Stock Price Prediction Using LLM-Based Sentiment Analysis

1st Qizhao Chen
Graduate School of Information Science
University of Hyogo
Kobe, Japan
brandonchan787@gmail.com

2nd Hiroaki Kawashima
Graduate School of Information Science
University of Hyogo
Kobe, Japan
kawashima@gsis.u-hyogo.ac.jp

***Abstract*—This paper examines the effectiveness of recent large language model-based news sentiment estimation for stock price forecasting with the combination of latest transformer-based prediction models. To achieve a better accuracy in sentiment classification, experiments are designed to compare six different models (GPT 4, Llama 3, Gemma 2, Mistral 7b, FinBERT, VADER) in financial news sentiment classification, and it was found that recent large language models can outperform FinBERT and VADER, which are the most commonly used models in financial sentiment analysis. Based on the experiment results, Llama 3, with relatively stable performance, is chosen to classify the news sentiments of the selected companies. Informer, Transformer, TCN, LSTM, SVR, Random Forest and Naive Forecast are used to predict the stock prices with different sliding window sizes. Experiments with different scenarios are designed to evaluate the prediction ability of news sentiment. Results show that adding news sentiment data can indeed improve the stock price prediction. Informer, one of the state-of-the-art transformer models for long-term prediction tasks, yields the best performances in most cases. Ablation study of Informer suggests that the generative style decoder plays an important role in performance improvement.**

***Index Terms*—time series forecasting, sentiment analysis, Transformer, Informer, LLM**

I. Introduction

The objective of this paper is to study whether adding news sentiment in stock price prediction can improve the forecasting performances by applying large language models (LLMs) to classify the sentiments because the accuracy of sentiment classification has significant impact on stock price prediction. Based on the objective, we predict the stock prices using machine learning models and financial news sentiment. Traditionally, researchers tend to use technical indicators only, derived from historical price and volume data, to predict the stock prices [1]–[4]. This method only focuses on past information and market behavior. However, news sentiment provides insights into the market's reaction to current events, which may not be reflected in historical price data, especially those unpredictable situations like natural disasters and political upheavals. Thus, it is necessary to include news sentiment in stock price prediction to better capture the market behavior but how well it can help in forecasting need to be further investigated in this paper.

Sentiment analysis is used to determine the emotional tone of texts using the natural language processing techniques. There are mainly two approaches to conduct sentiment analysis. The first approach is the lexicon-based approach, which is to utilize pre-prepared dictionaries or sentiment lexicons to determine the sentiments of texts. One of the most popular models of this approach is VADER (Valence Aware Dictionary and Sentiment Reader). Many research papers related to finance sentiment analysis [5]–[9] also apply VADER to classify financial texts. The second approach is machine learning-based approach, which treats sentiment analysis as a classification problem. Machine learning models can be trained on labeled datasets to forecast the sentiments of the texts. Recently, due to their superior performance in natural language processing tasks, Transformer-based models have become widely used in sentiment classification. One of the examples is BERT (Bidirectional Encoder Representations from Transformers). FinBERT may be the most popular BERT model in finance sentiment analysis. FinBERT, proposed in [10], is a pre-trained BERT model for financial text sentiment.

Recently, AI chatbots such as ChatGPT, backed by LLMs, become very popular. They follow an instruction in the way of a prompt and then offers a detailed response. Sentiment analysis can also be done using this kind of chatbot or the LLMs backing them. An input (prompt) is entered and the chatbot can give you the sentiments for the corresponding texts. Sometimes, it is difficult to have an accurate sentiment classification when using models like VADER and BERT to classify financial texts because finance domain has its own unique terminologies. For example, bear and bull are two types of animals but in finance, these two terms are used as upward market (bullish market) or downward market (bearish market). On the other hand, while generative LLMs are widely used in many tasks, its prediction ability for sentiments related to financial sources is still not clear. Thus, in this paper, LLMs are applied to classify the sentiments of financial news headlines. Experiments are conducted to compare the performances of six different models in sentiment classification. To clarify the effectiveness of prompt-based generative LLMs compared to widely used LLMs in the financial domain, models chosen for comparison include GPT 4, LLaMA 3, Gemma 2, Mistral 7b, FinBERT and VADER.

When it comes to models to predict the stock prices, in the past decade, many researchers utilized machine learning models like Support Vector Machine (SVM), Random Forest and Gradient Boosting (e.g., XGBoost) [6], [8], [11], [12]. Since stock price has non-linear behavior, although these machine learning models can handle non-linearity to some extent, such as kernel trick in SVMs, they fail to capture the stock price movement in highly volatile market. Due to these reasons, Recurrent Neural Network (RNN) (e.g., Long Short Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM)) becomes a very popular candidate for researchers to forecast the stock price [2], [5], [9], [13], [14] because RNN, with neural network structure, can better handle non-linearity and capture the temporal dependencies between data points while traditional machine learning models like SVM treat each data point, including fixed-length history, independently. Researchers also utilize Convolutional Neural Network (CNN) to forecast stock price [15] although CNN is used more commonly in image recognition. One type of CNNs designed for time series forecasting is Temporal Convolutional Network (TCN). TCN, proposed by [16], has proven to be able to outperform RNN in time series forecasting. Furthermore, some hybrid models such as CNN-LSTM are also used in stock price forecasting [1].

Recently, Transformer-based models are also used in time series forecasting although originally Transformer is applied in natural language processing. Related research show that Transformer can outperform LSTM in stock price prediction [3], [17]. Various extensions are also proposed to handle different tasks. In time series forecasting, one of the newly proposed models is Informer. Informer, proposed by [18], mainly utilizes ProbSparse and distilling self-attention to reduce the computation cost of Transformer and makes it better for long sequence time series forecasting. The distilling technique can also reduce the sequence length progressively without losing important information. In the original paper of Informer model, researchers applied Informer model to predict time series of electricity transformer temperature, electricity consuming load and climatological data in the US. However, its effectiveness in stock price forecasting is not obvious as stock prices affected by a greater variety of factors, including sentiments.

In this paper, we compare recent transformer models, including Informer, and more traditional machine learning models commonly used in stock price forecasting to examine the effectiveness of recent transformer-based models in stock price prediction and also their affinity with sentiment features.

The main contributions of this research are:

1) We verify the accuracy of recent generative large language models for sentiment classification of financial news headlines and show that models like Llama 3 can outperform FinBERT and VADER models, two most commonly used models in financial sentiment analysis.
2) We conduct thorough comparison of transformer models including its latest extension and verify the performance in stock price prediction with financial news sentiment. We show that Informer model outperforms all other models such as LSTM and basic Transformer models.

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

II. Related Work

Stock price prediction is an active research area in finance, computer science and statistics. Various models, different sentiment analysis techniques, different features or forecast horizons are used by researchers to try to achieve a better prediction performance.

1) Use of Technical Indicators: Traditionally, technical indicators are popular features for researchers to predict the stock prices and machine learning models like LSTM, SVR and ensemble methods like Random Forest and Gradient Boosting are often used in prediction. For example, Phuoc et al. apply LSTM with technical indicators (SMA, RSI, MACD) to predict the stock price trend in Vietnam. The results show that LSTM has a high accuracy of 93% for most of the stocks used [19]. Jabed applies LSTM, Facebook Prophet and Random Forest Regressor to predict the stock prices with moving averages and the results demonstrate that LSTM yields the best performance, especially in forecasting the next day's closing price [20]. Gür applies SVM, XGBoost and LSTM to predict the daily stock price of Turkish Airlines and the results show that LSTM can outperform SVM and XGBoost [21]. Henrique et al. use SVR to predict stock prices for large and small capitalisation in three different markets with technical indicators. The results show that SVR can predict the stock prices and the linear kernel used in SVR is better than RBF and polynomial kernels [22]. Jayaswara et al. also apply SVR with linear and RBF kernels to predict Central Asia Bank's closing price for the next 10 days. The results show that SVR with linear kernel has a good prediction performance [23]. Orsel & Yamada apply a linear Kalman filter and different varieties of LSTMs (LSTM, Bi-LSTM, CNN-LSTM) to forecast stock prices. What they found is that a simple linear Kalman filter is already good enough to forecast low-volatile stock prices for the next day while LSTMs outperform Kalman filter for high-volatile stock price forecasting [24].

2) Use of Macroeconomic Parameters: Although used less commonly, macroeconomic parameters are also proved to have prediction ability in stock movement. Haque et al. incorporate macroeconomic variables with historical stock prices to predict the close price. The macroeconomic variables include Volume, Consumer Price Index, Unemployment Rates and Interest Rates. Prediction models include Light GBM, XGB Regressor and Decision Tree. The results show that inclusion of macroeconomic variables improves the model performances [25]. Weng et al. also investigate the use of macroeconomic variables in predicting the one-month ahead price for major US stock and sector indices. They compare the performances between four ensemble methods and three time series models (ARIMA, GARCH, LSTM). The results show that four ensemble methods outperform the three time-series models and incorporating the macroeconomic variables can improve the model accuracy [26].

3) Use of Sentiment Data: On top of technical indicators and macroeconomic parameters, researchers also study how financial news and sentiment data can affect the stock movement. Maqbool et al. study the impact of financial news on stock price using three different methods (VADER, TextBlob, Flair) to classify the news sentiments. The results show that MLP Regressor with financial news sentiments can predict the stock price with an accuracy of 90% [7]. Mishra et al. analyze how news and historical prices affect the stock market. They use BERT to classify the sentiments of Financial Phrasebank dataset with a 83% percent accuracy. LSTM and ARIMA models are used for stock price prediction. The results show that LSTM outperforms ARIMA model [27]. Sun et al. use the stock comment sentiments and technical indicators to predict the next day's closing price in Chinese stock market. The prediction models are LSTM and Bi-LSTM. The results show that incorporating sentiment information does not improve prediction accuracy all the time for every stock. Bi-LSTM model with sentiment data has the best performance [28].

III. Methodology

A. Sentiment Analysis

Figure 1 shows the overall procedure of the sentiment classification. Sentiments of financial news headlines of three companies (Apple, Tencent, Toyota) from three stock markets are classified using LLM by giving the prompt instruction classify the sentiment of this piece of news headline: [Headline Input]. Sentiment is "positive", "negative" or "neutral". Return only the sentiment of the news headline." The output sentiment is either positive, negative or neutral.

In order to apply the appropriate LLM to conduct the sentiment classification, performances of six different models (Table I) are compared using two different datasets. Llama is a LLM developed by Meta AI from February 2023 and the latest version is Llama 3.1. GPT, a LLM created by OpenAI, was launched on March 14, 2023. Gemma, a LLM developed by Google, is derived from the same research and technology used to create the Gemini Models. Its initial release is on February 21, 2024. The latest version is Gemma 2. FinBERT, which was put forward by [10], is a pre-trained BERT model in the finance domain to analyze sentiment of financial text. VADER is a lexicon and rule-based sentiment analysis model which is specifically applied to sentiments in social media. VADER will output mainly four scores, compound, pos, neu and neg. Compound score is calculated by summing the valence scores of each word in the lexicon and then normalized to be between -1 and 1. pos, neu and neg scores are ratios for proportion of texts that fall in each category (positive, neutral, and negative, respectively) [29].

The following two datasets are used in the experiment.

Labelled financial news dataset, used in [30], consists of 400 pieces of news with labels positive, negative and neutral. Data was annotated by several readers with a competent background in statistics, Indian stock market and data science.

TABLE I
SENTIMENT CLASSIFICATION MODELS

Model	Model Version
Llama	llama3-8b-8192
GPT	gpt-4-0613
Gemma	gemma2-9b-it
Mixtral	mistral-7b-bnb-4bit
FinBERT	-
VADER	-

Financial Phrasebank, created by [31], is a sentiment dataset of sentences of financial news. It consists of about 2000 news sentences, which are annotated by 16 people with background in finance and there are 5 to 8 annotations per sentence. Sentiment labels are positive, negative and neutral.

B. Stock Price Prediction

The overall structure of the stock price prediction is illustrated in Figure 2. Three features (technical indicators, macroeconomic parameters, news sentiments) are used as inputs. News sentiment is acquired by leveraging LLMs to classify the financial news headlines of the target companies. Seven different machine learning models are applied to forecast the close price of future one day and 16 days. To evaluate the impact of news sentiment on stock price prediction, model performances are evaluated on three scenarios: without sentiment (**w/o sentiment**), with ground truth polarity score (**GT polarity**), and with predicted sentiment (**predicted sentiment**). Polarity score measures the overall sentiment of a sentence, ranging between -1 and 1 while predicted sentiments are either positive, negative or neutral.

C. Prediction Model

Models used to predict the stock price include Informer, LSTM, TCN, Transformer, Support Vector Regression (SVR), Random Forest (RF) and Naive Forecast. Naive forecast predicts the future value with previous time step. The hyperparameters (e.g., number of layers, activation function, and dropout rate of the models) are determined using the Optuna Tuner [32]. The following is a brief introduction of the Transformer-based models and traditional models used in stock price prediction.

1) Transformer: Transformer, proposed by [33], is composed of encoder and decoder. Encoder is used for input sequence processing and it mainly consists of self-attention mechanism and feed-forward neural network. Decoder mainly consists of masked self-attention mechanism and feed-forward neural network. Different from the vanilla Transformer model which consists of both encoder and decoder structures, only the Transformer encoder structure is used in this paper. Transformer encoder structure can effectively capture temporal dependencies in time series forecasting.

2) Informer: Informer model addresses the limitations of Transformer model in handling long sequences by introducing three innovative modules, ProbSparse self-attention, self-attention distilling and generative style decoder. ProbSparse

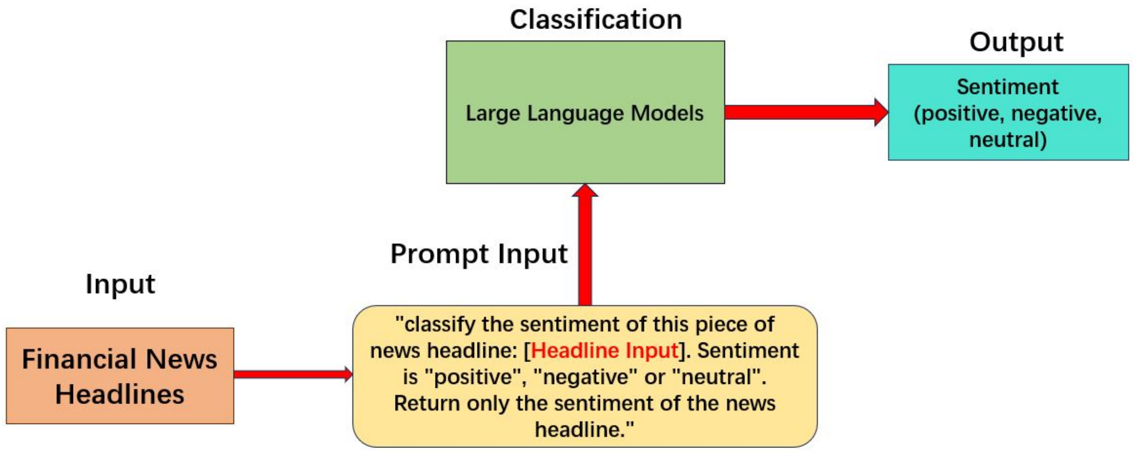

Fig. 1. Sentiment Classification

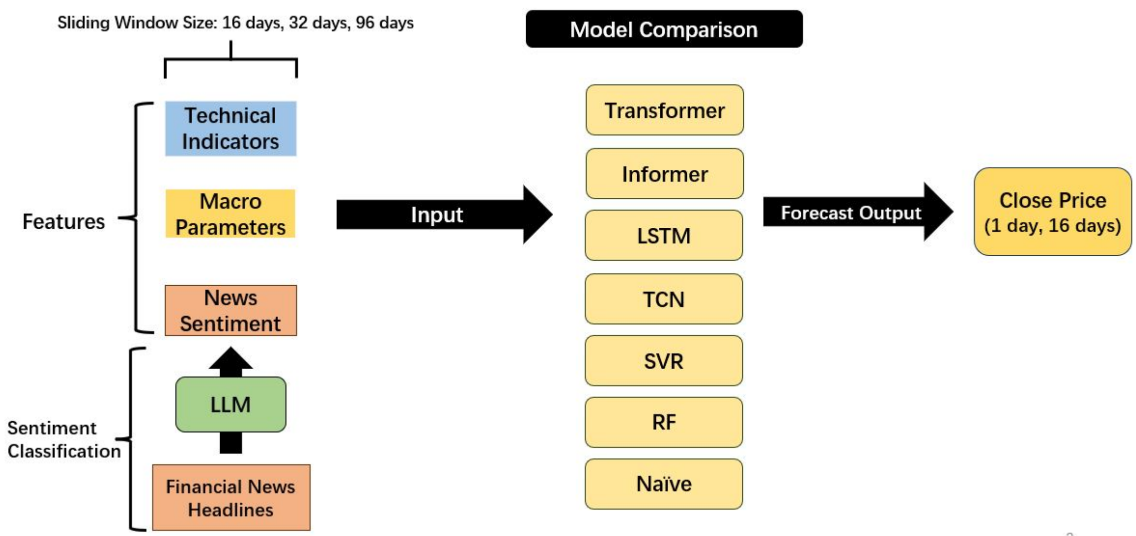

Fig. 2. Stock Prediction

self-attention helps reduce computation costs and memory usage. Self-attention distilling helps receiving long sequence input by trimming the input's time dimension. Generative style decoder helps acquiring long sequence output with only one forward step needed. For Informer model, the same structure as in the original paper of Informer [18] will be used.

3) Long Short-Term Memory: Long Short-Term Memory (LSTM), proposed by [34], is designed to better capture the long-term dependencies and handle the vanishing/exploding gradient problem of RNN models. It consists of a series of memory cells. Each memory cell mainly consists of three components, Forget Gate, Input Gate and Output Gate. Forget Gate controls how much information from the previous hidden state will be removed from the memory cell. Input Gate controls how much information will be added to the memory cell. Output Gate controls how much information will be output from the memory cell.

4) Temporal Convolutional Network: Temporal Convolutional Network, proposed by [35], is designed for sequential data, especially for time series. Different from Recurrent Neural Network, TCN uses convolutional layers with dilation to capture long-term dependencies without dramatically increasing the computational cost, making sure that the network can look back at multiple previous time steps. Use of causal convolutions can ensure that the output at any time is only affected by the past values.

D. Data

The historical stock price data of Toyota (01/05/2015 to 08/05/2024), Apple (30/03/2016 to 08/05/2024) and Tencent (24/07/2017 to 08/05/2024) can be collected from Yahoo Finance by using the Python library yfinance. Financial news headlines of the companies from the same period as the stock price data are collected by using News Feed and News Sentiment data API. Although stock price includes Open Price, Close Price, High Price and Low Price, in this paper, we choose to predict the close price, which is the price at the end of each trading day and also represents the final consensus of the traders and investors each day. It is considered to be the most important price.

Different window sizes are used to study how the models perform with different input lengths and output lengths. The sliding window size of input is 16 days, 32 days and 96 days while the sliding window size of output is 1 day and 16 days. For example, the historical 32 days' data are used to predict the stock price one day ahead. 70% of the data are used for training and 30% of the data are used for testing. MinMax Scaler is applied to both the training and testing datasets to scale the data between 0 and 1 to avoid that the features with larger values dominate the prediction, generating biased results.

E. Features

Apart from news sentiment, in order to fully capture the stock price movement, technical indicators and macroeconomic parameters will also be added to predict the stock price. Technical indicators are often used by traders to predict the stock trend and they are mathematical patterns calculated from historical stock data. Technical indicators used in this paper are shown in Table II. Simple moving average (**SMA**) is the average price of a stock over a specified period. SMA can be used to predict the future stock movement based on the historical stock prices. Relative Strength Index (**RSI**) is a momentum indicator to measure the speed and change of stock price movement. It provides short-term buy and sell signals by tracking the overbought and oversold levels of a stock. On-balance Volume (**OBV**) is a cumulative indicator to measure buying and selling pressure. It adds volume on up days and subtracts volume on down days. **DMI-ADX** is used to measure the strength and direction of a trend. All the technical indicators can be calculated using the Python library Pandas TA.

For macroeconomic parameters, we mainly choose gold price and US Dollar to Japanese Yen Exchange rate. Gold price is the barometer of the economy. First of all, gold can hedge against the inflation. Investors tend to purchase golds to preserve values when the inflation is rising because the purchasing power of currency is decreasing. Furthermore, investors tend to buy more golds during times of economic uncertainty. Therefore, rising gold price can indicate concerns about the economic stability. Exchange rate reflects the relative value of one currency against another. A depreciating currency can lead to higher prices for imported goods and services, which in turn can lead to inflation. Besides, a stable and strong currency can reflect confidence in the economic outlook of a country.

TABLE II
FEATURE TABLE

Technical Indicators	SMA, RSI, OBV, DMI-ADX
Macroeconomic Parameters	Gold Price, USD/JPY Exchange rate

F. Evaluation Metrics

Model performances of stock price prediction are evaluated using two metrics, MSE and MAE. Mean Squared Error (MSE) is the average of the distance between the predictions and the actual values. Mean Absolute Error (MAE) measures the absolute value of the difference between the predicted values and the actual values. In the experiments, the lower the values of MSE and MAE, the better performances the models have.

IV. RESULTS

A. Sentiment Classification

We compare the performances of four generative LLMs together with FinBERT and VADER in sentiment classification using two different datasets. The results of the experiments are shown in Table III.

TABLE III
SENTIMENT CLASSIFICATION ACCURACY SCORE

Model	Financial Phrasebank	Lablled Financial Data
GPT 4	0.966	0.716
Llama 3	0.893	0.779
Gemma 2	0.897	0.799
Mistral 7b	0.813	0.770
FinBERT	0.920	0.513
VADER	0.580	0.582

Based on the experimental results, all LLMs have really good performances in two datasets but models tend to perform better when using Financial Phrasebank dataset. When using the Financial Phrasebank Dataset, GPT 4 performs best, followed by FinBERT, which is expected as FinBERT is pretrained on this dataset. However, when tested on the Labelled Financial Data, all other LLMs outperform both FinBERT and VADER, suggesting that FinBERT struggles to produce consistent results across different datasets. Due to the stable performances of Llama 3 and Gemma 2 and because of the fact that the performances of these two models have trivial differences, we decided to use Llama 3 to classify the sentiments of the financial news headlines of the selected companies. Besides, we also try to fine-tune the large language models but the results indicate that fine-tuning cannot improve the model performances in this case.

B. Stock Price Prediction

Tables VI to VIII show 1-day prediction results and Tables IX to XI show 16-day prediction results. Model with the best performance in each scenario is in **bold font**, same model with the best performance in each look-back period is marked with Asterisk ($\star$). All numbers are percentages, and only two decimal places are shown. MAE is also calculated and its values indicate the same model performances as MSE so only the MSE is shown below.

Generally speaking, for most of the time, predictions with sentiment data have better performances than those without sentiment data, which indicates that adding news sentiment is more likely to improve stock price prediction. Besides, it is more likely for models to demonstrate better performances with ground truth polarity than using the predicted sentiment. For example, in Table IX, models with ground truth polarity always have the best performances.

For most of the time, Informer has the best performances among all models, which is followed by Transformer. Sometimes, Transformer can outperform Informer, like in Table VI when forecasting Apple's 1-day close price with ground truth polarity. In most cases, Informer model is the only model which can outperform Naive forecast. Furthermore, results show that TCN generally can outperform LSTM.

When predicting short-term time series (1-day close price), the Naive Forecast performs nearly the best among all models, second only to Informer. However, when forecasting longer term sequences like 16-day close price, it is more likely for other models to outperform Naive Forecast.

C. Ablation Study of Informer Model

To further analyze the key components of the Informer model, we conducted an ablation study. In each experiment, one important component of the Informer model is removed. This allows us to observe how the model responds to the absence of each component. Informer model mainly has three innovative modules, ProbSparse Attention, Self-Attention Distilling and Generative Style Decoder. Apart from these three modules, we also remove the encoder of Informer to see how well the model will perform with only a decoder because the encoder input contains more information than the decoder input. In the experiments, we use two look-back periods (16 days and 32 days) to predict the future stock prices in 1 day, 8 days and 16 days. The results are shown in Table IV and Table V. Based on the results, we can see that when predicting shorter time sequences (1 day prediction), simple structure model like decoder only model is good enough to generate a good result and it can outperform Informer model with full structure. When predicting longer time sequences (8 days & 16 days predictions), Informer model with full structure is more likely to outperform other model structures. Besides, when we remove generative style decoder from Informer, the performances tend to become the worst. Therefore, generative style decoder seems to play the most important role in Informer to improve prediction accuracy. Generative inference in Informer decoder combines the historical time stamps and the output sequences as the decoder input.

TABLE IV
INFORMER ABLATION STUDY: MEAN SQUARED ERROR (%), LOOK-BACK: 16 DAYS

	1 day	8 days	16 days
Informer	0.057	0.052	0.065
Informer (Decoder Only)	0.013	0.063	0.114
with Full Attention	0.064	0.039	0.051
w/o Self-attn Distilling	0.015	0.063	0.058
w/o ProbSparse Attn & Distilling	0.028	0.058	0.151
w/o Generative Decoder	0.080	0.101	0.126

TABLE V
INFORMER ABLATION STUDY: MEAN SQUARED ERROR (%), LOOK-BACK: 32 DAYS

	1 day	8 days	16 days
Informer	0.032	0.057	0.039
Informer (Decoder Only)	0.034	0.131	0.166
with Full Attention	0.054	0.075	0.077
w/o Self-attn Distilling	0.008	0.068	0.091
w/o ProbSparse Attn & Distilling	0.034	0.090	0.052
w/o Generative Decoder	0.070	0.093	0.133

V. CONCLUSION

This paper proposed integrating the news sentiment data to forecast the stock prices. News sentiments are predicted using the Llama 3 model. Six different machine learning models are used for comparison to forecast the stock prices. Experiments with and without sentiment data are also conducted. The main findings of this paper are as follows. (1) Generative LLM-based sentiment classification can outperform commonly used models like FinBERT and VADER. (2) Models with sentiment data as input have better prediction performances. (3) Informer model yields the best results in most scenarios. (4) Through ablation study, generative style decoder in Informer is the most important component to improve prediction performance. For future work, this research can be improved from several aspects. First, only three stocks are used in this paper, we will increase the number of stocks for analysis to get more general results. Furthermore, since for most of the time, GT polarity score outperforms predicted sentiment, which suggests that there may be limitations to perform sentiment analysis solely on news headlines. Therefore, we will try to use news articles in sentiment analysis to include more information. Besides, only one LLM prompt is considered in this paper, model performances can be evaluated using different prompts in the future.

REFERENCES

[1] J. M.-T. Wu, Z. Li, N. Herencsar, B. Vo, and J. C.-W. Lin, "A graph-based CNN-LSTM stock price prediction algorithm with leading indicators," *Multimedia Systems*, vol. 29, no. 3, pp. 1751–1770, Jun. 2023. [Online]. Available: https://doi.org/10.1007/s00530-021-00758-w

[2] L.-P. Chen, "Using Machine Learning Algorithms on Prediction of Stock Price," *Journal of Modeling and Optimization*, vol. 12, pp. 84–99, Dec. 2020.

[3] Q. Wang and Y. Yuan, "Stock price forecast: Comparison of lstm, hmm, and transformer," in *Proceedings of the 2nd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2023)*. Atlantis Press, 2023, pp. 126–136.

[4] N. Malibari, I. Katib, and R. Mehmood, "Predicting stock closing prices in emerging markets with transformer neural networks: The saudi stock exchange case," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, 2021.

[5] B. A. Abdelfattah, S. M. Darwish, and S. M. Elkaffas, "Enhancing the prediction of stock market movement using neutrosophic-logic-based sentiment analysis," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 19, pp. 116–134, Jan. 2024.

TABLE VI
MEAN SQUARED ERROR (%): APPLE, 1 DAY PREDICTION

Lookback	Scenario	Transformer	Informer	LSTM	TCN	SVR	RF	Naive
96 days	w/o sentiment	0.87	**0.11**	0.72	0.40	2.32	1.84	0.15
	GT polarity	**0.00**⋆	**0.00**⋆	0.27⋆	0.19⋆	2.05⋆	1.43⋆	0.15
	predicted sentiment	0.94	**0.05**	0.63	0.35	2.17	1.47	0.15
32 days	w/o sentiment	0.96	**0.06**	0.64	0.47	2.06	0.36⋆	0.15
	GT polarity	**0.00**⋆	0.01⋆	0.07⋆	0.07⋆	0.76⋆	0.39	0.15
	predicted sentiment	1.00	**0.02**	0.71	0.39	1.32	0.38	0.15
16 days	w/o sentiment	0.41	**0.04**	0.65	0.41	1.76	0.37	0.15
	GT polarity	**0.00**⋆	**0.00**⋆	0.14⋆	0.08⋆	0.65⋆	0.31⋆	0.15
	predicted sentiment	0.42	**0.04**	0.72	0.34	1.18	0.37	0.15

TABLE VII
MEAN SQUARED ERROR (%): TENCENT, 1 DAY PREDICTION

Lookback	Scenario	Transformer	Informer	LSTM	TCN	SVR	RF	Naive
96 days	w/o sentiment	7.06	**0.03**⋆	0.77	0.46	1.38⋆	0.45	0.18
	GT polarity	3.57	**0.03**⋆	0.73	0.46	1.71	0.45	0.18
	predicted sentiment	0.49⋆	**0.08**	0.45⋆	0.19⋆	3.19	0.11⋆	0.18
32 days	w/o sentiment	3.07	**0.04**	0.76	0.73	1.67	0.56	0.18
	GT polarity	2.56	**0.03**⋆	0.66	0.71	1.79	0.56	0.18
	predicted sentiment	0.19⋆	**0.07**	0.37⋆	0.20⋆	0.72⋆	0.15⋆	0.18
16 days	w/o sentiment	0.13⋆	**0.00**⋆	0.27	0.18⋆	0.36⋆	0.14⋆	0.18
	GT polarity	1.91	**0.12**	0.30	0.18⋆	1.57	0.62	0.18
	predicted sentiment	0.22	**0.04**	0.24⋆	0.22	0.38	0.14⋆	0.18

TABLE VIII
MEAN SQUARED ERROR (%): TOYOTA, 1 DAY PREDICTION

Lookback	Scenario	Transformer	Informer	LSTM	TCN	SVR	RF	Naive
96 days	w/o sentiment	0.08⋆	**0.01**	0.17⋆	0.11	0.38⋆	0.24	0.04
	GT polarity	0.14	**0.04**	1.14	0.31	0.89	0.90	0.04
	predicted sentiment	0.39	**0.00**⋆	0.42	0.09⋆	4.19	0.22⋆	0.04
32 days	w/o sentiment	0.08	**0.03**	0.29⋆	0.08⋆	0.24⋆	0.07⋆	0.04
	GT polarity	**0.01**⋆	0.05	0.42	0.25	1.03	0.68	0.04
	predicted sentiment	0.53	**0.01**⋆	0.59	0.11	2.66	0.08	0.04
16 days	w/o sentiment	0.11	**0.07**	0.13⋆	0.12	0.16⋆	0.06⋆	0.04
	GT polarity	0.05⋆	**0.01**⋆	0.59	0.21	0.28	0.57	0.04
	predicted sentiment	0.19	**0.03**	0.45	0.09⋆	2.46	0.06⋆	0.04

TABLE IX
MEAN SQUARED ERROR (%): APPLE, 16 DAYS PREDICTION

Lookback	Scenario	Transformer	Informer	LSTM	TCN	SVR	RF	Naive
96 days	w/o sentiment	1.10	**0.20**	2.10	1.86	3.07	7.23	2.10
	GT polarity	0.04⋆	**0.01**⋆	1.62⋆	0.82⋆	2.13⋆	2.15⋆	2.10
	predicted sentiment	3.43	**0.37**	2.39	1.76	2.58	7.28	2.10
32 days	w/o sentiment	2.33	**0.19**	2.85	1.80	2.58	4.48	2.10
	GT polarity	0.05⋆	**0.01**⋆	0.43⋆	0.32⋆	0.89⋆	0.86⋆	2.10
	predicted sentiment	1.85	**0.08**	2.75	1.77	1.95	4.48	2.10
16 days	w/o sentiment	2.45	**0.09**	2.14	1.89	2.14	2.68	2.10
	GT polarity	0.34⋆	**0.08**⋆	0.50⋆	0.28⋆	0.68⋆	0.56⋆	2.10
	predicted sentiment	1.09	**0.13**	1.90	2.00	1.79	2.67	2.10

TABLE X
MEAN SQUARED ERROR (%): TENCENT, 16 DAYS PREDICTION

Lookback	Scenario	Transformer	Informer	LSTM	TCN	SVR	RF	Naive
96 days	w/o sentiment	1.21⋆	**0.33**	1.60	1.48	2.06	0.82⋆	1.79
	GT polarity	3.58	**0.08**⋆	1.41⋆	1.52	1.47⋆	1.62	1.79
	predicted sentiment	5.83	**0.27**	2.19	0.79⋆	6.50	0.82⋆	1.79
32 days	w/o sentiment	0.64⋆	**0.16**	2.19	1.50	1.80⋆	0.97⋆	1.79
	GT polarity	3.49	**0.07**⋆	2.01⋆	1.42	1.95	1.89	1.79
	predicted sentiment	0.69	**0.20**	2.73	1.10⋆	2.43	0.97⋆	1.79
16 days	w/o sentiment	1.10	**0.10**	1.57⋆	1.32	1.48⋆	0.92⋆	1.79
	GT polarity	10.31	**0.06**⋆	2.02	1.42	1.97	1.70	1.79
	predicted sentiment	0.95⋆	**0.20**	1.61	1.28⋆	1.73	0.93	1.79

TABLE XI
MEAN SQUARED ERROR (%): TOYOTA, 16 DAYS PREDICTION

Lookback	Scenario	Transformer	Informer	LSTM	TCN	SVR	RF	Naive
96 days	w/o sentiment	5.52	**0.26**	0.84	0.75*	0.97	1.03	0.62
	GT polarity	2.34	**0.11***	0.70*	0.81	0.97	1.03	0.62
	predicted sentiment	1.36*	**0.45**	1.62	1.00	5.10*	1.03	0.62
32 days	w/o sentiment	2.57	**0.10**	0.95	0.80	0.49*	0.96	0.62
	GT polarity	0.30*	**0.16**	0.85*	0.77*	0.49*	0.96	0.62
	predicted sentiment	0.64	**0.07***	0.92	0.77*	4.73	0.96	0.62
16 days	w/o sentiment	0.17*	**0.10**	0.79*	0.69	0.44	0.59	0.62
	GT polarity	0.31	**0.19**	0.93	0.55*	0.44	0.59	0.62
	predicted sentiment	0.17*	**0.08***	0.90	0.69	0.44	0.59	0.62

[6] Q. Xiao and B. Ihnaini, "Stock trend prediction using sentiment analysis," *PeerJ Computer Science*, vol. 9, p. e1293, Mar. 2023.

[7] J. Maqbool, P. Aggarwal, R. Kaur, A. Mittal, and I. A. Ganaie, "Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach," *Procedia Computer Science*, vol. 218, pp. 1067–1078, 2023.

[8] P. Koukaras, C. Nousi, and C. Tjortjis, "Stock market prediction using microblogging sentiment analysis and machine learning," *Telecom*, vol. 3, pp. 358–378, May 2022.

[9] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*. Newark, CA, USA: IEEE, Apr. 2019, pp. 205–208.

[10] D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," 2019. [Online]. Available: https://arxiv.org/pdf/1908.10063

[11] N. Darapaneni, A. R. Paduri, H. Sharma, M. Manjrekar, N. Hindlekar, P. Bhagat, U. Aiyer, and Y. Agarwal, "Stock price prediction using sentiment analysis and deep learning for indian markets," Apr. 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2204.05783

[12] J.-X. Liu, J.-S. Leu, and S. Holst, "Stock price movement prediction based on stocktwits investor sentiment using FinBERT and ensemble SVM," *PeerJ Comput. Sci.*, vol. 9, no. e1403, p. e1403, Jun. 2023. [Online]. Available: https://peerj.com/articles/cs-1403/

[13] G. Mu, N. Gao, Y. Wang, and L. Dai, "A stock price prediction model based on investor sentiment and optimized deep learning," *IEEE Access*, pp. 51 353–51 367, 2023.

[14] S. Wu, Y. Liu, Z. Zou, and T.-H. Weng, "S_i_lstm: stock price prediction based on multiple data sources and sentiment analysis," *Connection Science*, vol. 34, pp. 44–62, Dec. 2022.

[15] S. Chen and H. He, "Stock prediction using convolutional neural network," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 435, p. 012026, Nov. 2018.

[16] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018.

[17] L. Costa and A. Machado, "Prediction of stock price time series using transformers," in *Anais do II Brazilian Workshop on Artificial Intelligence in Finance*. Porto Alegre, RS, Brasil: SBC, 2023, pp. 85–95.

[18] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," 2021.

[19] T. Phuoc, P. T. K. Anh, P. H. Tam, and C. V. Nguyen, "Applying machine learning algorithms to predict the stock price trend in the stock market – the case of vietnam," *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, Mar. 2024. [Online]. Available: https://creativecommons.org/licenses/by/4.0

[20] M. I. K. Jabed, "Stock market price prediction using machine learning techniques," *American International Journal of Sciences and Engineering Research*, vol. 7, no. 1, p. 1–6, Feb. 2024.

[21] Y. E. Gür, "Makine öğrenimi ve derin öğrenme algoritmalarını kullanarak hisse senedi fiyat tahmini: Havacılık sektörüne yönelik bir örnek çalışma," *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, Oct. 2023.

[22] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *J. Finance Data Sci.*, Apr. 2018.

[23] D. G. Jayaswara, I. Slamet, and Y. Susanti, "Prediction of central asia bank's stock price using support vector regression method," *ICRSE*, vol. 2, pp. 7–12, Apr. 2023. [Online]. Available: https://sunankalijaga.org/prosiding/index.php/icrse/article/view/883

[24] O. E. Orsel and S. S. Yamada, "Comparative study of machine learning models for stock price prediction," 2022. [Online]. Available: https://arxiv.org/abs/2202.03156

[25] M. S. Haque, M. S. Amin, J. Miah, D. M. Cao, and A. H. Ahmed, "Boosting stock price prediction with anticipated macro policy changes," 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2311.06278

[26] B. Weng, W. Martinez, Y.-T. Tsai, C. Li, L. Lu, J. R. Barth, and F. M. Megahed, "Macroeconomic indicators alone can predict the monthly closing price of major U.S. indices: Insights from artificial intelligence, time-series analysis and hybrid models," *Appl. Soft Comput.*, vol. 71, pp. 685–697, Oct. 2018.

[27] P. Mishra, P. Pai, P. Singh, S. Kulkarni, and S. A. Weakey, "Analysis of effect of historical prices and news on the stock market," in *2021 International Conference on Communication information and Computing Technology (ICCICT)*. IEEE, Jun. 2021.

[28] T. Sun, J. Wang, P. Zhang, Y. Cao, B. Liu, and D. Wang, "Predicting stock price returns using microblog sentiment for chinese stock market," in *2017 3rd International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE, Aug. 2017.

[29] C. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, May 2014.

[30] G. Nath, A. Sood, A. Khanna, S. Wilson, K. Manot, and S. K. Durbaka, "On quantifying sentiments of financial news – are we doing the right things?" 2023. [Online]. Available: https://arxiv.org/abs/2312.14978

[31] P. Malo, A. Sinha, P. Takala, P. Korhonen, and J. Wallenius, "Good debt or bad debt: Detecting semantic orientations in economic texts," 2013.

[32] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: https://www.bioinf.jku.at/publications/older/2604.pdf

[35] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," 2016. [Online]. Available: https://arxiv.org/abs/1608.08242