A Novel Sentiment Correlation-based Method with Dual Transformer Model for Stock Price Prediction

Qizhao Chen
brandonchan787@gmail.com

University of Hyogo

Hiroaki Kawashima
University of Hyogo

Research Article

Keywords: time series forecasting, sentiment analysis, sentiment correlation, Transformer, Named Entity Recognition

Posted Date: May 12th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6479946/v1

Additional Declarations: No competing interests reported.

A Novel Sentiment Correlation-based Method with Dual Transformer Model for Stock Price Prediction

Qizhao Chen* and Hiroaki Kawashima

Graduate School of Information Science, University of Hyogo, Kobe, 651-2197, Japan.

*Corresponding author(s). E-mail(s): af24o008@guh.u-hyogo.ac.jp ;
Contributing authors: kawashima@gsis.u-hyogo.ac.jp;

Abstract

This paper proposes a novel Dual Transformer model to predict the target company's stock price using news sentiment from related companies. The Dual Transformer comprises two transformer structures, the Enhancement Transformer and the Forecast Transformer. The Enhancement Transformer is used to enhance the correlation strengths between related companies, while the Forecast Transformer is used for stock price forecasting. The polarity scores of these related companies, along with historical close prices, are used to forecast the next day's stock close price. Experimental results demonstrate that incorporating the news sentiment of related companies improves the prediction of the target company's stock price when using the proposed Dual Transformer model. In addition, the Dual Transformer model can outperform existing models such as Informer and LSTM.

Keywords: time series forecasting, sentiment analysis, sentiment correlation, Transformer, Named Entity Recognition

1 Introduction

Stock price prediction has long been a topic of great interest for financial analysts, investors, and researchers. Conventionally, these forecasts are based on historical stock performance, financial statements, and macroeconomic indicators. However, with the increasing availability of textual data, particularly news articles, sentiment analysis has become a valuable method for stock price prediction. The sentiments embedded in news articles often provide important information on market perceptions and investor behavior, which can affect stock movement.

Although recent studies have focused on investigating the sentiment in the news directly related to a target company [1–6], some research suggests that strong media sentiment toward a company, often caused by a significant real-world event, could have a spillover effect on related companies [7]. Therefore, the stock price of a company can also be affected by news related to its connected companies, such as competitors in the same industry or sector. In reality, no company can operate in isolation, and it forms a broader network of suppliers, customers, competitors, and business partners. Events that influence one company can also have an impact on others within the network, which can be used to predict the stock price.

Furthermore, incorporating sentiment analysis into stock price prediction models can significantly improve the precision of prediction [8–10]. However, effectively leveraging news sentiment poses challenges, particularly when there is minimal news coverage about target companies, as is often the case with stocks that have low trading volumes. In such situations, it becomes essential to consider alternative data sources, such as news about companies that are related to or have significant business ties to the target firms. Thus, exploring these indirect sentiment sources and developing methods to integrate them into predictive models is crucial to improving the reliability of forecasts.

Moreover, research on sentiment-driven stock price prediction and sentiment linkages between different companies has significant real-world implications. For example, in portfolio management, identifying sentiment linkages between companies can improve risk assessment and inform diversification strategies, reducing overexposure to highly correlated assets. In risk management, tracking sentiment among related companies can act as an early warning system for sector-wide or market-wide risks, helping mitigate potential losses.

When it comes to prediction models in stock price forecasting, traditionally, some frequently used models include recurrent neural network (RNN) (e.g., long-short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM)) [11–13], Convolutional Neural Network (CNN) [14, 15], Support Vector Machine (SVM) [16–19] and ensemble methods such as Random Forest [20]. The limitations of some classical machine learning models, such as SVM, are evident. These models fail to capture the complicated temporal and dynamic relationships within the stock price data. Although RNNs excel at capturing temporal dependencies, they also have limitations in stock price prediction. For example, RNNs struggle with highly volatile or noisy data, which is common in stock prices, potentially leading to over-fitting or reduced accuracy.

Currently, due to its outstanding performance in sequential tasks, Transformer becomes a popular candidate in stock price prediction. The Transformer model, proposed by [21], is a deep learning architecture widely used in natural language processing (NLP) and other sequential tasks, including time series forecasting. For example, some Transformer-based models have been shown to be able to outperform some classical models such as LSTM in stock price prediction [22–24]. Unlike traditional models such as RNNs that process sequences sequentially, Transformer uses self-attention mechanisms to handle tokens simultaneously, which allows it to capture relationships between words within the text, making it highly effective in understanding context and meaning across long sequences. This aspect of the Transformer model inspires us to use it to capture the relationships between companies based on their sentiment. However, in this research, the effectiveness of Transformer in capturing the relationship between companies needed to be further investigated.

This paper aims to explore whether the news sentiment of related companies can improve the prediction of the stock price of a target company with the proposed Dual Transformer model. The main contributions of this research are:

1. We propose a framework by incorporating news sentiment of related companies in stock price prediction and demonstrate that news sentiment of related companies offer predictive value for the target company's stock price movement when using the proposed model.
2. We propose a novel Dual Transformer model to increase the correlation strengths between related companies and enhance the performance of stock price prediction. We show that the Dual Transformer model can increase the correlation strength between companies and outperform models such as Informer and LSTM.
3. We introduce an innovative use of Named Entity Recognition (NER) to identify related companies, complemented by methods such as SHAP (Shapley Additive Explanations) to refine the related companies further.

The remainder of this paper is structured as follows. Firstly, related work is listed (Section 2). The methodology is then described (Section 3). In Section 4, the experimental results are presented. Finally, we conclude with Section 5.

2 Related Work

The use of sentiment analysis in stock price prediction has attracted growing attention in recent years, driven by the increasing availability of textual data and the rapid development of NLP. This review of the literature summarizes key contributions in these areas, focusing on news impact on stock price and sentiment analysis on stock price prediction.

There are almost no studies related to sentiment correlation of related companies, which may be due to the fact that the companies chosen in most research are well-known companies locally or globally, such as Google and Microsoft, and there are tons of news or comments about these companies online. Therefore, alternative data are not needed in these cases. Currently, most of the research still focuses on the application of the sentiments of target companies to their stock price forecasting.

Research on the application of sentiment analysis in stock price forecasting has explored various methods and data sources. Several studies focus on using social media sentiment for prediction. For example, Batra, Rakhi, and Daudpota [18] applied SVM to predict the sentiment of tweets about Apple from StockTwits, demonstrating a positive relationship between sentiments and stock price. Similarly, Nguyen, Thien, and Shirai [17] proposed Topic Sentiment Latent Dirichlet Allocation (TSLDA) to analyze social media sentiments, finding that incorporating sentiment data improves stock price predictions. Pagolu et al. [20] also analyzed tweets, combining human-annotated and machine-classified sentiments, and observed a strong correlation between Twitter sentiment and stock price movements.

Other research has explored the sentiment of financial news and its predictive power. Khedr, Ayman, and Yaseen [25] used Naive Bayes to extract sentiment polarity from financial news and combined it with historical stock prices, achieving an accuracy of up to 89.80%. Mehta, Pandya, and Kotecha [16] examined how public opinions in financial news influence stock prices. They used machine learning models, including Naive Bayes, Maximum Entropy (ME), SVM, and MLP, and found that LSTM was particularly effective for stock price prediction.

A comparative study by Li et al. [19] evaluated the performance of sentiment analysis against other approaches such as bag-of-words and sentiment polarity. They found that sentiment analysis models outperformed bag-of-words approaches at stock, sector, and index levels, while sentiment polarity models offered limited predictive value.

In terms of hybrid approaches, Srijiranon et al. [26] proposed a PCA-EMD-LSTM model that incorporates news sentiment and technical indicators to predict the next day's stock price. Their findings revealed that integrating news sentiment analysis significantly enhances the performance of LSTM-based models.

3 Methodology

Figure 1 provides an overview of the research setup presented in this paper.

There are basically two parts to perform the analysis. Part 1 is the related company selection. The definition of a related company is broad. However, in this research paper, we define the related companies as the companies which frequently co-occur in news articles concerning the target company because if a company frequently appears in news articles about the target company, it is likely that the two companies are connected. By applying Named Entity Recognition (NER), we can extract the companies from the news articles of the target company. The proposed methods for identifying related companies are more user-friendly for individuals with limited knowledge of finance. Unlike traditional fundamental analysis, such as reviewing financial statements, which can be time consuming, our method offers a more efficient alternative. Furthermore, if users are unfamiliar with certain companies, especially those from other countries, traditional analysis may require significant time and effort, making our approach even more advantageous.

Part 2 is the stock price prediction. We mainly use two types of feature to forecast the stock price. The first type is historical close price and the other type is polarity score of companies. We propose a Dual Transformer model to capture the interaction between companies and to forecast the stock price. To evaluate the effectiveness of this Dual Transformer model, we compare its performance against several deep learning models, such as Informer, and traditional machine learning models, including Random Forest. In addition, to investigate whether the number of companies used in the prediction affects the results, we design an experiment featuring various scenarios. The details of this experiment will be discussed in Section 3.6.

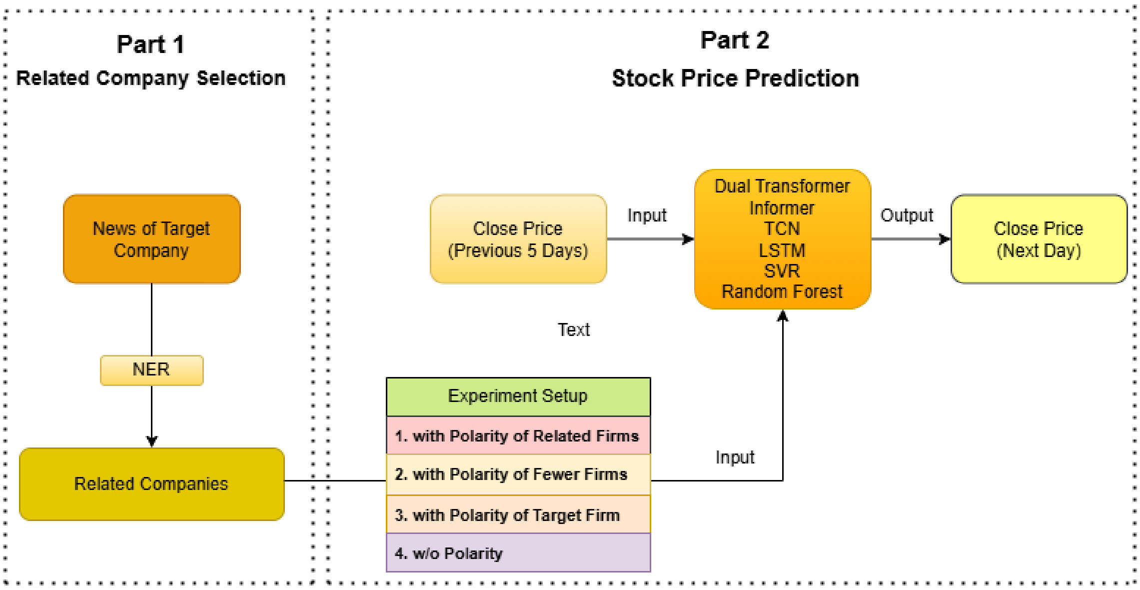


Fig. 1: Full Picture of Research Setup

3.1 Data

There are two types of data used in this research, news data and stock price data. The financial news data of companies was downloaded using the News Feed and News Sentiment data API [1] from websites such as Yahoo News, InvestorPlace and GlobeNewswire. The collected data table mainly includes the columns "date", "title", "content", "link", "symbols", and "sentiment". The "sentiment" column contains the polarity score of the corresponding news, which is calculated based on negative and positive mentions in texts normalized with possible values between -1 and 1. In sentiment analysis, the polarity score quantifies the sentiment expressed in a text and ranges from -1 to 1, where negative values represent negative sentiment, positive values represent positive sentiment and values close to 0 suggest neutral sentiment. Since multiple pieces of news are released each day, we use the average polarity score for each day to represent the overall sentiment of that day. These polarity scores will serve as input for predicting stock prices in this research paper.

[1] https://eodhd.com/financial-apis/stock-market-financial-news-api

For stock price data, the Python library yfinance helps import the stock price of companies from Yahoo Finance. The imported data mainly contain five features, "Open", "Close", "High", "Low", and "Volume". Only the close price ("Close"), the stock price at the end of each trading day, is used for prediction because it is often considered to be the most important price of the day and it is the final consensus of the stock value among traders and investors. Based on related research in stock price prediction [4, 5, 8, 9, 27], the stock prices of four to six companies are typically included in the analysis. In this paper, six companies' stock price from 01/05/2015 to 08/05/2024 are used, which include Apple (AAPL), HSBC (0005.HK), InfuSystem Holdings (INFU), Toyota (7203.T), Tencent (0700.HK) and Pepsi (PEP).

The selected companies represent diverse financial markets and sectors. Apple is an American company renowned for its innovative software and electronic products. HSBC is a British banking and financial services company headquartered in London. InfuSystem Holdings is an American healthcare company specializing in integrated therapy services and medical equipment management. Toyota, based in Japan, is one of the largest automotive manufacturers in the world. Tencent, a technology company headquartered in Shenzhen, China, owns business in social networks, gaming, and entertainment. PepsiCo is an American company with a strong presence in the food and beverage industry.

Historical close prices and polarity scores, using a sliding window of 5 days, will be utilized to predict the close price for the following day. 70% of the data (2015-2021) will be used for training and 30% of the data (2021-2024) will be used for testing. The MinMax Scaler will be applied to scale the feature values between 0 and 1 to avoid features with larger values dominating the prediction.

3.2 Methods to Identify Related Companies

In NLP, NER involves identifying and classifying entities such as organizations, people, dates, and other specific terms within a text. Through entity tagging, NER helps structure unstructured data for better data analysis. In Python, the NLP library spaCy makes it easy to perform NER tasks and the inputs are news articles of companies. In this paper, as shown in Figure 2, the entity names are extracted from news articles. The companies are then selected from the list of entity names based on their frequency of appearance in news articles, ranked from highest to lowest.

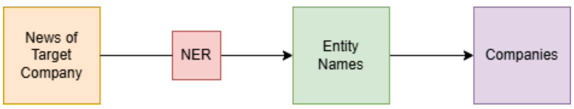


Fig. 2: NER Procedure

Furthermore, in one of our experimental settings in the prediction of stock prices, we need to reduce the number of related companies. Here, we will use the Shapley value to measure the feature importance of the polarity scores from different companies and choose only the top five with the highest scores.

The Shapley value (Equation 1), which originates from cooperative game theory, is used to measure the contribution of players in a coalition. Shapley value of a feature is its contribution to the payout. For example, in machine learning, the "players" are the features and the "payout" is the model's prediction. SHAP (SHapley Additive exPlanations), introduced by Lundberg and Lee in 2016 [28], is a framework and methodology to quantify the contributions of the features of machine learning models based on the idea of Shapley value. The Python implementation of SHAP is available. In this research, the SHAP values are generated using the Random Forest with SHAP's TreeExplainer in Python. This choice is based on experimental results that show minimal differences in feature rankings derived from SHAP values between different models.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \tag{1}$$

where ϕ_i is the Shapley value for the feature i, S is a subset of features that do not include i, N is the set of all features, $|S|$ is the size of the subset S, $f(S)$ it is the prediction of the model using only the features in subset S, $f(S \cup \{i\})$ is the prediction of the model using the features in subset S plus the feature i.

3.3 Prediction Model

Inspired by [29], where a Dual-path Actor Interaction (DualAI) framework is proposed to model group activities in sports, in this paper, we propose a Dual Transformer model to model interaction between companies and predict stock prices. As the name suggests, this model connects two transformers (Figure 3), the Enhancement Transformer (**ET**) and the Forecast Transformer (**FT**). The reason for using this model structure is to effectively process both feature extraction and prediction tasks. The ET focuses on capturing key relationships and patterns within the news sentiment. The FT focuses on mapping the refined features of the ET to accurate stock price predictions. Generally speaking, this model structure separates the responsibilities of feature extraction and prediction, allowing each transformer to specialize in its respective task.

As shown in Figure 4, the ET consists only of an encoder (without positional encoding), while FT (Figure 5) includes both an encoder and a decoder, similar to the vanilla Transformer model [21].

3.3.1 Enhancement Transformer

The ET takes in the news sentiment of companies as input, and output sentiment features with the same dimensions as the input data. The ET is trained as an autoencoder, as explained later in Section 3.4.

To find an optimized way to feed the data to the ET, we investigated how the input configurations of the Enhancement Transformer impact its performance in extracting meaningful features from company polarity scores. Specifically, letting W be the sliding window size, N be the number of companies, and a tuple (d, n) denote the input shape to the model, where d is the dimension of an input vector and n is the number of input vectors to the Transformer, we tested two different input configurations: $(d, n) = (W, N)$ and $(d, n) = (N, W)$.

In the first configuration $(d, n) = (W, N)$, as illustrated in Figure 4, each input vector consists of the polarity scores of one company over a given time period W. Before passing this into the attention layers, we first project the sliding window size W into a hidden representation of size $d_{model} = 32$, resulting in an input shape of (d_{model}, N).

This approach enables the model to emphasize inter-company dependencies, capturing how sentiment trends across different companies interact and influence one another. This is particularly relevant for industries where companies exhibit strong correlations due to market conditions, sector-wide trends, or shared economic factors.

For the second configuration $(d, n) = (N, W)$, the input is structured such that the polarity scores of different N companies are combined as an input vector and sequential W input vectors are prepared across the time window W. In this case, we project N dimensions into hidden size d_{model}, transforming the input into (d_{model}, W) before feeding it into the attention layers.

This transformation allows the model to focus on capturing temporal dependencies within each company's sentiment trends while preserving company-wise feature relationships. The Transformer then processes this representation to extract meaningful patterns in sentiment evolution and its potential impact on stock price movements.

In the first configuration, the number of companies N does not need to be fixed in advance when training the model, allowing for greater flexibility. This is because the Transformer operates along the temporal dimension (window size W), treating each company as an independent sequence input. Since the model learns to process sequences of a fixed length W, and not across companies, it can flexibly handle varying numbers of companies at inference time. This flexibility is important, as the number of related companies can vary depending on the target company. Moreover, in one of our experimental setups, we intentionally reduce the number of companies to investigate how this affects the model's performance.

Additionally, this setup enhances the explainability of the model, as the goal of the Enhancement Transformer is to capture the inter-company relationships and interactions more effectively. By structuring the input such that each vector represents the polarity scores of a single company across the sliding window, this configuration naturally focuses on learning how different companies influence each other, achieving exactly what the Enhancement Transformer is designed to do.

Based on the reasons mentioned above, we choose to continue with the first configuration. Specifically, we denote $X \in \mathbb{R}^{W \times N}$ as the vectors of N companies with W time steps. ET begins by passing the input X into a multi-head attention mechanism (MultiHeadAttention), which captures the dependencies between different companies H_1 (Equation 2). Since companies have no ordering relation, there is no positional encoding. A residual connection is applied by adding H_1 and X, followed by layer normalization (LayerNorm) to stabilize training and output H_2 (Equation 3). Next, H_2 is processed by a feed-forward neural network (FFNN) to further learn nonlinear representations, resulting in H_3 (Equation 4). Another residual connection is applied by adding H_2 and H_3, followed by layer normalization, producing H_4 (Equation 5). Finally, a linear feed-forward network (Linear) is used to further improve the learning capacity of the companies and output $Y \in \mathbb{R}^{W \times N}$ (Equation 6).

$$H_1 = \text{MultiHeadAttention}(X) \tag{2}$$

$$H_2 = \mathrm{LayerNorm}(H_1 + X) \quad (3)$$
$$H_3 = \mathrm{FFNN}(H_2) \quad (4)$$
$$H_4 = \mathrm{LayerNorm}(H_2 + H_3) \quad (5)$$
$$Y = \mathrm{Linear}(H_4) \quad (6)$$

Figure 7 shows the feature correlation before and after ET processing. Figure 7a is the correlation coefficient between polarity scores of companies before ET processing and Figure 7b is the correlation coefficient of features output by ET. These two figures were generated using all the training data. Based on the experimental results, after ET processing, the correlation between the sentiment features of different companies is generally stronger, with more cells showing yellow or green shades. This improvement is likely due to ET's ability to capture inter-company interactions, which helps to infer and fill in missing sentiment data for certain companies. This is one of the key motivations behind its design. As the missing data is filled, additional information is introduced, leading to higher correlation coefficients between company sentiment features. The second heat map also shows clearer patterns or clusters of correlation, indicating that the Transformer model may have grouped certain companies based on sentiment similarity.

3.3.2 Forecast Transformer

The Forecast Transformer (FT) (Figure 5) uses an encoder-decoder architecture for stock price prediction and the detailed illustration of the input dimension is shown in Figure 6. The encoder processes a 5-day sliding window of historical market data that contains three key elements: (1) daily close prices, (2) sentiment features, and (3) time features (day of week, month, and day of month). These inputs are combined through a special embedding layer, where each component is processed separately [30]. The numerical features (prices and polarity scores) pass through a 1D convolutional embedding layer, while the time features are transformed via temporal embedding. All features are then projected into a shared 512-dimensional space and combined with positional information using element-wise summation before being processed by the encoder's attention layers. Experimental results demonstrate that this combined input of price data, sentiment scores, and temporal context yields better predictions than using sentiment scores alone, as the model can learn richer patterns from these complementary information sources.

Note that, we use the second input configuration discussed in Section 3.3.1 for the FT as opposed to the ET to focus on capturing temporal dependencies. This is because FT is designed to predict the stock price. Therefore, we want the model to capture the evolution of sentiments at different times. Specifically, the input dimension of the encoder is (d_{model}, W), where $W = 5$ represents the size of the sliding window of 5 days.

The decoder in the Forecast Transformer receives two key inputs to predict future stock prices. First, it takes the most recent five-day close prices to create an input sequence like [Day 1 Close, Day 2 Close, Day 3 Close, Day 4 Close, Day 5 Close]. This approach uses the last five known prices as its starting reference points. The number of known prices used for reference is a hyperparameter, defined as pre_len. In the case mentioned above, pre_len is equal to five.

Second, it processes time features (week, month, and day) for all positions in the sliding window sequence. These inputs are transformed through embedding layers: The close price passes through a 1D convolutional layer applied along the time dimension, while the time features are processed by temporal embedding. Both are projected to the same 512-dimensional space and combined with positional encoding before being fed to the decoder layers. This design allows the model to maintain temporal context while focusing on predicting future values, using the last known prices as its starting reference points.

Furthermore, the Forecast Transformer uses a hidden size of $d_{model} = 512$, which has been found to be an effective choice for this task. Reducing d_{model} leads to a degradation in prediction performance, emphasizing the importance of selecting an appropriate hidden size to balance model capacity and computational efficiency.

3.4 Model Training

We trained the Enhancement Transformer as an autoencoder using sentiment scores of different companies as inputs. An autoencoder is a type of neural network that learns to copy its input to its

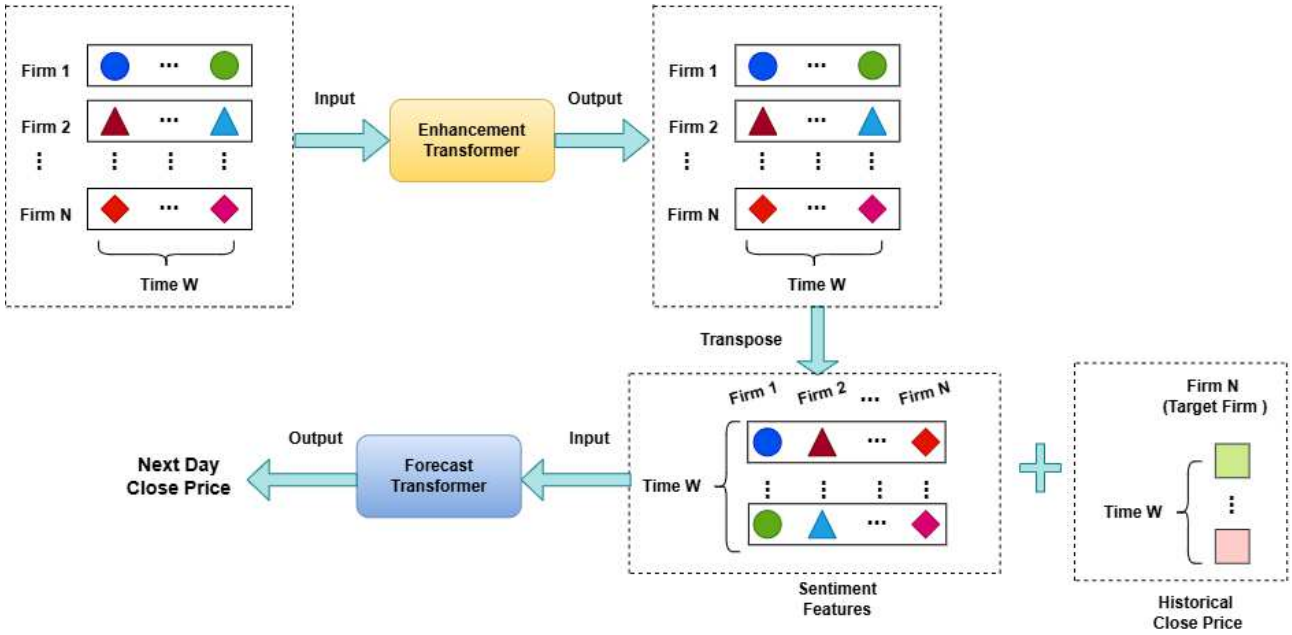


Fig. 3: Dual Transformer

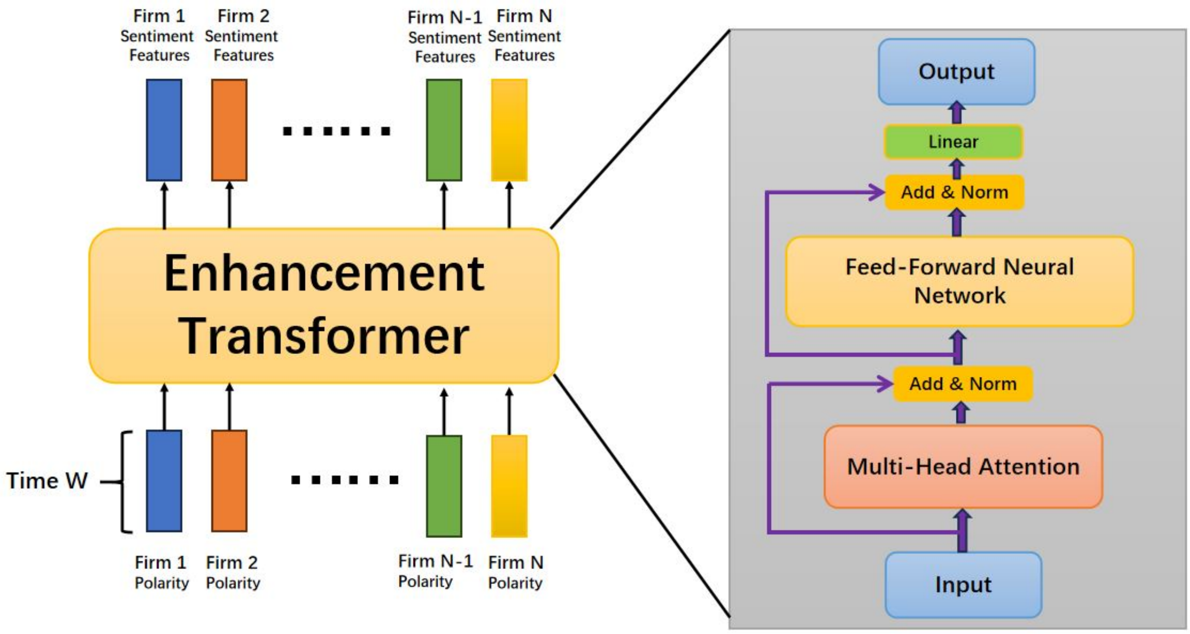


Fig. 4: Enhancement Transformer Structure

output. By training to recreate the input as accurately as possible, the model learns important patterns and structures in the data. Autoencoders are often used for tasks like noise reduction, filling in missing data, or learning compressed features [31].

At each step, the model receives a sequence of sentiment data and is trained to reconstruct

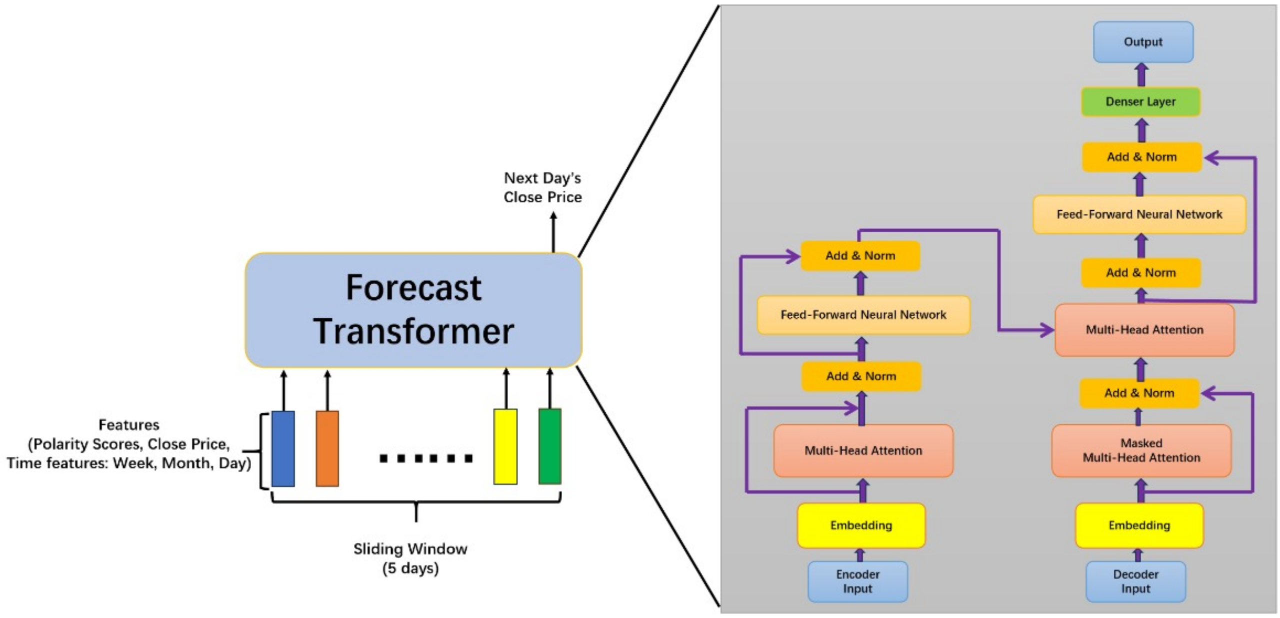


Fig. 5: Forecast Transformer

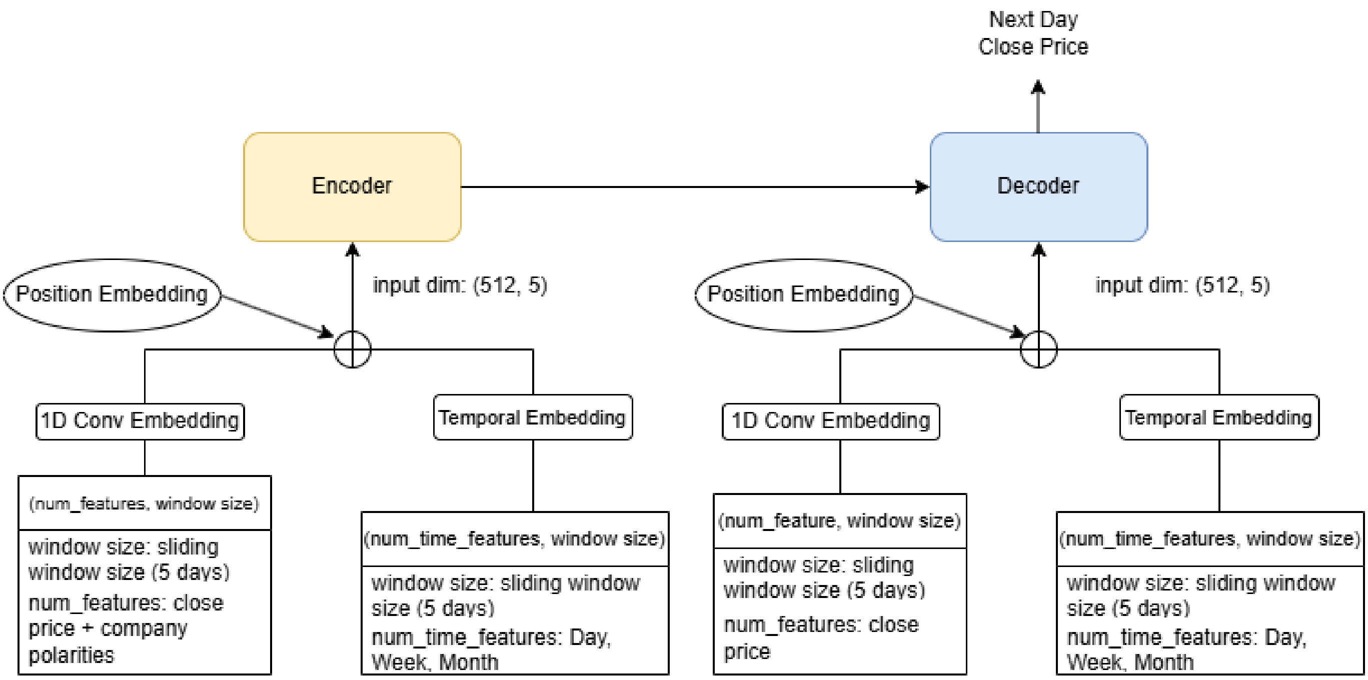
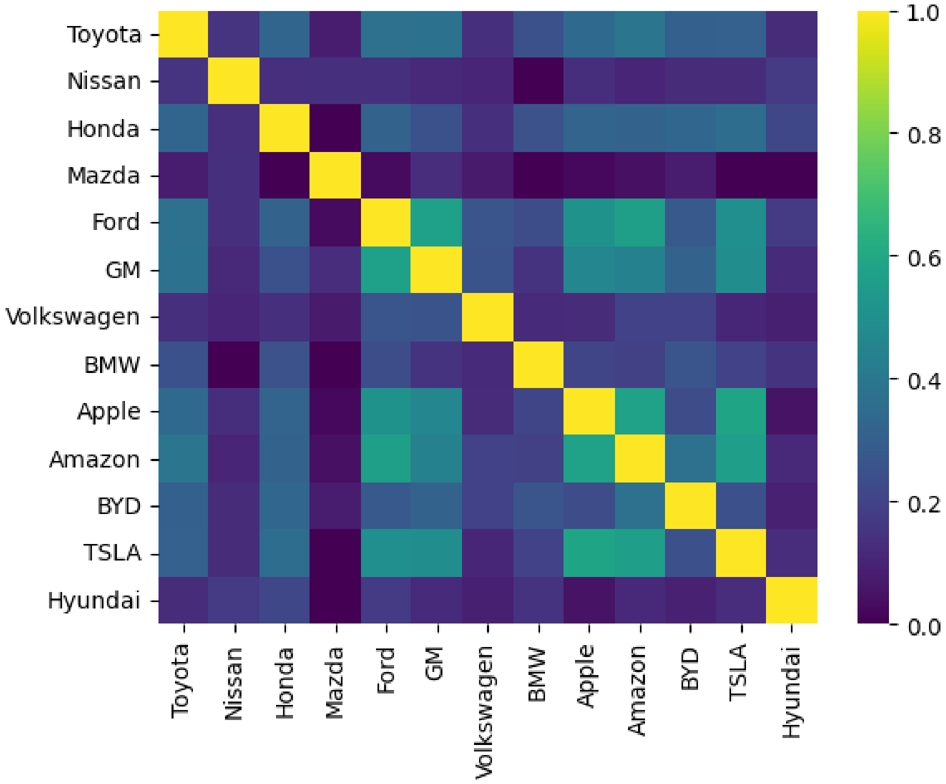
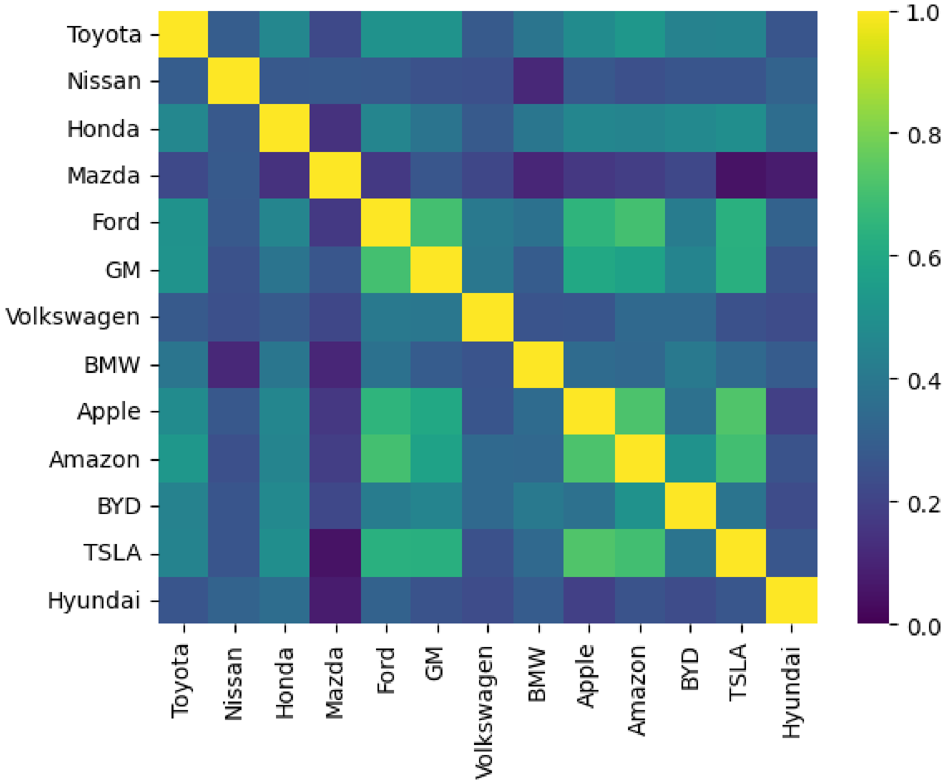


Fig. 6: Forecast Transformer: Detailed Illustration

the same sequence, with the reconstruction loss measured by mean squared error (MSE). Through this process, the Transformer learns to capture the underlying relationships between the sentiment scores of different companies. By capturing these relationships, the Enhancement Transformer



(a) Sentiment Correlation Without Enhancement Transformer Processing



(b) Sentiment Correlation After Enhancement Transformer Processing

Fig. 7: Comparison of Correlation Differences Before and After Processing

can also infer and fill in missing sentiment data that may be absent in the original dataset.

For the Forecast Transformer model, we trained it to predict stock prices based on both sentiment scores and historical close prices. The model takes sequences of combined sentiment and price information as input and outputs the predicted next-day close prices. Similar to the Enhancement Transformer, MSE is used to measure the prediction loss during training. Through this setup, the model learns to capture how changes in sentiment and past prices jointly influence future stock movements.

3.5 Baseline models

To evaluate the effectiveness of Dual Transformer, Informer, TCN, LSTM, SVR and Random Forest are chosen as baseline models for comparison, and the reason to choose these models is because they are used frequently in stock price prediction. For deep learning models such as LSTM and TCN, Optuna Tuner, proposed by [32], can be used to optimize the hyperparameters such as dropout rate, kernel size and number of channels. For machine learning models such as SVR and Random Forest, GridSearchCV can be used to optimize hyperparameters such as kernel function type. We use a sliding window of five days for the historical close prices and corresponding daily sentiment scores as input features. For each time step, the sentiment and close price are concatenated to form a feature vector, resulting in a sequence of five such vectors. This sequence is then fed into models such as LSTM, TCN, SVM, and Random Forest for predicting the stock price of the next day. Informer Model shares the same input setup as the Forecast Transformer. The following is a brief introduction to the baseline models.

Informer: Informer, proposed by [30], has mainly three innovative modules compared to the vanilla Transformer model and is designed for prediction of long sequences. The first innovative component is ProbSparse self-attention, which helps reduce time complexity and memory usage. The second innovative component is self-attention distilling, which trims the time dimension of inputs and helps receiving long sequence input. The third innovative module is called the generative style decoder, which combines historical data of a certain length with a predictive sequence filled with zeros as the input of the decoder. The informer model has been shown to be able to outperform the Transformer model [22, 27]. In this research, the same Informer structure as in the original paper [30] is applied.

Long Short Term Memory (LSTM): LSTM is a type of Recurrent Neural Network (RNN) designed to tackle the issue of vanishing or exploding gradients in RNNs. It mainly consists of the Forget Gate, the Input Gate, and the Output Gate, which are used to control the flow of information and retain memory over long periods.

Temporal Convolutional Network (TCN): TCN is a type of convolutional neural network (CNN) designed for sequential modeling. Unlike RNNs, TCN leverages dilated convolutions to capture long-term dependencies in time series data.

Support Vector Regression (SVR): SVR is an extension of the Support Vector Machine (SVM) and can be used to solve the regression problem. SVR works by finding a hyperplane to best fit the data within a defined tolerance margin.

Random Forest: Random Forest is a type of ensemble method. The main logic of Random Forest is to build multiple decision trees and then combine their results based on the majority vote or the average of all trees to improve the accuracy of the prediction.

3.6 Experimental Setup

In order to study the effectiveness of using news sentiment of related companies in stock price prediction, experiments are set up in four scenarios, with polarity score of both target and related companies (**with Polarity of Related Firms**), with polarity of target company only (**with Polarity of Target Firm**) and without polarity score (**w/o Polarity Score**). Furthermore, to examine whether changes in the number of related companies impact the model's performance, we also test the models by reducing the number of related companies (**with Fewer Related Firms**). Specifically, we rank the companies based on their Shapley values and select only the top five with the highest scores.

In addition to the polarity score, we also use historical close prices as input to predict the next day's close price. The input features are based on a sliding window of the past 5 days, meaning that the data from the previous 5 days is used to predict the close price for the following day.

4 Experiments

4.1 Related Companies Selection

Here, we use Toyota as an example to illustrate the analysis and results. First, we used NER to extract the organization names from Toyota news. Table 1 shows the top ten entity names in Toyota news. Then we selected only companies from the entity names, such as Tesla, Ford, and GM. Due to the limitation of the API we used to collect the news data, only publicly listed companies are chosen from the entity names. For each target company, we chose about ten related companies from the extracted entities based on their frequency of occurrence. Table 2 shows the related companies of all target companies. Based on the results, our proposed method can identify not only companies within the same industry or sector as the target company, but also companies from different industries that are connected to the target company. These connections may arise from investor behavior, broader market trends, or other factors that are not apparent to the general public. These connections include relationships such as Apple, Microsoft, Tesla and IBM for Pepsi, Apple and Amazon for Toyota, Northwest Pipe, Citizens & Northern, Westamerica, CECO Environmental, and Luokung Technology for InfuSystem.

We also calculated the SHAP values (Section 3.2) for the related companies after training the Random Forest model (Figure 8). Then we selected the top five companies (Amazon, Apple, BYD, GM, Ford) in this ranking together with close price as input to evaluate the model performances in stock forecasting.

Table 1: Top 10 Extracted Entity Names Using NER

Entity Name	Count
NYSE	4721
EV	4220
Tesla	2982
TM	2757
Ford	1766
GM	1591
Reuters	1192
Zacks Investment Research	965
Lexus	955
Free Stock Analysis	880
...	...

4.2 Stock Price Prediction

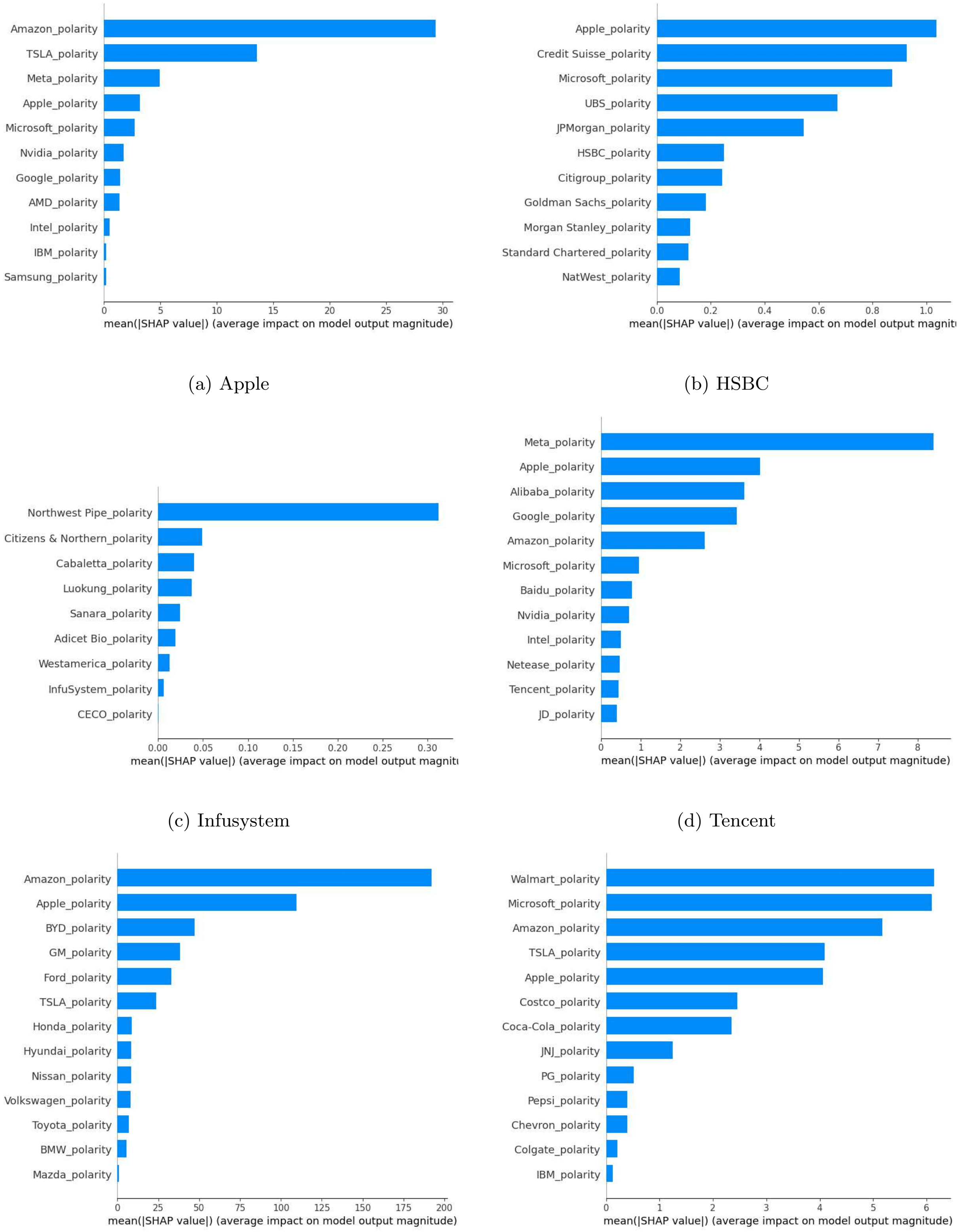
The mean squared error (MSE) is used to evaluate the performance of the models. MSE is the average squared difference between the observed values and the predicted values. A smaller MSE indicates better model performance.

Table 3 to 8 presents the prediction results in terms of MSE for the six selected companies. All values are expressed as percentages. The proposed Dual Transformer model is applied only in scenarios involving polarity scores of related companies, since the input to the Enhancement Transformer is the polarity score. The MSE values of the Dual Transformer are highlighted in red if the Dual Transformer outperforms the other models.

Based on the results, the main findings are summarized as follows. First, the proposed Dual Transformer consistently outperforms other models across all companies for almost all the time when using features with the polarity of related firms. When using fewer related firms, the Dual Transformer achieves the lowest MSE values in some cases (Table 5 and 6), indicating that removing companies with minimal contributions to prediction is likely to enhance performance. However, in most cases, the Dual Transformer yields the best result with polarity of more related companies.

In some instances, the Informer model with polarity scores as input exhibits comparative results with the Dual Transformer. For example, for HSBC (Table 4), the MSE value of Informer and Dual Transformer are closely aligned when using fewer related firms. In addition to the Dual Transformer, the Informer model shows a greater tendency to performance improvement in some cases when polarity scores are added as input compared to other models (Table 3 and 6), which indicates that models with a self-attention mechanism are better equipped to capture the intercompany relationships that influence stock prices.

For LSTM, TCN, SVR and Random Forest, adding polarity score on top of historical close price cannot improve prediction performance. In general, traditional models such as SVR and Random Forest perform worse than deep learning



(a) Apple

(b) HSBC

(c) Infusystem

(d) Tencent

(e) Toyota

(f) Pepsi

Fig. 8: Feature Importance Ranking in Terms of SHAP values

Table 2: Related Companies

Target Company	Related Company
Apple	Google, Amazon, Microsoft, Tesla, Samsung, Nvidia, Meta, Intel, AMD, IBM
HSBC	JPMorgan, Standard Chartered, Morgan Stanley, UBS, HSBC, Citigroup, Goldman Sachs, Credit Suisse, NatWest, Apple, Microsoft
Infusystem	Northwest Pipe, Citizens & Northern, Cabaletta Bio, Luokung Technology, Sanara MedTech, Adicet Bio, Westamerica, CECO Environmental
Tencent	NetEase, Baidu, JD, Alibaba, Microsoft, Apple, Google, Amazon, Meta, Nvidia, Intel
Toyota	Nissan, Honda, Mazda, Ford, GM, Volkswagen, BMW, Tesla, Hyundai, BYD, Apple, Amazon
Pepsi	Coca-Cola, Walmart, Costco, Colgate, Chevron, P&G, Johnson & Johnson, Apple, Microsoft, Amazon, Tesla, IBM

models for this task, which demonstrates the limitations of traditional machine learning models in capturing the complicated relationships present in financial data. The performance of Random Forest shows a trivial difference when polarity scores are added as input.

Generally, in this experimental setup, the proposed Dual Transformer model exhibits state-of-the-art performance in forecasting stock prices by effectively leveraging the polarity of related firms. Additionally, the results highlight the importance of intercompany relationships and advanced Transformer-based architectures in financial prediction tasks.

4.3 Ablation Study

We conducted an ablation study to evaluate the performance of various model configurations on stock price forecasting task for Apple, Pepsi, Toyota, and Tencent. The model configurations include Dual Transformer, Forecast Transformer (FT) alone, Enhancement Transformer (ET) combined with the Informer model, and the Informer model alone. The study is conducted in two feature setups: with polarity of related firms and with fewer related firms. MSE is used for performance evaluation.

Figure 9 to 12 illustrate the results of the ablation study. The MSE values are highlighted in red when the dual-structure models outperform the single-structure model. Generally, after adding the ET, the performance of the models tends to improve. The ET combined with Informer achieves competitive results compared to the Dual Transformer. This ablation study also highlights the robustness and effectiveness of the Dual Transformer, particularly in leveraging the polarity of related firms for improved prediction accuracy.

4.4 Further Discussion

In this section, we use a method called event study to further illustrate how the relationships between companies captured by the Enhancement Transformer match the real-world situations.

In finance, the event study can help identify companies whose cumulative abnormal returns follow a pattern similar to that of the target company. This similarity suggests that the events may have comparable impacts on these companies, indicating a potential connection between them.

The company relationships identified through event studies can be used for comparison with those captured by the Enhancement Transformer. The details of the event study will be explained below, followed by a comparison of the relationships identified by the event study and the Enhancement Transformer.

4.4.1 Event Study

In finance, event study is often used to estimate how asset prices react to announcements of events that contain new information relevant to the value of the underlying assets. To conduct an event study, the main idea is to compare the asset price that occurred as a consequence of the announcement of the event with a hypothetical asset price that would have occurred if no event had happened. The former is called the realized return (return calculated using the stock prices in the event window), and the latter is called the expected return (return calculated using the stock prices in the estimation period). That is, we want to calculate the difference between the realized and expected return, called the abnormal return (Equation 7). The expected return is calculated using the market model (Equation 8). The final step is to calculate the cumulative abnormal return (CAR) (Equation 9). Companies that show similar patterns in CAR during the event study are likely connected. Additionally, if the CARs of some companies frequently move in opposite directions, it may suggest a relationship where favorable news for one translates to unfavorable news for another. For example, negative news for

Table 3: Mean Squared Error (%): Apple

Features	Dual Transformer	Informer	TCN	LSTM	SVR	RF
with Polarity of Related Firms	0.002	0.032	0.179	0.286	0.568	0.184
with Fewer Related Firms	0.004	0.029	0.160	0.233	0.544	0.181
with Polarity of Target Firm	-	0.072	0.152	0.158	0.404	0.180
w/o Polarity Score	-	0.071	0.142	0.150	0.372	0.178

Table 4: Mean Squared Error (%): HSBC

Features	Dual Transformer	Informer	TCN	LSTM	SVR	RF
with Polarity of Related Firms	0.022	0.041	0.091	0.078	0.246	0.042
with Fewer Related Firms	0.041	0.049	0.060	0.059	0.246	0.043
with Polarity of Target Firm	-	0.035	0.082	0.043	0.114	0.043
w/o Polarity Score	-	0.030	0.047	0.049	0.114	0.043

Table 5: Mean Squared Error (%): Infusystem

Features	Dual Transformer	Informer	TCN	LSTM	SVR	RF
with Polarity of Related Firms	0.032	0.083	0.156	0.115	0.342	0.093
with Fewer Related Firms	0.029	0.038	0.333	0.147	0.329	0.092
with Polarity of Target Firm	-	0.040	0.081	0.089	0.113	0.092
w/o Polarity Score	-	0.067	0.083	0.081	0.113	0.092

Table 6: Mean Squared Error (%): Pepsi

Features	Dual Transformer	Informer	TCN	LSTM	SVR	RF
with Polarity of Related Firms	0.026	0.065	0.298	0.953	0.205	0.136
with Fewer Related Firms	0.022	0.028	0.150	0.143	0.252	0.136
with Polarity of Target Firm	-	0.077	0.376	0.153	0.210	0.138
w/o Polarity Score	-	0.088	0.126	0.129	0.217	0.137

Table 7: Mean Squared Error (%): Toyota

Features	Dual Transformer	Informer	TCN	LSTM	SVR	RF
with Polarity of Related Firms	0.009	0.027	0.137	0.171	0.181	0.050
with Fewer Related Firms	0.010	0.026	0.106	0.134	0.187	0.049
with Polarity of Target Firm	-	0.016	0.047	0.121	0.187	0.049
w/o Polarity Score	-	0.020	0.038	0.044	0.083	0.047

Table 8: Mean Squared Error (%): Tencent

Features	Dual Transformer	Informer	TCN	LSTM	SVR	RF
with Polarity of Related Firms	0.028	0.047	0.315	0.413	0.203	0.100
with Fewer Related Firms	0.035	0.106	0.245	0.321	0.201	0.099
with Polarity of Target Firm	-	0.041	0.103	0.107	0.201	0.099
w/o Polarity Score	-	0.054	0.095	0.100	0.201	0.099

Table 9: Ablation Study, MSE (%): Apple

Features	Dual Transformer	FT	ET + Informer	Informer
with Polarity of Related Firms	0.002	0.016	0.003	0.032
with Fewer Related Firms	0.004	0.017	0.025	0.029

Table 10: Ablation Study, MSE (%): Pepsi

Features	Dual Transformer	FT	ET + Informer	Informer
with Polarity of Related Firms	0.026	0.035	0.037	0.061
with Fewer Related Firms	0.022	0.028	0.038	0.044

Table 11: Ablation Study, MSE (%): Toyota

Features	Dual Transformer	FT	ET + Informer	Informer
with Polarity of Related Firms	0.009	0.021	0.026	0.027
with Fewer Related Firms	0.010	0.022	0.023	0.026

Table 12: Ablation Study, MSE (%): Tencent

Features	Dual Transformer	FT	ET + Informer	Informer
with Polarity of Related Firms	0.028	0.044	0.039	0.047
with Fewer Related Firms	0.030	0.034	0.022	0.106

one company could translate into positive news for its competitors.

The event study must specify the event date (business day right after the event announcement), the estimation period (period to calculate the expected return), and the event window (period to calculate the realized return). Here, we used before and after 10 days surrounding the event date as the event window and the last 130 days to last 11 days before the event as the estimation period (Figure 9). The length of the estimation period is arbitrary, but in practice, the length of the estimation period typically ranges from 100 to 250 trading days.

$$AR_{i,t} = R_{i,t} - ER_{i,t}, \quad t \in [-10, 10] \tag{7}$$

where $AR_{i,t}$ is the abnormal return for the stock i on day t, $R_{i,t}$ is the realized return for the stock i on day t, calculated using the stock prices in the event window, $ER_{i,t}$ is the expected return for the stock i on day t, which is calculated using Equation 8.

$$ER_{i,t} = \alpha_i + \beta_i R_{m,t} + \epsilon_{i,t} \tag{8}$$

where parameters such as α_i and β_i can be obtained using the stock data in the estimation period $t \in [-130, -11]$. α_i is the intercept term, β_i is the slope coefficient, representing the sensitivity of the return of the stock i to the market return, $R_{m,t}$ is the return of the market index on day t, $\epsilon_{i,t}$ is the error term. Examples of market index include S&P 500 (US), Hang Seng Index (Hong Kong), Taiwan Capitalization Weighted Stock Index (Taiwan), Swiss Market Index (Switzerland) and Korea Composite Stock Price Index (South Korea). Stocks from different countries are associated with the market index of their respective exchanges. For example, Tencent, listed on the Hong Kong Stock Exchange, uses the Hang Seng Index for calculations.

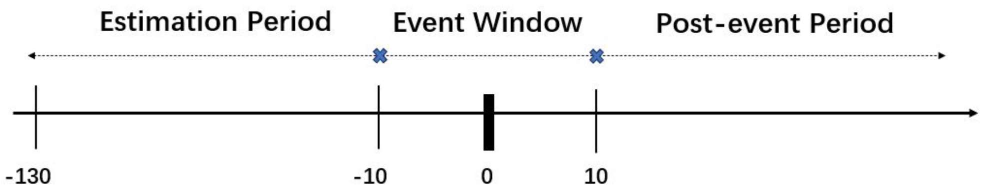


Fig. 9: Event Study

$$CAR_{i,t} = \sum_{t'=-10}^{t} AR_{i,t'}, \quad (9)$$

where $CAR_{i,t}$ is the cumulative abnormal return for the stock i during the event window, $AR_{i,t'}$ is the abnormal return for the stock i on day t'.

4.4.2 Analysis of Enhancement Transformer with Event Study Results

Here, we use related companies of Toyota as an example. Figure 11 shows the results of the event study using Toyota's scandal on 3 June 2024. On this date, Toyota acknowledged irregularities in safety tests conducted by its subsidiaries, Daihatsu and Hino Motors. Figure 12 shows the event study using the Russia-Ukraine War, which started on 24 February, 2022. These two figures reveal changes in cumulative abnormal returns for companies such as BYD, Nissan, Volkswagen, and Honda after these events, which indicate relationships between Toyota and other companies.

Figure 10 uses attention scores from the transformer model (second layer) to measure the interaction between companies based on their polarity scores. The attention scores here are computed using data from 2022 to 2024, which covers the event dates mentioned above. Attention scores are used to represent the importance assigned to each input feature when making predictions, providing information on the relationships between companies. By aggregating and analyzing these scores, we can quantify the degree to which companies are related within the learned representation of the transformer. In this research, we used two layers and four attention heads in the calculation of self-attention. The attention scores generated by each layer represent the combined contributions from all heads (Equation 2), with the attention scores averaged across all heads.

According to Figure 11, after Toyota's scandal, the abnormal returns of most of auto companies dropped. Although the magnitude of the abnormal returns varied, their overall direction or trend was largely consistent. In particular, Honda, Nissan, and Mazda exhibited remarkably similar movements in their abnormal returns. Similarly, Ford, General Motors, and BMW showed similar patterns, as did Toyota and Volkswagen. These relationships are also captured in the transformer results (Figure 10). For example, Toyota shows a relatively strong positive interaction with Honda and Volkswagen, suggesting that their sentiment trends are closely related. Companies such as Volkswagen and BMW exhibit a delayed or less pronounced reaction compared to Japanese automakers such as Nissan and Honda, potentially reflecting regional market differences or varying levels of exposure to the scandal.

Furthermore, based on Figure 10, Tesla stands out with predominantly negative or weak interaction with other companies, such as Toyota, Mazda, Honda and Ford, which is consistent with its unique pattern in the event study (Figure 11 and 12). This reflects the reality that Tesla focuses primarily on the electric vehicle market, while others remain more diversified in traditional and hybrid vehicles.

BYD and Tesla exhibit positive interactions and relatively similar patterns in the event study, highlighting their shared focus on the electric vehicle market.

Apple displays weak but noticeable interactions with Tesla in some aspects, reflecting their alignment with innovation and technology-driven sentiment, even though they operate in different industries.

In general, based on the results and analysis, the proposed transformer model can effectively capture and enhance the relationships and interactions between companies. In addition, the

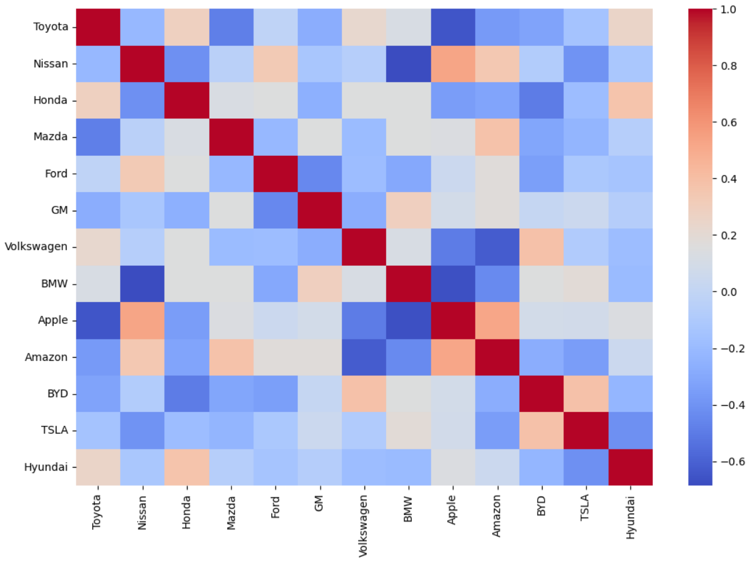


Fig. 10: Attention Scores between Companies with ET Processing (Different From Figure 7b)

consistency between some results of the event study and the transformer demonstrates the transformer's ability to uncover meaningful relationships between entities, complementing traditional event study analyses.

5 Conclusion and Limitations

This paper proposed a novel sentiment correlation-based method and a Dual Transformer model to forecast the stock prices of six companies. Some key findings of this paper are as follows. (1) The Enhancement Transformer (encoder transformer) of our proposed Dual Transformer model can help increase the correlation strength of the companies' polarity scores. (2) The proposed Dual Transformer model with sentiments of related companies as input can outperform models like Informer and LSTM. (3) Models with attention mechanisms, such as Informer, are more likely to improve prediction performance when sentiment data is added as input, which demonstrates the strength of self-attention mechanisms in capturing intercompany relationships, which are crucial for understanding the influence of related firms on stock prices.

This research has certain limitations. The API used to collect news data is limited to publicly listed companies, with a bias toward American companies, resulting in less comprehensive coverage for other regions. To address this, alternative data sources will be explored in future work. Another limitation of this study lies in the NER process, which is leveraged to extract companies mentioned in financial news. While NER effectively identifies potential companies, the final selection of the relevant company is performed manually. This manual step introduces subjectivity and may limit scalability, particularly when processing large volumes of financial news.

For future study, automating this step can improve efficiency and consistency. Generative large language models, such as GPT and LLaMA, can be used to better understand the context and relevance of company mentions in financial news, potentially improving entity recognition accuracy and reducing manual intervention.

Declarations

- **Funding Declaration** No funding was received for conducting this study.
- **Conflict of interest/Competing interests** The authors declare that there are no conflicts of interest or competing interests associated with this research.
- **Ethics approval and consent to participate** Not applicable

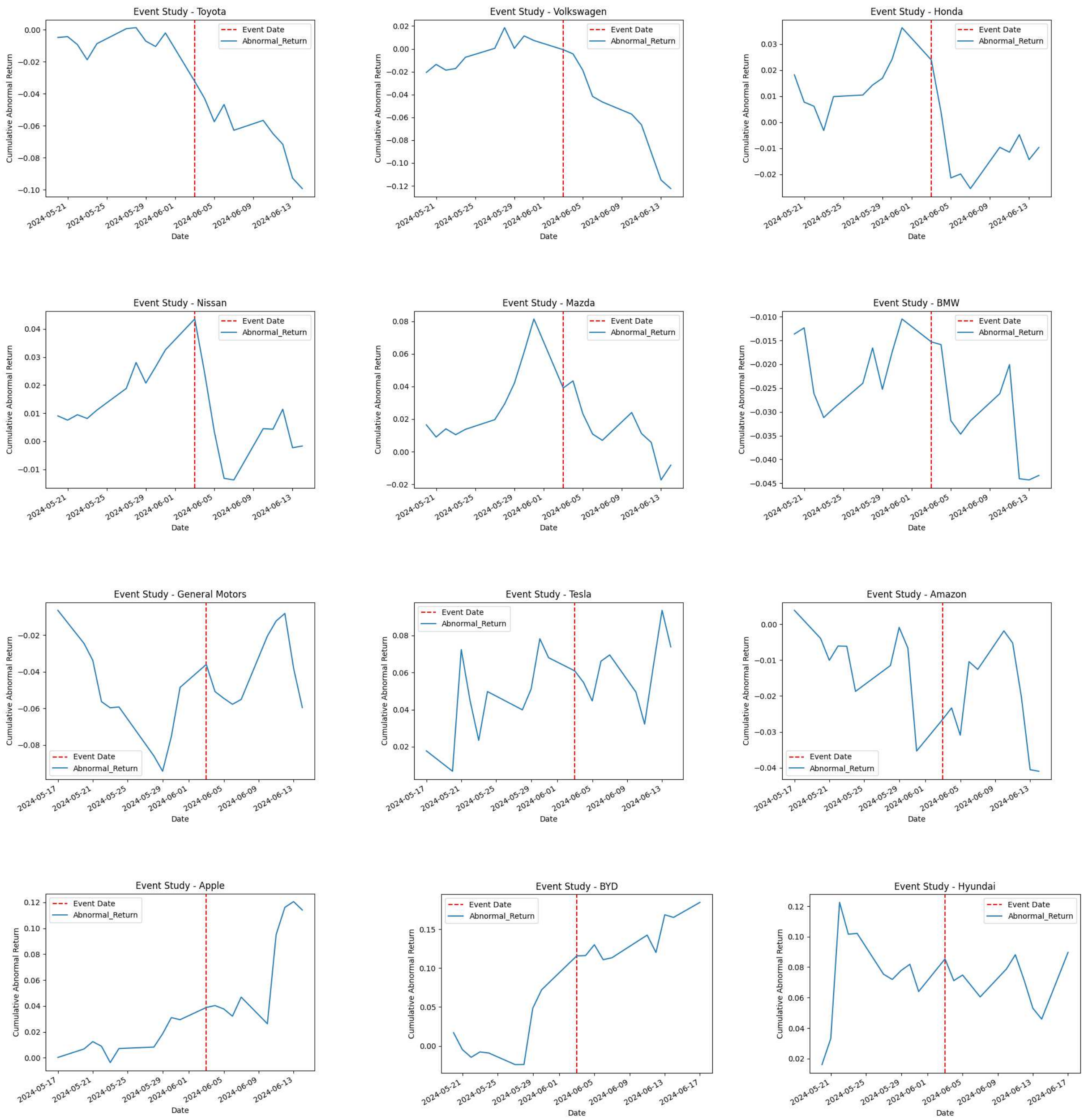


Fig. 11: Event Study - Toyota Scandal: Related Companies of Toyota

- **Consent for publication** All authors consent to the publication of this work in its current form.
- **Data availability** The news datasets used in this study were collected using the News Feed and News Sentiment data API.
- **Materials availability** Not applicable.
- **Code availability** GitHub Repository
- **Author contribution** Qizhao Chen conceptualized the idea, conducted the experiments, and drafted the manuscript. Hiroaki Kawashima contributed to the conceptualization of the idea and provided critical revisions to the manuscript.

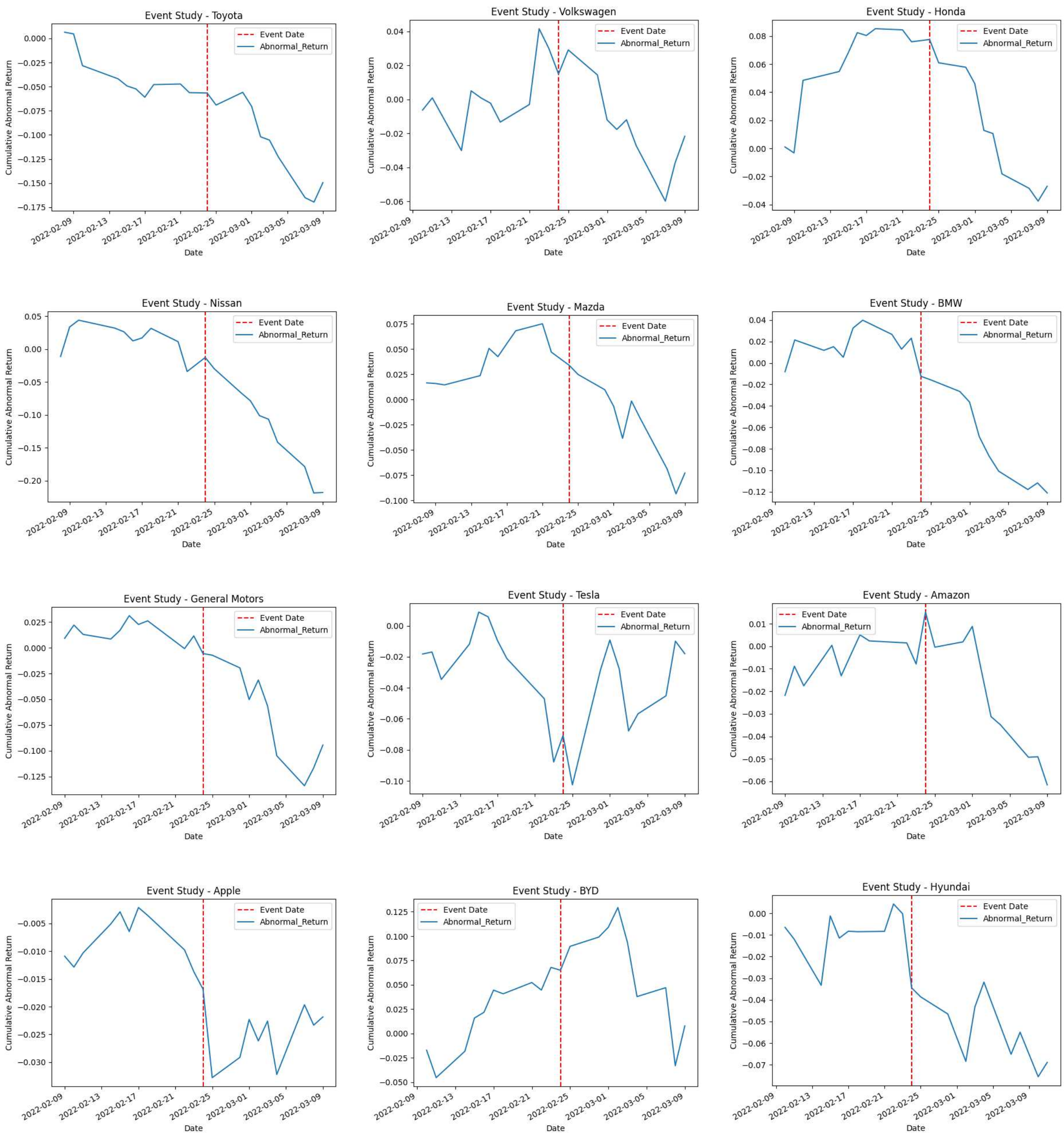


Fig. 12: Event Study - Russia-Ukraine War: Related Companies of Toyota

References

[1] Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., Anastasiu, D.C.: Stock price prediction using news sentiment analysis. In: 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), pp. 205–208 (2019). https://doi.org/10.1109/BigDataService.2019.00035

[2] Cristescu, M.P., Nerisanu, R.A., Mara, D.A.,

Oprea, S.-V.: Using market news sentiment analysis for stock market prediction. Mathematics **10**(22) (2022) https://doi.org/10.3390/math10224255

[3] Luan, Y., Zhang, H., Zhang, C., Mu, Y., Wang, W.: Stock price prediction with sentiment analysis for chinese market. In: Proceedings of the Joint Workshop of the 7th Financial Technology and Natural Language Processing, the 5th Knowledge Discovery from Unstructured Data in Financial Services, and the 4th Workshop on Economics and Natural Language Processing, pp. 167–177. Association for Computational Linguistics, Torino, Italia (2024). https://aclanthology.org/2024.finnlp-1.16

[4] Xiao, Q., Ihnaini, B.: Stock trend prediction using sentiment analysis. PeerJ Comput. Sci. **9**, 1293 (2023) https://doi.org/10.7717/peerj-cs.1293

[5] Darapaneni, N., Paduri, A.R., Sharma, H., Manjrekar, M., Hindlekar, N., Bhagat, P., Aiyer, U., Agarwal, Y.: Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets (2022). https://doi.org/10.48550/arXiv.2204.05783

[6] Kolasani, S.V., Assaf, R.: Predicting stock movement using sentiment analysis of twitter feed with neural networks. J. Data Anal. Inf. Process. **08**(04), 309–319 (2020) https://doi.org/10.4236/jdaip.2020.84018

[7] Wan, X., Yang, J., Marinov, S., Calliess, J.-P., Zohren, S., Dong, X.: Sentiment correlation in financial news networks and associated market movements. Scientific Reports **11**(1) (2021) https://doi.org/10.1038/s41598-021-82338-6

[8] Mu, G., Gao, N., Wang, Y., Dai, L.: A stock price prediction model based on investor sentiment and optimized deep learning. IEEE Access **11**, 51353–51367 (2023) https://doi.org/10.1109/ACCESS.2023.3278790

[9] Maqbool, J., Aggarwal, P., Kaur, R., Mittal, A., Ganaie, I.A.: Stock prediction by integrating sentiment scores of financial news and mlp-regressor: A machine learning approach. Procedia Computer Science **218**, 1067–1078 (2023) https://doi.org/10.1016/j.procs.2023.01.086

[10] John, A., Latha, T.: Stock market prediction based on deep hybrid rnn model and sentiment analysis. Automatika **64**(4), 981–995 (2023) https://doi.org/10.1080/00051144.2023.2217602

[11] Moghar, A., Hamiche, M.: Stock market prediction using lstm recurrent neural network. Procedia Computer Science **170**, 1168–1173 (2020) https://doi.org/10.1016/j.procs.2020.03.049

[12] Reddy, K.R., Kumar, B.T., Ganesh, V.R., Swetha, P., Sarangi, P.K.: Stock market prediction using recurrent neural network. In: 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), pp. 1–4 (2022). https://doi.org/10.1109/CCET56606.2022.10080154

[13] Islam, S.B., Hasan, M.M., Khan, M.M.: Prediction of stock market using recurrent neural network. In: 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 0479–0483 (2021). https://doi.org/10.1109/IEMCON53756.2021.9623206

[14] Mehtab, S., Sen, J.: Stock price prediction using cnn and lstm-based deep learning models. In: 2020 International Conference on Decision Aid Sciences and Application (DASA), pp. 447–453 (2020). https://doi.org/10.1109/DASA51403.2020.9317207

[15] Chen, S., He, H.: Stock prediction using convolutional neural network. IOP Conference Series: Materials Science and Engineering **435**(1), 012026 (2018) https://doi.org/10.1088/1757-899X/435/1/012026

[16] Mehta, P., Pandya, S., Kotecha, K.: Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Computer Science **7**, 476 (2021) https://doi.org/10.7717/peerj-cs.476

[17] Nguyen, T.H., Shirai, K.: Topic modeling based sentiment analysis on social media for stock market prediction. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 1354–1364 (2015). https://doi.org/10.3115/v1/P15-1131

[18] Batra, R., Daudpota, S.M.: Integrating stocktwits with sentiment analysis for better prediction of stock price movement. In: 2018 International Conference on Computing, Mathematics and Engineering Technologies (ICoMET), pp. 1–5 (2018). https://doi.org/10.1109/ICOMET.2018.8346382 . IEEE

[19] Li, X., Xie, H., Chen, L., Wang, J., Deng, X.: News impact on stock price return via sentiment analysis. Knowledge-Based Systems **69**, 14–23 (2014) https://doi.org/10.1016/j.knosys.2014.04.022

[20] Pagolu, V.S., Reddy, K.N., Panda, G., Majhi, B.: Sentiment analysis of twitter data for predicting stock market movements. In: 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), pp. 1345–1350 (2016). https://doi.org/10.1109/SCOPES.2016.7955659 . IEEE

[21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need (2023). https://arxiv.org/abs/1706.03762

[22] Chen, Q., Kawashima, H.: Stock price prediction using llm-based sentiment analysis. In: Proceedings of the IEEE BigData 2024, pp. 4828–4835. IEEE, Washington DC, USA (2024). https://doi.org/10.1109/BigData62323.2024.10825946

[23] Ji, Y., Luo, Y., Lu, A., Xia, D., Yang, L., Wee-Chung Liew, A.: Galformer: a transformer with generative decoding and a hybrid loss function for multi-step stock market index prediction. Sci. Rep. **14**(1), 23762 (2024) https://doi.org/10.1038/s41598-024-72045-3

[24] Li, T., Liu, Z., Shen, Y., Wang, X., Chen, H., Huang, S.: MASTER: Market-Guided Stock Transformer for Stock Price Forecasting (2023). https://doi.org/10.48550/arXiv.2312.15235

[25] Khedr, A.E., Yaseen, N., *et al.*: Predicting stock market behavior using data mining technique and news sentiment analysis. International Journal of Intelligent Systems and Applications **9**(7), 22 (2017) https://doi.org/10.5815/ijisa.2017.07.03

[26] Srijiranon, K., Lertratanakham, Y., Tanantong, T.: A hybrid framework using pca, emd and lstm methods for stock market price prediction with sentiment analysis. Applied Sciences **12**(21), 10823 (2022) https://doi.org/10.3390/app122110823

[27] Lu, Y., Zhang, H., Guo, Q.: Stock and market index prediction using Informer network (2023). https://doi.org/10.48550/arXiv.2305.14382

[28] Lundberg, S., Lee, S.-I.: A Unified Approach to Interpreting Model Predictions (2017). https://arxiv.org/abs/1705.07874

[29] Han, M., Zhang, D.J., Wang, Y., Yan, R., Yao, L., Chang, X., Qiao, Y.: Dual-AI: Dual-path Actor Interaction Learning for Group Activity Recognition (2022). https://doi.org/https://arxiv.org/abs/2204.02148

[30] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting (2021). https://doi.org/10.48550/arXiv.2012.07436

[31] Michelucci, U.: An Introduction to Autoencoders (2022). https://arxiv.org/abs/2201.03898

[32] Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019). https://doi.org/10.1145/3292500.3330701