For each scenario, the model was trained and subsequently analyzed in terms of both predictive performance and interpretability.

a) **Scenario 1**: A single-stream TFT model combining both price and sentiment features as a unified input.

b) **Scenario 2**: The proposed Dual-Stream TFT model that processes price and sentiment features in separate streams.

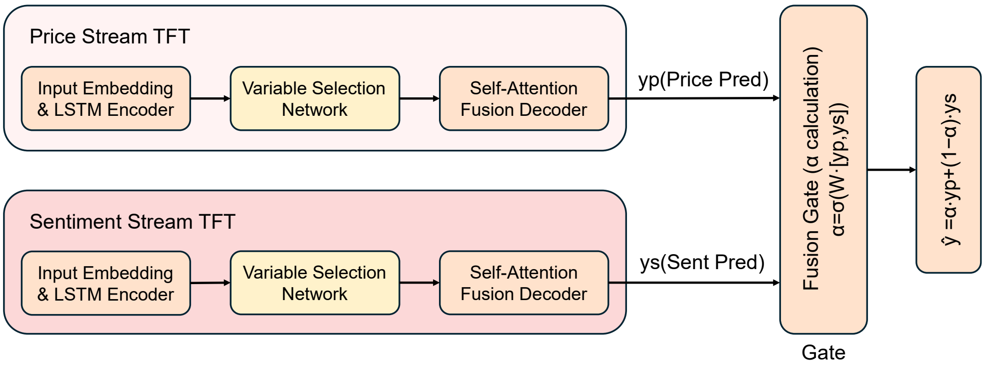


Figure 1. Proposed Model System Architecture

3.4 Experimental Environment and Settings

Model hyperparameters were optimized using Optuna, a framework for automated hyperparameter search. Key hyperparameters included batch size, hidden layer size, dropout rate, and learning rate. Separate optimization runs were performed for each experimental scenario to ensure the best configuration for each model variant.

All experiments were conducted using Python 3.10, PyTorch 2.1.0 (CUDA 11.8), and Lightning 2.5.1, with random seed fixed at 42. However, slight variations may occur due to nondeterminism in LSTM models and GPU operations.

4 Experimental Results and Analysis

4.1 Performance Comparison and Quantitative Evaluation

In this study, we conducted a quantitative evaluation of the TFT across the two experimental scenarios described earlier. Predictive performance was assessed using three metrics: Mean Absolute Error(MAE), Root Mean Squared Error(RMSE), and prediction volatility. These metrics were computed on both the validation and test datasets to ensure robust evaluation of the model's accuracy and generalization capability.

Table 1. Performance Comparison by Scenario

Scenario	Val_MAE	Val_RMSE	Val_Vol	Test_MAE	Test_RMSE	Test_Vol
Scenario 1	319.2764	485.3020	483.9223	358.3081	568.6622	562.9303
Scenario 2	320.7794	484.1341	483.5529	350.0146	555.6910	555.5461

As a result of the performance evaluation, the Dual-Stream TFT model (Scenario 2), which leverages both technical price indicators and sentiment data, achieved the lowest MAE, RMSE, and volatility scores on the validation set. However, it recorded the highest performance on the test set across all metrics. These findings indicate that the dual-stream architecture contributes meaningfully to improving both the prediction accuracy and generalization capability of the TFT model.

4.2 Encoder Attention Analysis

Figures 1 through 3 illustrate the encoder attention distributions for each experimental scenario:

a) **Figure 2** : The attention appears more dispersed across early (0–2) and mid-range time steps, yet lacks a consistent or interpretable pattern, indicating unstable temporal focus within the single-stream structure that includes sentiment data.

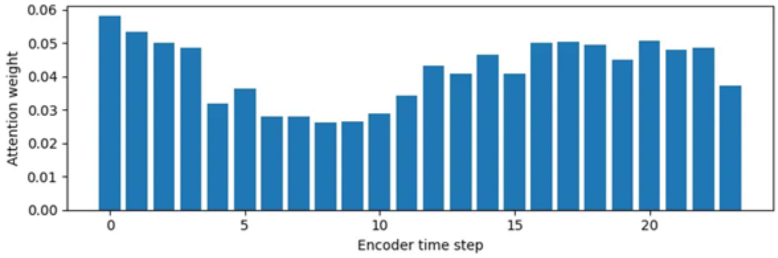
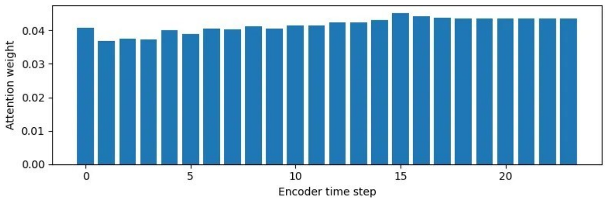
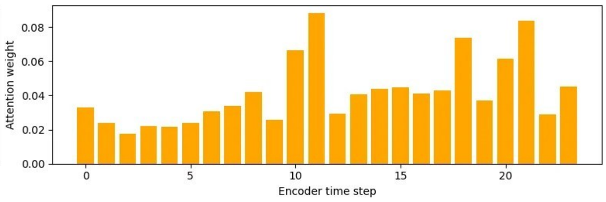


Figure 2. Encoder attention distribution in Scenario 1

b) **Figure 3** : The Dual-Stream encoder attention distribution in Scenario 2 reveals distinct temporal focus across the two streams. In the price stream (a), the attention weights are relatively evenly distributed but show a gradual increase toward the later time steps (15–23), indicating that recent price information played a more prominent role in the model's predictions. In contrast, the sentiment stream (b) exhibits sharp attention peaks at time steps 10, 17, and 21, visually confirming that sentiment signals from news and social media at these points had a strong influence on the prediction outcome.



(a)



(b)

Figure 3. Encoder attention distribution in Scenario 2 (a: price stream, b: sentiment stream)

5 Conclusion and Future Work

The proposed Dual-Stream TFT model demonstrated marginal differences on the validation set compared to the single-stream baseline (Scenario 1), with a 0.47% increase in MAE and a 0.24% decrease in RMSE. However, on the test set, the Dual-Stream model outperformed the baseline, showing a 2.31% reduction in MAE and a 2.28% reduction in RMSE. Moreover, the standard deviation of prediction errors (volatility) decreased by 1.31%, indicating improved generalization and greater stability in real-world scenarios.

Encoder attention analysis further supports these findings: the price stream exhibited a flattened distribution with gradually increasing attention toward recent time steps, alleviating mid-sequence overemphasis. Meanwhile, the sentiment stream revealed distinct attention peaks during the mid-to-late sequence, highlighting temporally critical points that significantly contributed to the prediction. These results enhance the model's interpretability by uncovering the temporal importance of each data modality.

Future research will focus on enhancing the dynamic fusion mechanism and integrating various exogenous variables as additional input streams to evaluate the scalability and robustness of multi-modal prediction frameworks.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) under Grant RS-2022-NR-069955(2022R1A2C1092178).

References

1. B. Lim, S. Ö. Arik, N. Loeff, and T. Pfister: Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting, arXiv preprint arXiv:1912.09363 [cs.LG]. (2019)
2. G. B. Haritha and N. B. Sahana: Cryptocurrency Price Prediction using Twitter Sentiment Analysis, arXiv preprint arXiv:2303.09397 [cs.CY]. (2023)
3. D. Perera, J. Lim, S. Gunraku, and W. H. Lim: Price Prediction of Cryptocurrency Using User Sentiments and Quantitative Data, in *Proc. 2024 Int. Conf. Data Sci. Adv. Analytics*, (2024)