IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.DOI

Transformer-Gated Recurrent Unit Method for Predicting Stock Price based on News Sentiments and Technical Indicators

AGUS TRI HARYONO, RIYANARTO SARNO, (SENIOR MEMBER, IEEE), KELLY ROSSA SUNGKONO
Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

Corresponding author: Riyanarto Sarno (e-mail: riyanarto@if.its.ac.id)

This research was funded by Asian Development Bank under HETI ADB Project; The Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program; Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Dana Departemen Program; and Institut Teknologi Sepuluh Nopember (ITS) under project scheme of the Publication Writing and IPR Incentive Program(PPHKI).

ABSTRACT News sentiment can influence stock prices indirectly, in addition to the technical indicators already used to analyze stock prices. The information quantification of news sentiment by considering time sequence data in the stock analysis has been the primary issue; this article proposes methods for quantifying news sentiments by considering time sequence data. The news sentiment quantification uses a daily confidence score from the classification model. The active learning model uses to build a classification model considering time sequence data, which results in sentiment indicators. Then the sentiment indicators are utilized by stock price forecasting using the proposed Transformer Encoder Gated Recurrent Unit (TEGRU) architecture. The TEGRU consists of a transformer encoder to learn pattern time series data with multi-head attention and pass it into the GRU layer to determine stock price. The accuracy mean absolute percentage error (AcMAPE) uses to evaluate forecasting models sensitive to the misclassification of stock price trends. Our experiment showed that the sentiment indicator could influence stock issuers based on the increased performance of the stock price forecasting model. The TEGRU architecture outperformed other transformer architecture on five feature scenarios. In addition, TEGRU presented the best-fit parameters to produce low financial risk for each stock issuer.

INDEX TERMS active learning, financial loss, quantifying news sentiment, stock price prediction, transformer time series

I. INTRODUCTION

SENTIMENT analysis research is increasingly developing, making stock price analysis break through the limitation of fundamental and technical indicators as forecasting features. The sentiment analysis results in qualitative data cannot be combined with technical indicators in quantitative data to forecast stock prices. The processing of converting qualitative into quantitative data type is needed [1]. In previous research, quantitative sentiment represents positive or negative stock market conditions [2]. The average sentiment value was used as data training to produce stock price forecasting [3] and the stock trend prediction model [4]. The stock market is represented by financial news sentiments and multi aspects, such as corporate, stock, market, and economy news [5]. The multi-aspect sentiment news represents stock market conditions better and analyzes aspects significantly influencing stock price forecasting.

The stock price forecasting model is learned by time series data, which means time sequence data is essential. The aspect sentiment classification model commonly uses large data text that does not consider time sequence data. Therefore, an aspect sentiment classification model needs to be produced using training data that evaluates the data time sequence so that the sentiment aspect results follow the stock transaction data. The active learning method is utilized to extract the aspect sentiment value corresponding to the time sequence of knowledge. Active learning can measure model knowledge of the test data. If model knowledge has a low value on some of the test data, then the test data become data training on the next model training [6]. Active learning is already used in stock market data to produce weight on calculating the distance between stock transactions and text data. The threshold on active learning based on company pairs distance with highest and lowest scores, and precision score [7]. In previous research, active learning is used to select valuable time series data samples to reduce training complexity [8]. Our research uses active learning to produce a monthly model appropriate to time sequence data. The monthly model is used as aspect sentiment feature extraction, and the indicators represent knowledge according to the time sequence of stock prices.

Various methods have already been used to forecast stock prices, from statistics [9], machine learning [10], and deep learning [11]. The Transformer architecture became a breakthrough in the deep learning method because the Transformer had a multi-head attention layer to learn a pattern of data input and an encoder-decoder section with different input layers to resolve transformation data tasks [12]. The Transformer is already used in Natural Language Processing (NLP) [13], Vision [14], and Time Series [15], but few used transformer architectures for forecasting stock price tasks. The Gated Three-Tower Transformer (GT3) is a developed Transformer architecture for stock market prediction using numerical and text input. Two Transformers are used to learn numerical data, channel-wise and time-wise embedding, and one Transformer to learn daily social text with Bidirectional Encoder Representations from Transformers (BERT) sentence embedding [16]. The Recurrent Neural Network (RNN) architecture has an excellent reputation for stock forecasting [17] [18]. The characteristics of RNN architecture can learn patterns in sequence data. Combination Transformer and RNN architecture are already used in stock price prediction. The FDG-Trans is built from frequency decomposition, LSTM, GRU, and Transformer architecture. The stock price as input data is transformed into another form using Complete Ensemble Empirical Mode Decomposition (CEEMD). Then, the sequence matrix passes to GRU, LSTM, and Transformer layer [19]. Our research proposes a combination of Transformer and GRU architecture, in which the Transformer is a learner of input features, and GRU is a learner of sequence data to predict stock prices. The input data is numerically represented by indicator technical and news sentiment in quantitative form.

Evaluation model regression performance used Mean Square Error (MSE), Mean Square Percentage Error (MSPE), and Mean Absolute Percentage Error (MAPE) [18]. The evaluation method only calculates errors on actual and predicted values without considering value movement classification. The evaluation method by considering trend classification is vital in evaluating regression forecasting model performance with financial risk. Even though error value model regression is low, if trend classification is miss predicted, monetary loss is inflicted.

In previous studies, the quantification sentiment in stock forecasting did not use multi-aspect with data time sequence attention to represent stock market conditions. To improve the performance of stock price forecasting, in this paper, we present feature extraction that uses an active learning model to produce a multi-aspect sentiment indicator with time sequence data attention. The sentiment and technical indicators are utilized for stock price forecasting using Transformer Encoder Gated Recurrent Unit (TEGRU). The TEGRU can extract information with a multi-head attention layer and recognize pattern sequence data. The evaluation performance of the forecasting model uses Accuracy Mean Absolute Percentage Error (AcMAPE), which have sensitive to the misclassification of stock price trends.

The significant contributions of this paper are summarized as follows.

1) The feature extraction method uses an active learning model to produce a multi-aspect sentiment indicator with time sequence data consideration to describe detailed stock market conditions.
2) The Transformer Encoder Gated Recurrent Unit (TEGRU) architecture, which consists of a Transformer Encoder as extraction information and GRU as a recognized sequence data pattern, is used in stock price forecasting.
3) The Accuracy Mean Absolute Percentage Error (AcMAPE) is an evaluation performance method on a forecasting model, which have sensitive to the misclassification of stock price trends.
4) The impact of the TEGRU feature parameter on stock price forecasting is analyzed to produce a low financial risk model for each stock issuer.

The rest of the article is organized as follows: Section 2 reviews the related work; Section 3 presents the proposed architecture; Section 4 experiments and results, Section 5 discusses, followed by a conclusion in Section 6.

II. RELATED WORKS

Various ideas and opinions have already solved forecasting stock prices. However, forecasting stock price still has appeal to resolve with a minimum result of financial risk. A brief review of the literature on predicting stock prices based on features is presented in this section, followed by some related works on the active learning method and associated issues of transformer architecture implemented with data time series.

A. ASPECTS SENTIMENT CLASSIFICATION

The aspects in the sentence are the attributes contained in the topic of discussion in the text sentiments. The aspect categorization process can use the semantic relation method or deep learning models [5]. The semantic relation method would extract terms from the text and then measure semantic similarity to the aspect term list. The higher the semantic similarity, the higher the value obtained [20]. The aspect categorization model is trained using data sets with aspect labels in the deep learning method. Therefore, the aspect categorization is determined by the model that has been trained.

A sentiment is an opinion that contains feelings for an object in the text, and adjectives are commonly used as sentiment expressions in sentences. Sentiment classification has been used to accomplish dataset political [21], review [22], hate speech [23], and financial [24]. In political datasets, sentiment classification is resolved using a lexicon list sentiment. The appearance of the word lexicon in the text will determine the sentiment label. Review, hate speech, and financial datasets use deep learning methods as sentiment classifications. The sentiment classification model is trained using a sentiment-label dataset.

B. STOCK PRICES PREDICTIONS

Various approaches have already been made to stock analysis. A classical method such as ARIMA is used to forecast stock prices, but the limitation is that ARIMA cannot recognize significant nonlinear data [9]. Three phases, which are model identification, parameter estimation, and diagnostic checking, are used in the ARIMA method. To break the limitations of the classical method, ARIMA combined with SVM machine learning can deal with highly nonlinear data [25]. ARIMA is used as a feature selection tool, and SVM is used to build the forecasting model. In addition, deep learning is used to gain better results, such as with CNN and LSTM. In previous research, the CNN-BiLSTM combined layer was used to build a stock price forecasting model, with the CNN layer processing raw input through an extraction feature, which was then passed to the BiLSTM layer to learn predicted stock prices [26]. The feature can be a technical indicator, sentiment analysis, or both.

Historical stock transaction processing using statistical methods such as the moving average yielded information about volatility, trend, and momentum. The stock trader uses the volatility, trend, and momentum as a signal to buy or sell stock. In previous research, a pre-processing step is used to reduce the dimension or select an essential feature of the technical indicator. Principal Component Analysis (PCA) and the LASSO method are implemented to minimize 53 indicators and used to forecast the Shanghai Composite Index. The result is that LASSO had better prediction performance than PCA [11]. The other method is feature decomposition, such as Empirical Wavelet Transform (EWT), which adapts stock prices into several sub-layers. On forecasting daily stock closing prices in the United States and China, the combined method EWT-dpLSTM-PSO-ORELM performed better [27].

Complex factors influence closing stock prices. News sentiment is one of them. The relationship between news sentiment and stock prices in the US market has been discovered [28]. Challenge quantifies the effects of news sentiment on stock forecasting. Sentiments provide qualitative data, whereas stock forecasting needs quantitative data as a feature. In previous research, sentiment results in quantitative data can be provided by some techniques. TextBlob and VADER lexicon libraries can determine the polarity value of text with a range of -1 to 1. RNN and BERT methods provide polarity sentiment with a scale of 0 to 1 [1]. Combined technical indicators and news sentiment are already used to resolve stock forecasting tasks. The hybrid model, which uses CNN to extract sentiment polarity and combines technical indicators to build an LSTM forecasting model, produced good results when forecasting the Shanghai Stock Exchange [29].

C. ACTIVE LEARNING OF TRANSFORMER FINE-TUNING

Deep learning methods require active learning to ensure the resulting model can learn automatically with as little human intervention as possible. The output of active learning is to determine which data can be automatically fed into the model as training data and which requires human labeling [6]. In addition, active learning can produce ever-evolving models based on the latest data. The active learning model closely matched time series data, which depended on the time sequence.

Transformer architecture is a profound learning model breakthrough, particularly for text-based topics. BERT is a pre-trained model of the transformer architecture encoder section. Developing a BERT pre-trained model aims to create a general pre-trained model that can handle various NLP tasks [13]. The BERT model has a large variant, siebert/sentiment-roberta-large-english variant in English, with good performance. Fine-tuned from the RoBERTa-large model [30], siebert/sentiment-roberta-large-english had an outstanding performance on text classification after being tested on 15 data sets [31]. IndoBERT is an Indonesian language variant of the BERT model. IndoBERT outperformed 8 of 12 classification and sequence labeling tasks after being trained on 15 data sets containing over 275 million sentences [32].

D. GATED RECURRENT UNIT (GRU)

Time series forecasting already uses deep learning methods like recurrent neural networks. The simplest variant of RNN architecture to predict stock prices is the gated recurrent unit (GRU) [11]. The GRU layer had slightly better than the LSTM layer on stock price trend prediction [33]. The most straightforward architecture because, compared with LSTM, GRU only has two gates in a cell: update and reset gate. The update gate u_t had a task to decide the amount of information from input x_t and previous output $h_{(t-1)}$ to be retained in the current state using activation function σ. The information from the previous state would be released more frequently if the update gate had a lower value. The reset gate r_t with activation function σ had to control how many previous pieces of information to forget; if the forget gate had a higher value, less information would be ignored. The update and reset gates information is saved into the memory unit in the

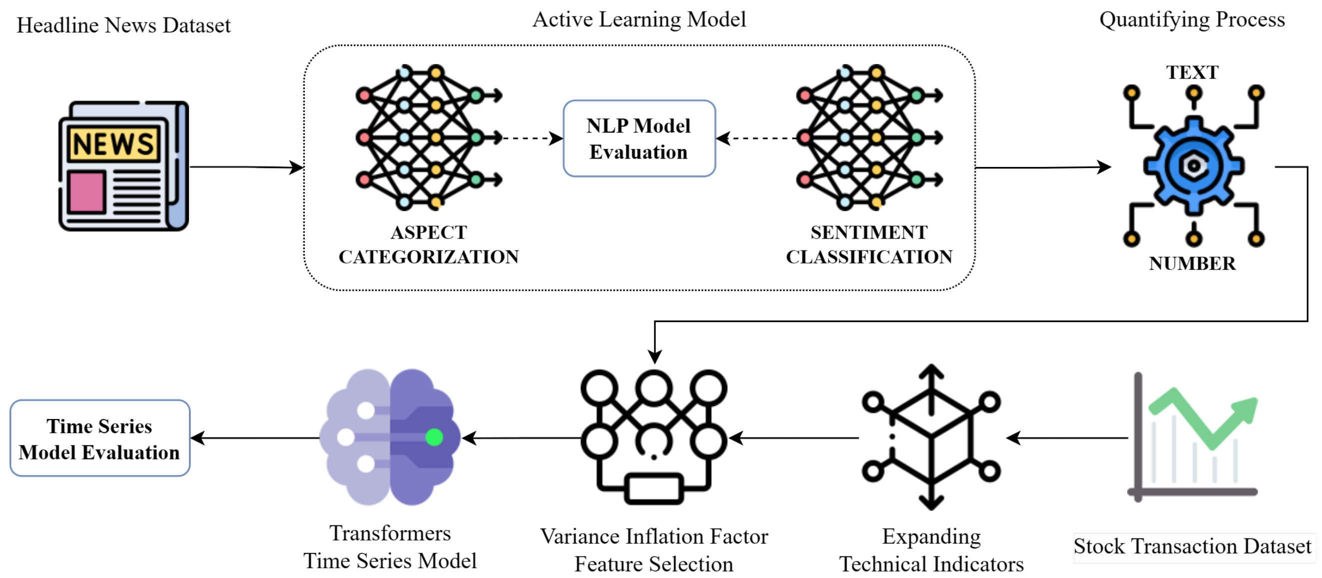


FIGURE 1. Proposed architecture for stock forecasting.

candidate current state $\widetilde{h}_t$ using $tanh$ activation function and then passed to the next cell until the final output h_t is reached. Each gate has a random weight represented by W_u, W_r, $W_{\widetilde{h}}$. Implementation of the GRU cell uses formulation (1) to (4).

$$u_t = \sigma(W_u \times [h_{(t-1)}, x_t]) \tag{1}$$

$$r_t = \sigma(W_r \times [h_{(t-1)}, x_t]) \tag{2}$$

$$\widetilde{h}_t = tanh(W_{\widetilde{h}} \times [r_t \times h_{(t-1)}, x_t]) \tag{3}$$

$$h_t = (1 - u_t) \times h_{(t-1)} + u_t \times \widetilde{h}_t \tag{4}$$

E. TRANSFORMER TIME SERIES

The architecture of the Transformer has already been implemented in the text, image, voice, and time series domains [34]. Transformer on time series tasks is implemented using the original Transformer's architecture, which includes encoder and decoder sections. The encoder section consists of an input layer, a positional encoding layer, and four encoder layers, including a self-attention layer, a normalization layer, and a feed-forward layer. The decoder part begins with the input layer and pass-through four decoder layers, which contain the self-attention layer, normalization layer, encoder-decoder attention layer, and feed-forward layer, before reaching the decoder output. The encoder and decoder parts are linked via the encoder output. On influenza-like illness (ILI) data sets, the transformer time series architecture outperformed ARIMA, LSTM, and Seq2Seq+attn [35].

FEDformer is another transformer of time series architecture that uses a decomposition block to replace the self-attention layer and a decoder part that includes seasonal and trend components. When tested on six data sets, FEDformer outperformed other transformer time series architectures such as AutoFormer, Informer, and Reformer [36] [15].

Our research uses active learning to create a monthly NLP model and reduce data training. NLP models accomplish two tasks: aspect categorization and sentiment classification. The confidence score from NLP models is extracted and used in the forecasting model. Aside from the sentiment feature, technical indicators are also used to predict the stock market's close price. The variation inflation factor as feature selection is used to reduce features from combined multi-aspect sentiment and technical indicators. Fig. 1 shows the architecture of the proposed method in this paper.

III. PROPOSED METHOD

A. DATASET

This sub-section explains the collection, labeling, and preparation of the dataset used in this research. The dataset consists of time series and text; each is prepared independently.

1) Dataset of news headlines

The text dataset used in this work was gathered from the Indonesian news portal at kontan.co.id under the categories investment, finance, industry, national, and international. The scraping technique collects data between 1 January 2017 to 31 March 2022 and produces a title, URL, description, postdate, category, and subcategory. Statistics from collected data are shown in Table 1.

To evaluate performance prediction, task classification in text requires ground truth. The dataset already has ground truth for aspect categorization, but sentiment classification does not have the ground truth label. A transfer learning method can label a dataset as ground truth; however, another issue is that only some highly-accuracy models are available in Indonesian. Labeling ground truth for the sentiment classification task in this research uses transfer learning with

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3298445

TABLE 1. Statistic dataset of headline news.

Category	Sentiment	Count
Industry	Negative	16,765
	Positive	47,417
International	Negative	15,084
	Positive	13,452
Investment	Negative	29,394
	Positive	50,570
Finance	Negative	11,054
	Positive	27,371
National	Negative	28,963
	Positive	41,262
Total		281,332

an English language-trained model. The labeling method is shown in Fig. 2.

The cleaned news title will be translated from Indonesian to English using a translator application. The resulting translated news title will be passed into the sentiment labeling using a trained model siebert/sentiment-roberta-large-english. The label sentiment complements the ground truth of the headline news dataset. A trained model siebert/sentiment-roberta-large-english has an accuracy of 93.2% in a binary sentiment classification task, positive or negative. Table 2 shows an example of a text dataset used to build a model. Preprocessing phase, transform to lowercase, remove HTML tags, and remove the date. The BERT tokenizer is used to tokenize cleaned sentences.

2) Dataset of stock prices

The Python yfinance library is used to collect historical stock market transactions. The results are daily transactions containing information about the transaction date, open price, lowest price, highest price, close price, volume of transactions, dividend, and split stock. Each issuer's Initial Public Offering (IPO) date determines the amount of data provided.

In previous research, forecasting Indonesian stock prices used issuers with large market capitalizations, such as BBCA, BBRI, TLKM, BMRI, and ASII [17]. The similarity of issuers with large market capitalization is that the stock price chart has increased since the IPO date, and the stock price tends to be stable within a year. In this research, in addition

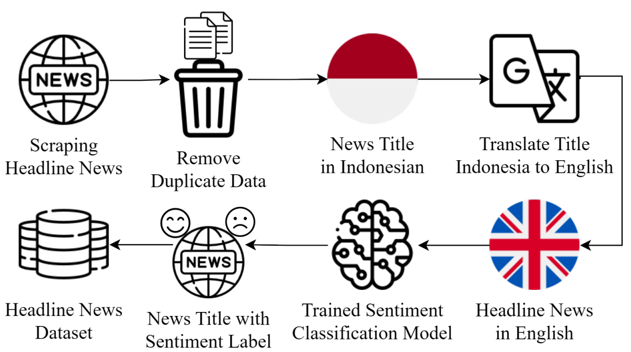


FIGURE 2. The method of Indonesian headline news dataset sentiment labeling.

TABLE 2. Example headlines news on each category and sentiment.

Category	Headline	Sentiment
Industry	Mahaka Media has not felt an increase in the request for videotron advertising in the campaign season. **(EN)** *Mahaka Media belum rasakan peningkatan permintaan iklan videotron di musim kampanye*. **(ID)**	Negative
	Karimun Wagon's exports to Pakistan continue to be improved. **(EN)** *Ekspor Karimun Wagon ke Pakistan terus ditingkatkan*. **(ID)**	Positive
International	Missed the estimated, factory output in Singapore dropped 0.2% in September. **(EN)** *Meleset dari perkiraan, output pabrik di Singapura merosot 0,2% pada September*. **(ID)**	Negative
	The US economy is the third quarter of 2018 slowing down, but still strong. **(EN)** *Ekonomi AS kuartal III 2018 melambat, namun tetap kuat*. **(ID)**	Positive
Investment	Vale Indonesia (INCO) performance can be depressed if unable to overcome this sentiment. **(EN)** *Kinerja Vale Indonesia (INCO) bisa tertekan jika tak mampu atasi sentimen ini*. **(ID)**	Negative
	NFC Indonesia is targeting two cities for expanding digital cloud advertising business. **(EN)** *NFC Indonesia incar dua kota untuk ekspansi bisnis digital cloud advertising*. **(ID)**	Positive
Finance	Not optimal, the distribution of KUR through multifinance is threatened with stagnant. **(EN)** *Tak optimal, penyaluran KUR melalui multifinance terancam mandek*. **(ID)**	Negative
	Although LDR increases, BI ensures that the condition of the banking liquidity is sufficient. **(EN)** *Meski LDR meningkat, BI pastikan kondisi likuiditas perbankan cukup*. **(ID)**	Positive
National	Central Kalimantan DPRD members who were netted by the KPT of the KPK were 8 people. **(EN)** *Anggota DPRD Kalteng yang terjaring OTT KPK sebanyak 8 orang*. **(ID)**	Negative
	The new rules of the sweetened condensed milk label will protect consumers and producers. **(EN)** *Aturan baru label susu kental manis akan melindungi konsumen dan produsen*. **(ID)**	Positive

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3298445

TABLE 3. Issuer stock transaction statistical.

Issuer	First Date	Last Date	Count	Min	Max	Mean
BBCA	08/06/2004	12/05/2022	4,448	108.71	8,200.00	2,404.27
BBRI	10/11/2003	12/05/2022	4,598	56.69	4,940.00	1,560.73
TLKM	28/09/2004	12/05/2022	4,367	503.04	4,770.00	2,122.61
BMRI	14/07/2003	12/05/2022	4,683	182.84	8,950.00	3,083.63
ASII	17/10/2000	12/05/2022	5,397	47.65	7,867.77	3,431.97
INDX	28/05/2001	13/05/2022	5,239	35.00	18,000.00	845.25
SMMT	04/12/2007	13/05/2022	3,407	12.36	6,275.00	474.60
HITS	17/10/2005	13/05/2022	3,736	197.65	2,891.13	544.31

to issuers with large market capitalization, issuers with the highest percentage increase in stock price or also known as top gain, are used. INDX, SMMT, and HITS are the issuers with the highest gains in March 2022. Stock price forecasting uses a combination of extensive market capitalization and top gain issuers to validate the forecasting model on various characteristic issuers in a short period. Profile each stock issuer as shown in Table 3.

B. ACTIVE LEARNING AND QUANTIFYING ASPECT SENTIMENTS INDICATOR

Building a model classification text is usually done only once, whereas forecasting stock prices always requires the most recent data. The active learning method can renew information on the classification model through the re-learning process. Process re-learning can reduce the data in the training dataset because the model only uses data that the model does not know with certainty. The scenario to construct an active learning model is shown in Fig. 3. A pre-trained model BERT is used before building an active learning model to determine whether the fitted pre-trained model matches our dataset. Pretrained models used are indolem/indobert-base-uncased, cahya/bert-base-indonesian-522M, indobenchmark/indobert-base-p2, cahya/distilbert-base-indonesian, and cahya/roberta-base-indonesian-522M [32] [37]. Every pre-trained model used to construct aspect categorization and sentiment classification models with a scenario of split data are January 2017 to February 2022 as data training and March 2022 as data testing. Evaluation aspect categorization and sentiment classification models using confusion matrix and then calculated into accuracy, recall, precision, and f1-score. The pre-trained model with the best performance is used to build an active learning model.

$$H(X) = H(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{5}$$

The initiation to build an active learning model uses January 2017 as training data, and the testing data always uses March 2022. The March dataset was used to test the growth of learning model knowledge in each month. The model will make predictions and then analyze the results every month from February 2017 to February 2022 to determine the data testing that will become data training in the subsequent learning. The confidence score from the prediction result on each sentence was used to choose the data training on active learning, and the calculation used Eq. (5) to produce the entropy score $H(X)$, with p as probability score on each class. The entropy score can describe which data is unknown with the model, and the average entropy score each month is used as the threshold value. If sentence entropy scores more than the threshold value, data testing becomes training for subsequent learning.

Two-level model classification is needed to solve aspect-based sentiment analysis. The first level is the aspect categorization model with five classes, and the second is the sentiment classification model with two types. Both models

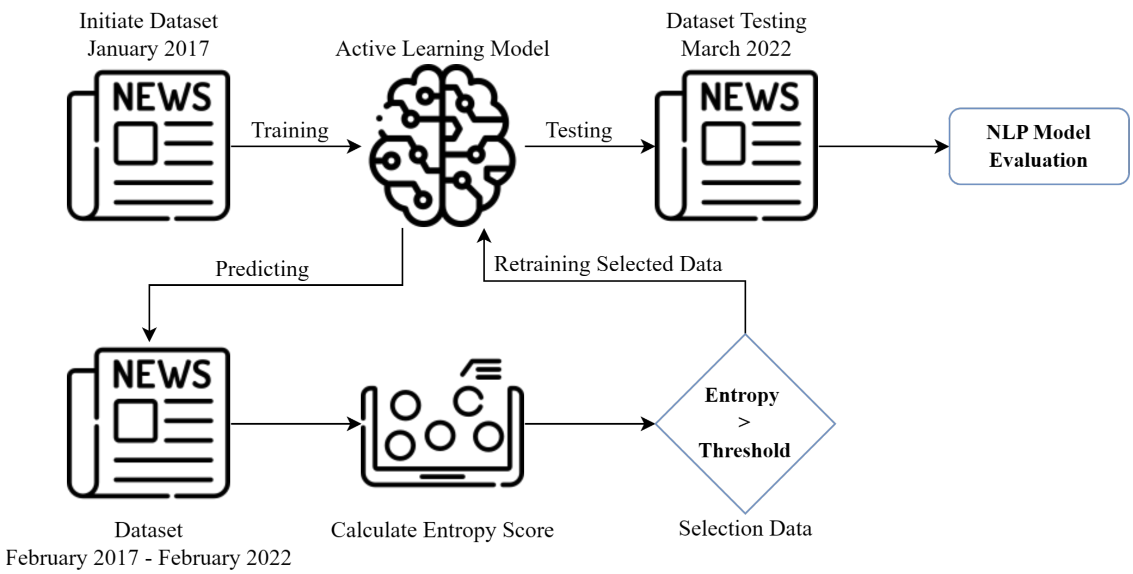


FIGURE 3. The architecture of the active learning model.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3298445

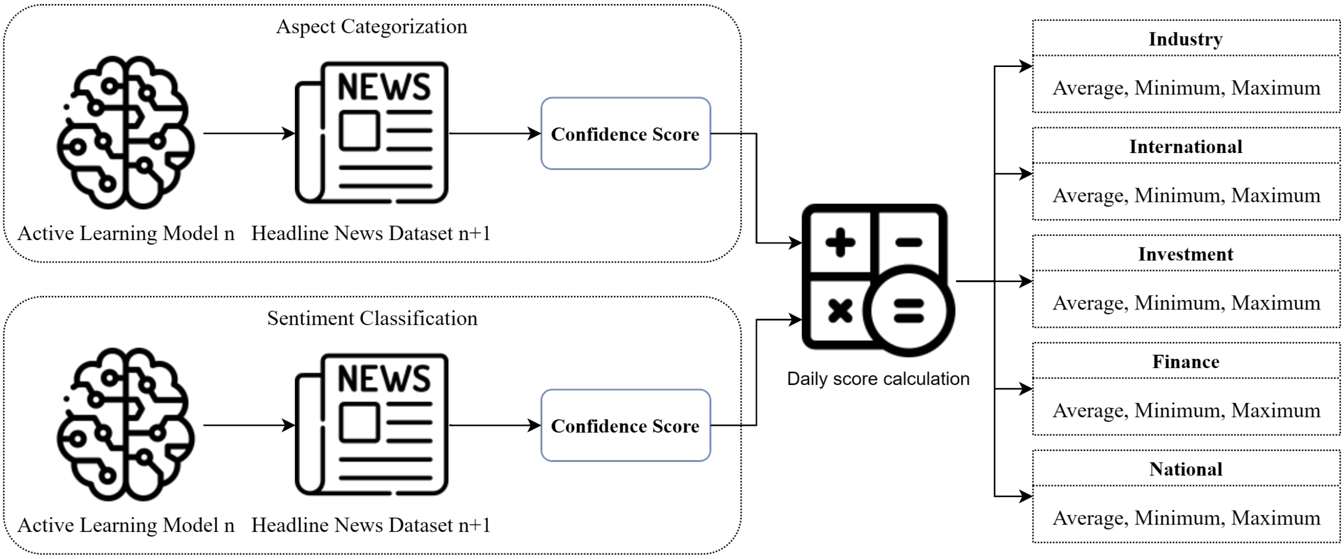


FIGURE 4. Quantifying sentiment label to daily aspect sentiment score. Model active learning (n) month to predict dataset (n+1) month.

use active learning to produce a model each month. As shown in Fig. 4, the active learning model from aspect categorization and sentiment classification is used to predict the dataset from February 2017 to March 2022, with a scenario model this month to predict the next month. The active learning classification models produce a confidence score. The confidence score is used to build aspect sentiment daily indicators, which are the average aspect score, the average aspect sentiment positive score, the average aspect sentiment negative score, the maximum aspect sentiment positive score, the maximum aspect sentiment negative score, the minimum aspect sentiment positive score, and the minimum aspect sentiment negative score. The aspect sentiment indicator with seven features on each aspect is more representative of daily sentiment than two features without aspect, such as the average sentiment positive and the average sentiment negative. In addition, this proposed method produces 35 aspect sentiment features that can utilize to forecast stock prices and combined with the technical indicator to form a quantitative dataset. Table 4 shows the aspect sentiment indicator illustration.

TABLE 4. Sentiment indicator example value.

News Indicator	Aspect				
	Industry	International	Investment	Finance	National
Aspect average	0.8885	0.9462	0.9345	0.9556	0.9887
Positive average	0.9832	0.8432	0.9957	0.7684	0.0
Positive minimum	0.9832	0.6625	0.9952	0.6492	0.0
Positive maximum	0.9832	0.9919	0.9962	0.8928	0.0
Negative average	-0.8061	-0.9506	-0.8604	-0.8810	-0.9551
Negative minimum	-0.6298	-0.8801	-0.6897	-0.6240	-0.9444
Negative maximum	-0.9823	-0.9878	-0.9662	-0.9775	-0.9658

C. EXPAND TECHNICAL INDICATOR

The historical stock transaction dataset consists of the transaction date, open price, lowest price, highest price, close price, and volume of transactions. The expanded stock transaction history can produce other technical indicators, such as momentum, volatility, and trend. This research uses the Python library TA to expand from 6 indicators to 43 additional technical indicators [38]. Descriptions of each expanded indicator are shown in Table 5.

D. MULTIVARIATE FEATURE SELECTION

Aspect sentiments and technical indicators are used as features to forecast each issuer's stock market close price. Correlation analysis is used to reduce multicollinearity between features and targets. Eq. (6) produces each indicator's variance inflation factor (VIF) value. The threshold VIF value is less than 5, which means multicollinearity between features is weak [39]. The number of iterations for selecting features with VIF values increased until all selected features had a VIF value less than the threshold. Correlation analysis features are used on each issuer to inform the selection feature, which significantly impacts the forecasting model.

$$VIF_i = \frac{1}{(1 - R_i^2)} \tag{6}$$

E. TRANSFORMER MODELS

The several baseline models are reproduced as a comparison to the proposed model. The Vanilla Transformer, Informer, and FEDformer are built using the official FEDformer repository. The FEDformer repository was chosen because FEDformer architecture outperformed other transformer families, such as Autoformer, Informer, LogTrans, and Reformer [15]. The implementation of the baseline and proposed model are discussed in this sub-section

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3298445

TABLE 5. Technical indicator expanding.

Indicator	Code	Type
Money Flow Index	MFI	Volume
Accumulation/Distribution Index	ADI	Volume
On-Balance Volume	OBV	Volume
Chaikin Money Flow	CMV	Volume
Force Index	FI	Volume
Ease of Movement	EMV	Volume
Volume-price Trend	VPT	Volume
Negative Volume Index	NVI	Volume
Volume Weighted Average Price	VWAP	Volume
Average True Range	ATR	Volatility
Bollinger Bands	BB	Volatility
Keltner Channel	KC	Volatility
Donchian Channel	DC	Volatility
Ulcer Index	UI	Volatility
Simple Moving Average	SMA	Trend
Exponential Moving Average	EMA	Trend
Weighted Moving Average	WMA	Trend
Moving Average Convergence Divergence	MACD	Trend
Average Directional Movement Index	ADX	Trend
Vortex Indicator	VI	Trend
Trix	TRIX	Trend
Mass Index	MI	Trend
Commodity Channel Index	CCI	Trend
Detrended Price Oscillator	DPO	Trend
KST Oscillator	KST	Trend
Ichimoku Kinko Hyo	Ichimoku	Trend
Parabolic Stop and Reverse	Parabolic SAR	Trend
Schaff Trend Cycle	STC	Trend
Relative Strength Index	RSI	Momentum
Stochastic RSI	SRSI	Momentum
True Strength Index	TSI	Momentum
Ultimate Oscillator	UO	Momentum
Stochastic Oscillator	SR	Momentum
William %R	WR	Momentum
Awesome Oscillator	AO	Momentum
Kaufman's Adaptive Moving Average	KAMA	Momentum
Rate of Change	ROC	Momentum
Percentage Price Oscillator	PPO	Momentum
Percentage Volume Oscillator	PVO	Momentum
Daily Return	DR	Other
Daily Log Return	DLR	Other
Cumulative Return	CR	Other

1) Transformer

The Vanilla Transformer consists of an encoder and a decoder parts, emulating as closely as possible to the original Transformer [12]. The encoder part is stacked with two encoder layers containing multi-head attention, feed-forward, and normalization. The encoder output passes to the decoder part, which contains the same as the encoder part, and then through linear transformation to solve the prediction task. The time series data is processed into input embedding on the encoder and decoder using the token and positional embedding. This architecture does not implement the masking method on the encoder and decoder parts. The Vanilla Transformer architecture will predict stock prices in Indonesian currency one step ahead.

2) Informer

The Informer architecture is another Transformer architecture to resolve long sequence time-series forecasting (LSTF) problems. The vanilla transformer had three significant problems that can tackle with Informer: time complexity and memory usage in each layer, memory bottleneck in receiving long sequence input, and time-consuming in predicting long outputs. The multi-head attention that contains self-attention on Vanilla Transformer changed to ProbSparse self-attention in Informer. The ProbSparse self-attention calculates diversity to catch domination pairs of dot products in the long tail self-attention distribution. The encoder section contains two encoder layers and one Conv1D layer. The decoder section contains one decoder layer; the final layer is a linear transformation. The token, position, and global time stamp embedding is used to build input embedding in Informer architecture [36]. In our experiment, the global time stamp embedding is daily to forecast stock prices the next day.

3) FEDformer

The FEDformer architecture is built to resolve limitations on trend capture in long-term time series forecasting. The FEDformer is a Transformer architecture integrated with a seasonal decomposition mechanism [15]. In Vanilla Transformer, Fourier Enhanced Blocks replace the self-attention layer. Our experiment used two encoders and one encoder to build FEDformer architecture. The encoder contains Fourier Enhanced Blocks, Mixture Of Expert Decomposition Blocks, and Feed-Forward. The decoder contains Fourier Enhanced Blocks, Mixture Of Expert Decomposition Blocks, Frequency Enhanced Attention, and Feed-Forward. The Frequency Enhanced Attention is used to combine output from the encoder part to the input decoder part. The FEDfomer is used to forecasting one-day stock price.

4) Transformer Encoder Gated Recurrent Unit (TEGRU)

The transformer architecture is already used to forecast time series data, such as the ILI dataset. The original transformer architecture consisted of an encoder and decoder section, whereas our Transformer only uses an encoder section to

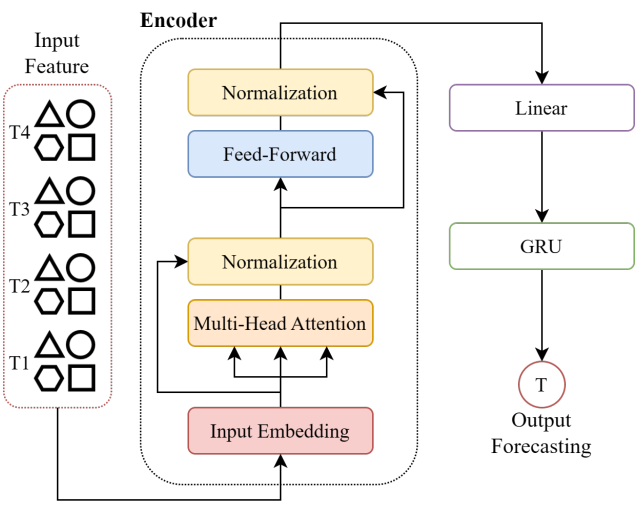


FIGURE 5. Architecture of Transformer Encoder-GRU stock forecasting model.

capture time series data patterns. As shown in Fig. 5, the transformer encoder is compiled by input embedding, a normalization layer, a multi-head attention layer, and a feed-forward layer. Input embedding can be univariate or multivariate with various time lags. The input embedding layer processes the time series features into a vector of dimensions, and then the result passes into the next layer. The Transformer multi-head attention extracts meaningful information from time series data.

$$MultiHead(Q,K,V) = Concat(h_1,\dots,h_H)W^o \quad (7)$$

$$h_a = Attn(QW_a^Q, KW_a^K, VW_a^V) \quad (8)$$

Eq. (7) and (8) show that multi-head attentions consist of some attention heads (h). The query (Q), key (K), and value (V) metrics are produced by multiplying a sequence matrix input embedding with weight matrices W^Q, W^K, and W^V. Two 1D convolution layers stack the feed-forward layer with a ReLU activation function and then pass it into the decoder section. In this research, GlobalAveragePooling1D and the GRU layer utilize separate decoders to compute forecasting stock prices one step ahead. The GRU layer learned the sequence of time series data based on encoder section results that contain self-attention mechanisms.

F. SCENARIO STOCK FORECASTING FEATURE

The proposed architecture stock forecasting compared with other transformer architecture, such as FEDformer, Informer, and Transformer original. Each architecture tested to forecast stock one step ahead uses five feature scenarios and six time-lag variations and evaluates the performance model with four methods. The evaluation methods are R2, RMSE, MSE, and MAPE. The time lag variants are 5, 10, 20, 100, and 200. The first feature scenario is univariate data: the close stock price yesterday to predict the close price tomorrow. The second feature scenario is the technical indicator original, such as open price, highest price, lowest price, close price, and stock transaction volume. The third feature scenario, the multivariate feature, uses section III-C to expand five original technical indicators to produce 43 other technical indicators. The fourth feature scenario uses a combination of close stock price and daily aspect sentiment score to test influencing aspect sentiment on stock price forecasting performance. The fifth feature scenario uses a variety of expanded technical indicators and daily aspect sentiment scores to examine features significantly on stock price forecasting. The architecture with the best evaluation value in each evaluation method and the issuer is determined as the best stock price forecasting architecture.

G. EVALUATION METRICS

Performance evaluation implemented on each model. Aspect categorization and sentiment classification models use a confusion matrix to produce accuracy and an f1-score. The model closing stock price forecasting use mean squared error (MSE), mean absolute percentage error (MAPE) to measure performance prediction, coefficient of determination (R^2), and root mean square error (RMSE) is used to estimate model fit. Previous evaluation methods to measure prediction performance in time series data were not sensitive to trends such as ups and downs. Our proposed method combines the MAPE and misclassification scores to determine the trending class.

$$TS_{true} = [A_t - A_{t-1} \geq 0, 1, 0] \quad (9)$$

$$TS_{pred} = [P_t - A_{t-1} \geq 0, 1, 0] \quad (10)$$

Trend class determined by Eq. (9) and (10) with TS_{true} as ground truth data and TS_{pred} as prediction data. The TS value is produced by actual (A) and predicted (P) values. If the TS value exceeds 0, the trend class is up or 1. Otherwise, if the TS value is less than 0, the trending class is down or 0. After determining trend classes in the ground truth and prediction, accuracy score calculation uses a confusion matrix with variables True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Eq. (11) is used to produce an accuracy score, and Eq. (12) uses to calculate MAPE.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (11)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{A_t - P_t}{A_t}\right| \quad (12)$$

The combined accuracy error score and MAPE equation produced the accuracy mean absolute percentage error (AcMAPE), as shown in Eq. (13). The classification error score will be punished twice times the actual value. Even though MAPE scores less than five percent, if the trending class misses predictions, the model can cause economic loss.

$$AcMAPE = ((1 - Accuracy) \times 2) + (\frac{MAPE}{100}) \quad (13)$$

TABLE 6. Classification performance of the headline news used the entire dataset.

Task	Model	Evaluations			
		Acc	Pre	Rec	f1
combined	cahya/roberta-base-indonesian-522M	0.70	0.66	0.64	0.65
	cahya/distilbert-base-indonesian	0.75	0.71	0.70	0.71
	cahya/bert-base-indonesian-522M	0.74	0.71	0.69	0.70
	indolem/indobert-base-uncased	0.75	0.72	0.70	0.71
	indobenchmark/indobert-base-p2	0.77	0.74	0.73	0.74
aspect	cahya/roberta-base-indonesian-522M	0.86	0.86	0.86	0.86
	cahya/distilbert-base-indonesian	0.88	0.88	0.88	0.88
	cahya/bert-base-indonesian-522M	0.87	0.87	0.87	0.87
	indolem/indobert-base-uncased	0.87	0.88	0.88	0.88
	indobenchmark/indobert-base-p2	**0.88**	**0.89**	**0.89**	**0.89**
sentiments	cahya/roberta-base-indonesian-522M	0.85	0.83	0.81	0.82
	cahya/distilbert-base-indonesian	0.87	0.86	0.84	0.85
	cahya/bert-base-indonesian-522M	0.87	0.85	0.84	0.85
	indolem/indobert-base-uncased	0.87	0.86	0.84	0.85
	indobenchmark/indobert-base-p2	**0.88**	**0.87**	**0.85**	**0.86**

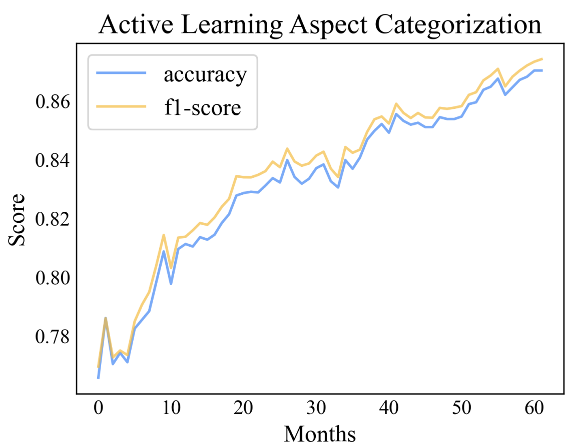
IV. EXPERIMENT RESULT

This section presents the results of active learning on the pre-trained model BERT and discusses forecasting stock prices using a variety of transformer models and various scenario datasets. Classification task performance measures used accuracy and f1-score.

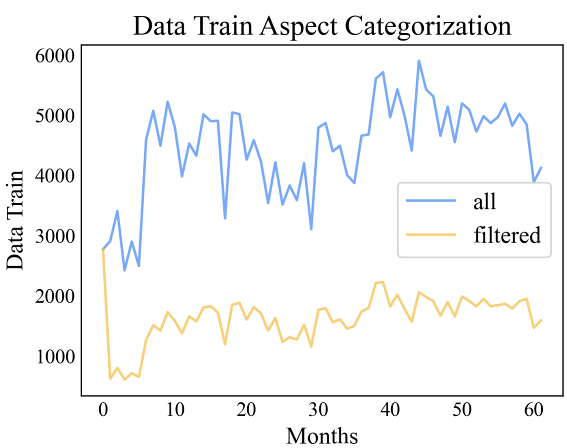
A. ACTIVE LEARNING ASPECT SENTIMENTS

Three scenario classifications are used to determine the best pre-trained model, an aspect-sentiments (combined) classification containing ten classes, aspects categorization with five classes, and a sentiments classification which conceives two classes. As shown in Table 6, the indobenchmark/indobert-base-p2 model is fitted with our dataset. In three scenario classifications, indobenchmark/indobert-base-p2 outperformed other pre-trained models because indobenchmark/indobert-base-p2 had the most extensive corpus and covered the formal Indonesian language. As presented in Table 6, split classification, such as aspect categorization and sentiment classification, are better than aspect-sentiments classification because the split classification model focuses more on small classes than the aspect-sentiment model, which contains larger classes amount.

Active learning builds into resolving two tasks: aspect categorization and sentiment classification using the pre-trained model indobenchmark/indobert-base-p2. Active learning produced 62 models is represent each month between January 2017 and February 2022. As shown in Fig. 6, the active learning model on each task knew to evolve with the newest information. Aspect categorization active learning on the last model had an accuracy of 0.87 and an f1-score of 0.87. Sentiment classification reached a performance accuracy of 0.88 and an f1-score of 0.85 on the last model. The performance of the active learning model is slightly lower than the best-split classification scenario in Table 6. However, as shown in Fig. 7, the active learning model can reduce the average amount of training data by up to 63.4% for aspect categorization and 58.0% for sentiment classification. Minor data usage on active learning indicates that Eq. (5)

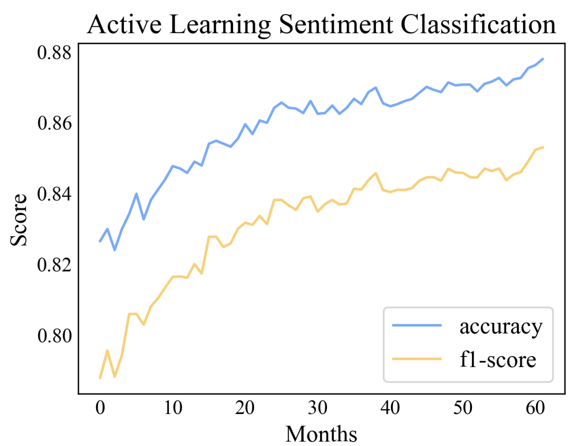


(a)

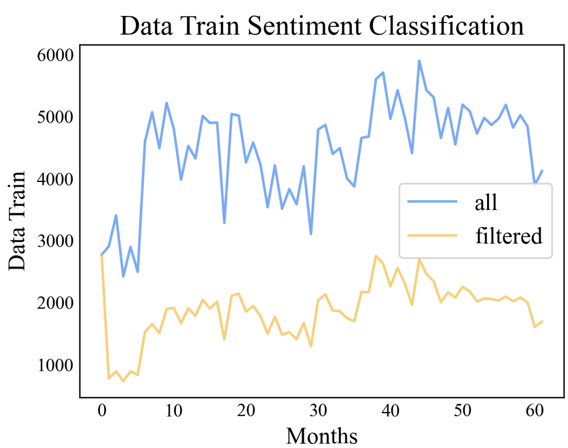


(b)

FIGURE 6. Evaluation performance active learning model on (a) aspect categorization and (b) sentiment classification.



(a)



(b)

FIGURE 7. Dataset training usage on active learning model with task (a) aspect categorization and (b) sentiment classification.

TABLE 7. Stock forecasting performance using the univariate close price dataset.

Model	Metrics	Issuer BBCA	BBRI	TLKM	BMRI
Transformer	RMSE	2516.30	1189.11	200.25	948.86
	MAPE (%)	39.41	29.22	4.56	13.41
Informer	RMSE	2735.99	1331.34	187.41	1159.95
	MAPE (%)	43.31	33.01	3.69	16.16
FEDformer	RMSE	235.01	186.92	124.23	275.03
	MAPE (%)	2.92	3.94	2.64	3.34
T.Encoder	RMSE	171.40	124.24	103.08	216.04
	MAPE (%)	2.07	2.61	2.37	2.70
TEGRU	RMSE	**123.53**	**95.12**	**83.20**	**169.37**
	MAPE (%)	**1.48**	**2.04**	**1.93**	**2.24**

Model	Metrics	Issuer ASII	INDX	SMMT	HITS
Transformer	RMSE	245.65	29.68	28.12	29.02
	MAPE (%)	3.38	33.90	10.26	3.96
Informer	RMSE	240.92	**13.92**	**15.68**	**27.45**
	MAPE (%)	3.09	**14.58**	**5.26**	**3.79**
FEDformer	RMSE	239.38	281.98	106.88	35.39
	MAPE (%)	3.23	337.14	70.42	4.98
T.Encoder	RMSE	221.75	94.04	55.71	32.61
	MAPE (%)	3.04	114.47	22.34	4.24
TEGRU	RMSE	**163.18**	87.71	44.71	29.85
	MAPE (%)	**2.25**	120.71	35.44	3.90

successfully analyzes training data that models do not know without significantly lowering performance.

B. STOCK PRICES FORECASTING

Stock forecasting performance evaluation employed RMSE and MAPE. The comparison performance on each forecasting model operates the average score RMSE and MAPE with different time lags, which are 5, 10, 20, 50, 100, and 200.

TABLE 8. Stock forecasting performance using the multivariate technical indicator dataset.

Model	Metrics	Issuer BBCA	BBRI	TLKM	BMRI
Transformer	RMSE	3043.19	1577.62	466.53	1780.70
	MAPE (%)	48.84	40.72	11.23	27.14
Informer	RMSE	3231.68	1681.73	464.19	1922.52
	MAPE (%)	52.11	43.46	10.89	29.16
FEDformer	RMSE	**263.21**	221.94	168.70	328.09
	MAPE (%)	3.33	4.70	3.84	4.09
T.Encoder	RMSE	468.37	152.44	122.58	352.24
	MAPE (%)	6.01	3.38	2.86	4.50
TEGRU	RMSE	270.57	**120.23**	**109.90**	**191.96**
	MAPE (%)	**3.32**	**2.60**	**2.45**	**2.44**

Model	Metrics	Issuer ASII	INDX	SMMT	HITS
Transformer	RMSE	930.11	1108.01	217.00	37.46
	MAPE (%)	13.94	849.41	120.28	5.73
Informer	RMSE	900.82	1038.64	203.82	36.09
	MAPE (%)	13.35	597.01	86.27	5.58
FEDformer	RMSE	297.36	2294.15	327.72	**36.39**
	MAPE (%)	4.06	953.17	74.76	5.24
T.Encoder	RMSE	418.23	1468.37	213.68	46.63
	MAPE (%)	5.83	543.00	35.34	7.10
TEGRU	RMSE	**230.33**	**161.60**	**64.39**	36.59
	MAPE (%)	**3.17**	**161.85**	**24.95**	**5.16**

TABLE 9. Stock forecasting performance using the multivariate expand technical indicator dataset.

Model	Metrics	Issuer BBCA	BBRI	TLKM	BMRI
Transformer	RMSE	4724.76	2697.95	1808.87	4169.97
	MAPE (%)	78.34	72.97	50.96	67.29
Informer	RMSE	4673.35	2692.07	1824.57	4140.29
	MAPE (%)	77.49	72.88	51.51	66.72
FEDformer	RMSE	**371.07**	**315.64**	228.08	503.37
	MAPE (%)	**4.78**	**6.90**	5.29	6.48
T.Encoder	RMSE	687.40	556.53	217.67	363.87
	MAPE (%)	8.48	11.12	4.83	4.72
TEGRU	RMSE	1.008.29	353.99	**168.73**	**308.54**
	MAPE (%)	13.46	7.66	**3.93**	**4.22**

Model	Metrics	Issuer ASII	INDX	SMMT	HITS
Transformer	RMSE	3470.75	1084.17	492.25	110.16
	MAPE (%)	55.40	1271.40	417.02	20.26
Informer	RMSE	3510.46	760.82	526.37	105.19
	MAPE (%)	56.06	772.84	435.83	18.63
FEDformer	RMSE	578.29	1.472.12	293.06	46.92
	MAPE (%)	8.00	761.50	76.54	7.51
T.Encoder	RMSE	389.35	944.70	275.45	51.90
	MAPE (%)	5.43	462.33	59.61	7.29
TEGRU	RMSE	**318.39**	**304.26**	**90.83**	**43.14**
	MAPE (%)	**4.53**	**168.94**	**34.28**	**6.56**

Various time lags measure model performance consistency in an inconsistent stock environment.

1) Univariate close price (UNI)

Stock forecasting univariate is used one feature as an input model. In our research, close stock prices with different time lags are utilized to forecast the next day's close price. As shown in Table 7, TEGRU, as the proposed model, outper-

TABLE 10. Stock forecasting performance using the multivariate combination of close price and aspect sentiment indicator dataset.

Model	Metrics	Issuer BBCA	BBRI	TLKM	BMRI
Transformer	RMSE	2132.28	905.29	472.01	817.69
	MAPE (%)	29.28	21.40	10.12	9.05
Informer	RMSE	2097.42	890.38	454.37	803.36
	MAPE (%)	28.91	21.12	9.98	8.98
FEDformer	RMSE	328.94	219.82	209.38	345.42
	MAPE (%)	3.82	4.51	4.54	4.24
T.Encoder	RMSE	316.68	175.78	173.80	286.81
	MAPE (%)	3.64	3.43	3.56	3.46
TEGRU	RMSE	**237.48**	**125.81**	**104.97**	**167.52**
	MAPE (%)	**2.76**	**2.48**	**2.22**	**2.12**

Model	Metrics	Issuer ASII	INDX	SMMT	HITS
Transformer	RMSE	1032.11	60.08	168.15	264.06
	MAPE (%)	18.55	27.52	35.29	66.46
Informer	RMSE	1014.86	59.39	166.93	260.15
	MAPE (%)	18.27	27.22	35.02	65.48
FEDformer	RMSE	311.90	**33.77**	**89.20**	38.14
	MAPE (%)	4.69	15.43	14.39	7.55
T.Encoder	RMSE	304.73	53.41	185.31	49.38
	MAPE (%)	4.42	15.00	17.60	8.16
TEGRU	RMSE	**184.13**	47.32	153.41	**35.58**
	MAPE (%)	**2.72**	**11.57**	**11.99**	**5.94**

TABLE 11. Stock forecasting performance using the multivariate combination of aspect sentiment and technical indicator dataset.

Model	Metrics	Issuer			
		BBCA	**BBRI**	**TLKM**	**BMRI**
Transformer	RMSE	2203.35	933.57	467.72	787.68
	MAPE (%)	30.26	22.10	10.21	8.92
Informer	RMSE	2207.67	929.29	449.79	767.74
	MAPE (%)	30.55	21.98	9.92	8.60
FEDformer	RMSE	**332.26**	227.27	207.99	328.65
	MAPE (%)	**3.89**	4.64	4.59	4.12
T.Encoder	RMSE	538.33	182.40	165.13	344.67
	MAPE (%)	538.33	182.40	165.13	344.67
TEGRU	RMSE	481.24	**135.62**	**158.52**	**245.29**
	MAPE (%)	5.65	**2.68**	**3.08**	**3.32**

Model	Metrics	Issuer			
		ASII	**INDX**	**SMMT**	**HITS**
Transformer	RMSE	1169.66	62.37	167.46	269.41
	MAPE (%)	21.29	29.99	35.25	67.82
Informer	RMSE	1129.86	57.82	165.12	263.09
	MAPE (%)	20.63	26.34	35.55	66.24
FEDformer	RMSE	299.47	**33.63**	**79.56**	**34.81**
	MAPE (%)	4.51	**14.89**	**13.50**	**6.73**
T.Encoder	RMSE	360.91	74.34	168.10	45.88
	MAPE (%)	5.20	28.32	17.41	7.91
TEGRU	RMSE	**215.37**	85.38	204.73	46.08
	MAPE (%)	**3.15**	33.93	16.40	8.61

formed the transformer encoder, full Transformer, Informer, and FEDformer on 5 of 8 stock issuers, especially on large market stock capitalization. The BBCA stock issuer had the lowest MAPE score with an RMSE score of 123.53, lower 27.93% than the transformer encoder, and 47.44% than the FEDformer architecture.

2) Multivariate original technical indicator (OTI)

Open price, lowest price, highest price, close price, and volume of transactions on previous days are used to forecast the closing price on the next day. As shown in Table 8, the proposed model TEGRU outperformed the baseline model on MAPE score, even though lousy performance score on INDX and SMMT issuer. The BMRI stock issuer had the lowest MAPE score with an RMSE score of 191.96, lower by 41.49% than the FEDformer architecture. Attaching the GRU layer as a decoder on transformer architecture boosts the performance of the transformer encoder on all stock issuers.

3) Multivariate expand technical indicator (ETI)

As mentioned in section III-C, an expanded technical indicator is used to stock forecasting the next day. Eq. (6) is utilized to determine the fitted feature of each stock issuer so that each stock issuer has a different number of features. As shown in Table 9, the proposed model TEGRU performs better than another transformer architecture, especially on 6 of 8 stock issuers. The TLKM stock issuer had the lowest MAPE score with an RMSE score of 168.73, lower by 26.02% than the FEDformer architecture. INDX and SMMT issuers still not yet gain better results.

4) Multivariate close price and aspect sentiment indicator (CAS)

The features combined one technical indicator, the close price, and 35 news sentiment indicators to forecast stock price in a one-time step. Appropriate feature on each issuer produced by process removed feature with high multi-collinearity with Eq. (6). Data usage ranged between January 2017 and March 2022 because of date range limitations on news sentiment. As shown in Table 10, the proposed model TEGRU outperformed the previous model on MAPE score. The BMRI stock issuer had the lowest MAPE score with an RMSE score of 167.52, lower by 51.50% than the FEDformer architecture. SMMT and INDX issuers gained better performance than only using the technical indicator on stock forecasting.

5) Multivariate expand technical and aspect sentiment indicator (ETS)

Concatenate expanded technical indicator and news sentiment produced 78 features to forecasting close stock price in one day. Eq. (6) uses to make features with lower multicollinearity. The proposed model TEGRU and FEDformer equally outperformed on 4 of 8 stock issuers, as shown in Table 11. TEGRU best fitted on BBRI, TLKM, BMRI, and ASII, whereas FEDformer best fitted on BBCA, INDX, SMMT, and HITS. The best performance TEGRU gained by BBRI issuer with a 2.68 MAPE score, whereas BBCA issuer is the best FEDformer performance with a 3.89 MAPE score.

V. DISCUSSION

Based on evaluating the stock forecasting model in subsections IV-B, TEGRU outperforms other Transformer architectures. Based on Table 7 to 11, the GRU layer success boosts the performance of the Transformer Encoder, which means the GRU layer better captures the pattern of sequence data than other Transformer decoders.

The influence of news sentiment on stock price forecasting depends on each issuer. As shown in Fig. 8, the average TEGRU performance evaluation uses eight stock issuers in five scenarios. The CAS feature scenario produces the best performance with a MAPE score is 5.22. Comparing UNI and CAS feature scenario, news sentiment significantly influence INDX and SMMT issuers. Issuers with the highest market capitalization are less affected by sentiment features. The AcMAPE score comparison confirms that news sentiment influences stock price forecasting. The CAS feature scenario yield outperformed other feature scenarios with an AcMAPE score is 1.00, shown in Fig 9. The CAS feature scenario had less error than the UNI feature scenario, with a deviation is 0.23. Features without news sentiment, such as UNI, OTI, and ETI, yield more errors than features with news sentiment. The CAS feature scenario with the lowest MAPE and AcMAPE score had a high risk to investors on trading because the CAS feature scenario had an accuracy score is 0.52. That indicates the investor had a chance of a 52% right decision on the stock transaction, which means the CAS

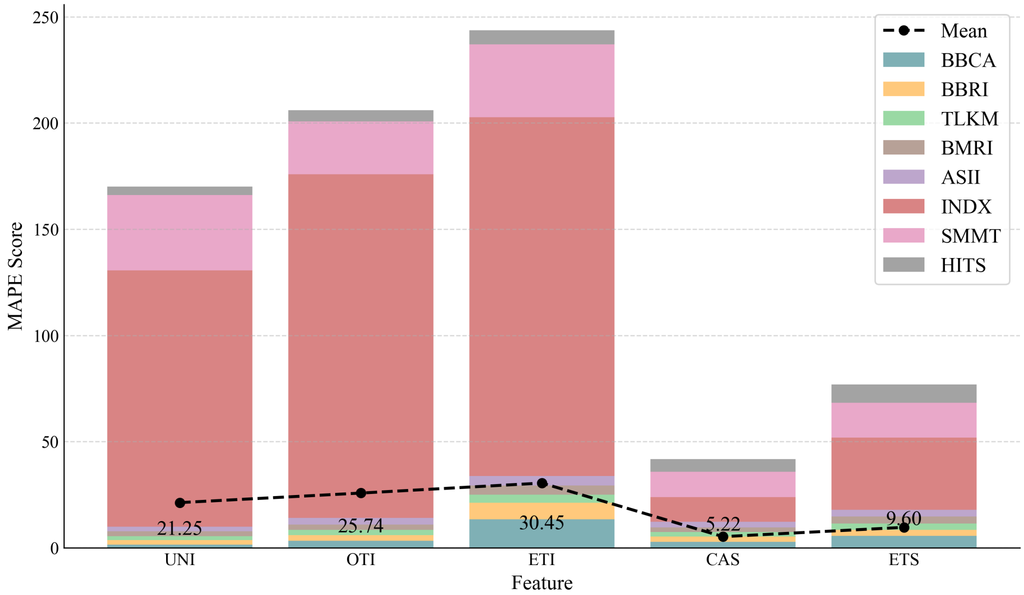


FIGURE 8. Feature scenario performance comparison using MAPE evaluation.

feature scenario cannot be used as an absolute scenario for all issuers.

Table 12 presents the best-fit parameter of each issuer with TEGRU architecture as a forecasting model. The criterion of the best-fit parameter had the best evaluation value of at least 3 of 5 evaluation methods. The lookback parameter used for stock forecasting varies significantly in the selected parameter scenario, indicates each issuer requires a fitted parameter to produce the best model performance. Based on the feature type, univariate features outperformed multivariate features, especially on evaluation R2, RMSE, MSE, and MAPE. Although the four methods evaluation performed excellently, AcMAPE showed that the forecasting result did not have good trend prediction accuracy. For example, a BBCA issuer with a lookback parameter is 200, and a univariate feature type had the highest MAPE score of 1.19%, but the AcMAPE score reached 1.02, which means the accuracy score is 0.50. The investor or trader had a risk in buying or selling stock because the model had a 50% performance to predict a correct stock trend. The failed stock trend prediction caused

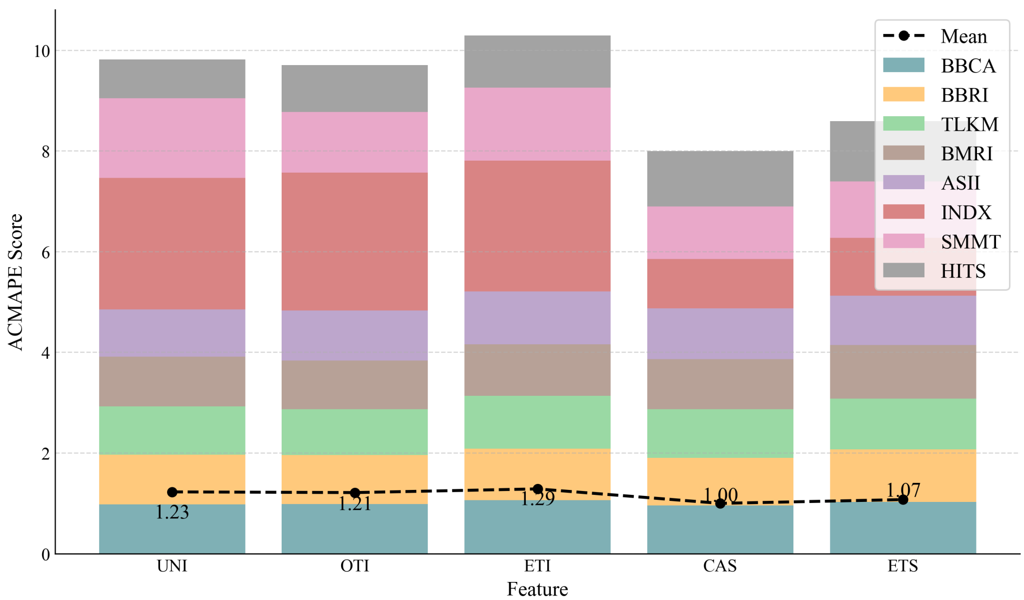


FIGURE 9. Feature scenario performance comparison using AcMAPE evaluation.

TABLE 12. TEGRU model best-fit parameter on each issuer for stock forecasting.

Issuer	Lookback	Type	Features	Evaluation				
				R2	RMSE	MSE	MAPE (%)	AcMAPE
BBCA	200	Univariate	1	0.99	99.26	9,852.55	1.19	1.02
	50	OTI	5	0.96	175.15	30,679.09	2.00	0.99
BBRI	50	Univariate	1	0.97	87.36	7,631.69	1.81	0.97
	50	OTI	5	0.97	100.17	10,033.85	2.04	0.94
TLKM	5	Univariate	1	0.97	75.19	5,652.93	1.73	0.93
	50	CAS	13	0.97	76.93	5,918.06	1.61	0.94
BMRI	10	Univariate	1	0.96	147.87	21,864.08	1.88	1.02
	10	OTI	5	0.97	143.09	20,475.32	1.86	0.93
ASII	50	Univariate	1	0.98	144.81	20,970.20	1.94	0.87
	5	CAS	12	0.87	165.52	27,396.96	2.43	0.91
INDX	100	Univariate	1	0.46	36.24	1,313.56	47.58	1.88
	10	CAS	13	0.59	44.34	1,966.05	10.34	1.15
SMMT	20	Univariate	1	0.97	26.51	702.85	18.29	1.41
	50	OTI	5	0.93	39.53	1,563.02	12.01	1.06
HITS	10	Univariate	1	0.93	28.59	817.60	3.48	0.92
	100	OTI	5	0.92	30.85	951.46	4.28	1.11

a financial loss in the MAE score.

The stock issuer, TLKM, ASII, and INDX uses aspect sentiment indicator to produce the best forecasting performance in multivariate feature. Fig. 10 shows the feature forecasting correlation in close price and aspect sentiment indicator. All correlation values were under 0.9, which means the features had a weak correlation. Industry, international, investment, finance, and national sentiment can influence stock issuers. The positive and negative sentiment with a mostly minimum daily score is selected to increase performance stock forecasting.

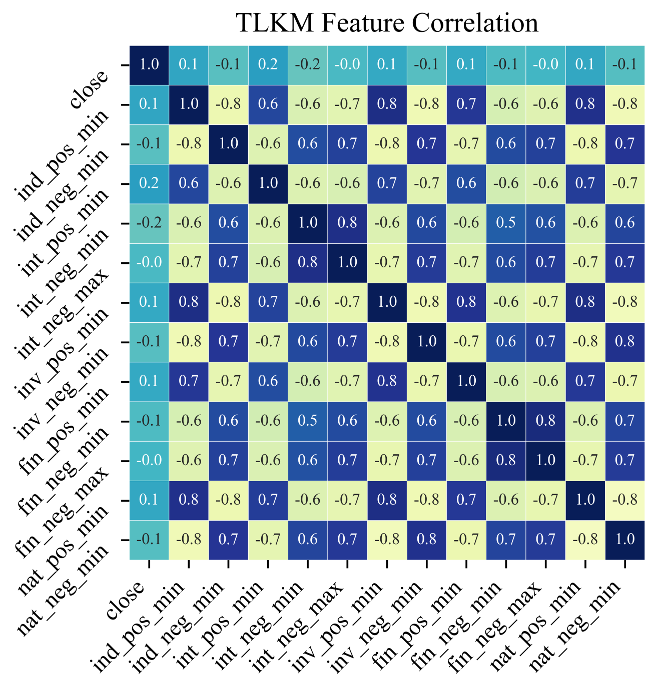


FIGURE 10. The feature forecasting correlation of TLKM stock issuer.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we presented quantifying multi-aspect news sentiment with considered time sequence data and then forecasting stock prices next days using sentiment and technical indicators. In feature extraction, a daily confidence score from the classification model is used to quantify sentiment. The active learning model produced a monthly classification model to represent model knowledge. The Transformer Encoder Gated Recurrent Unit (TEGRU) outperformed another Transformer architecture for stock price forecasting on eight stock issuers. The TEGRU model consists of the transformer encoder to learn time series pattern data with multi-head attention and passed into the GRU layer to determine stock price. The TEGRU used various parameters to produce the best model performance, and multi-aspect sentiment indicators affected the performance forecasting model top gains issuers. The Accuracy Mean Absolute Percentage Error (AcMAPE) can describe the risk of misclassification trends on stock price forecasting that implicated financial loss.

This study still has a flaw that can be addressed. Our future research selects news data that highly correlate with stock forecasting movement on each stock issuer and reduces active learning from monthly to weekly. In addition, use the AcMAPE score as a loss function to create a stock price forecasting model.

ACKNOWLEDGMENT

This research was funded by Asian Development Bank under HETI ADB Project; The Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program; Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Dana Departemen Program; and Institut Teknologi Sepuluh Nopember (ITS) under project scheme of the Publication Writing and IPR Incentive Program (PPHKI).

REFERENCES

[1] L. Nemes and A. Kiss, "Prediction of stock values changes using sentiment analysis of stock news headlines," *Journal of Information and*

Telecommunication, vol. 5, no. 3, pp. 375–394, 2021. [Online]. Available: https://doi.org/10.1080/24751839.2021.1874252

[2] Y. J. Chen, Y. M. Chen, and C. L. Lu, "Enhancement of stock market forecasting using an improved fundamental analysis-based approach," *Soft Computing*, vol. 21, no. 13, pp. 3735–3757, 2017. [Online]. Available: https://doi.org/10.1007/s00500-016-2028-y

[3] P. Mehta, S. Pandya, and K. Kotecha, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," *PeerJ Computer Science*, vol. 7, pp. 1–21, 2021. [Online]. Available: https://doi.org/10.7717/peerj-cs.476

[4] P. Y. Hao, C. F. Kung, C. Y. Chang, and J. B. Ou, "Predicting stock price trends based on financial news articles and using a novel twin support vector machine with fuzzy hyperplane," *Applied Soft Computing*, vol. 98, p. 106806, 2021. [Online]. Available: https://doi.org/10.1016/j.asoc.2020.106806

[5] A. T. Haryono, R. Sarno, and R. Abdullah, "Aspect-Based Sentiment Analysis of Financial Headlines and Microblogs Using Semantic Similarity and Bidirectional Long Short-Term Memory," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 3, pp. 233–241, 2022. [Online]. Available: https://inass.org/wp-content/uploads/2021/12/2022063020-2.pdf

[6] S. Budd, E. C. Robinson, and B. Kainz, "Survey paper A survey on active learning and human-in-the-loop deep learning for medical image analysis," *Medical Image Analysis*, vol. 71, p. 102062, 2021. [Online]. Available: https://doi.org/10.1016/j.media.2021.102062

[7] X. Du, L. Deng, and K. Qian, "Current market top business scopes trend—a concurrent text and time series active learning study of nasdaq and nyse stocks from 2012 to 2017," *Applied Sciences*, vol. 8, no. 5, 2018. [Online]. Available: https://doi.org/10.3390/app8050751

[8] J. Zhang and Q. Dai, "A cost-sensitive active learning algorithm: toward imbalanced time series forecasting," *Neural Computing and Applications*, vol. 34, no. 9, pp. 6953–6972, 2022. [Online]. Available: https://doi.org/10.1007/s00521-021-06837-3

[9] P. F. Pai and C. S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005. [Online]. Available: https://doi.org/10.1016/j.omega.2004.07.024

[10] X. Yuan, J. Yuan, T. Jiang, and Q. U. Ain, "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market," *IEEE Access*, vol. 8, pp. 22 672–22 685, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2969293

[11] Y. Gao, R. Wang, and E. Zhou, "Stock Prediction Based on Optimized LSTM and GRU Models," *Scientific Programming*, vol. 2021, pp. 1–8, sep 2021. [Online]. Available: https://doi.org/10.1155/2021/4055281

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 2017-Decem, no. Nips, 2017, pp. 5999–6009. [Online]. Available: https://doi.org/10.48550/arXiv.1706.03762

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, no. Mlm. Stroudsburg, PA, USA: Association for Computational Linguistics, oct 2019, pp. 4171–4186. [Online]. Available: http://dx.doi.org/10.18653/v1/N19-1423

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2010.11929

[15] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting," in *Proceedings of the 39th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, Eds., vol. 162. PMLR, 2022, pp. 27 268–27 286. [Online]. Available: https://doi.org/10.48550/arXiv.2201.12740

[16] J. Chen, T. Chen, M. Shen, Y. Shi, D. Wang, and X. Zhang, "Gated three-tower transformer for text-driven stock market prediction," *Multimedia Tools and Applications*, vol. 81, no. 21, pp. 30 093–30 119, 2022. [Online]. Available: https://doi.org/10.1007/s11042-022-11908-1

[17] S. Hansun and J. C. Young, "Predicting LQ45 financial sector indices using RNN-LSTM," *Journal of Big Data*, vol. 8, no. 1, 2021. [Online]. Available: https://doi.org/10.1186/s40537-021-00495-x

[18] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9713–9729, 2020. [Online]. Available: https://doi.org/10.1007/s00521-019-04504-2

[19] C. Li and G. Qian, "Stock Price Prediction Using a Frequency Decomposition Based GRU Transformer Neural Network," *Applied Sciences (Switzerland)*, vol. 13, no. 1, 2023. [Online]. Available: https://doi.org/10.3390/app13010222

[20] M. D. Pratama, R. Sarno, and R. Abdullah, "Sentiment Analysis User Regarding Hotel Reviews by Aspect Based Using Latent Dirichlet Allocation, Semantic Similarity, and Support Vector Machine Method," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 3, pp. 514–524, 2022. [Online]. Available: https://inass.org/wp-content/uploads/2021/12/2022063043-1.pdf

[21] B. Joyce and J. Deng, "Sentiment analysis of tweets for the 2016 US presidential election," *2017 IEEE MIT Undergraduate Research Technology Conference, URTC 2017*, vol. 2018-Janua, pp. 1–4, 2018. [Online]. Available: https://doi.org/10.1109/URTC.2017.8284176

[22] W. Jin, B. Zhao, L. Zhang, C. Liu, and H. Yu, "Back to common sense: Oxford dictionary descriptive knowledge augmentation for aspect-based sentiment analysis," *Information Processing and Management*, vol. 60, no. 3, p. 103260, 2023. [Online]. Available: https://doi.org/10.1016/j.ipm.2022.103260

[23] I. Ghozali, K. R. Sungkono, R. Sarno, and R. Abdullah, "Synonym based feature expansion for Indonesian hate speech detection," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 1, p. 1105, 2023. [Online]. Available: https://doi.org/10.11591/ijece.v13i1.pp1105-1112

[24] C. Qian, N. Mathur, N. H. Zakaria, R. Arora, V. Gupta, and M. Ali, "Understanding public opinions on social media for financial sentiment analysis using AI-based techniques," *Information Processing and Management*, vol. 59, no. 6, p. 103098, 2022. [Online]. Available: https://doi.org/10.1016/j.ipm.2022.103098

[25] C. Xiao, W. Xia, and J. Jiang, "Stock price forecast based on combined model of ARI-MA-LS-SVM," *Neural Computing and Applications*, vol. 32, no. 10, pp. 5379–5388, 2020. [Online]. Available: https://doi.org/10.1007/s00521-019-04698-5

[26] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Computing and Applications*, vol. 33, no. 10, pp. 4741–4753, 2020. [Online]. Available: https://doi.org/10.1007/s00521-020-05532-z

[27] H. Liu and Z. Long, "An improved deep learning model for predicting stock market price time series," *Digital Signal Processing: A Review Journal*, vol. 102, p. 102741, 2020. [Online]. Available: https://doi.org/10.1016/j.dsp.2020.102741

[28] E. Teti, M. Dallocchio, and A. Aniasi, "The relationship between twitter and stock prices. Evidence from the US technology industry," *Technological Forecasting and Social Change*, vol. 149, no. November, p. 119747, 2019. [Online]. Available: https://doi.org/10.1016/j.techfore.2019.119747

[29] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Systems with Applications*, vol. 178, no. March, p. 115019, 2021. [Online]. Available: https://doi.org/10.1016/j.eswa.2021.115019

[30] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv*, no. 1, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1907.11692

[31] J. Hartmann, M. Heitmann, C. Siebert, and C. Schamp, "More than a Feeling: Accuracy and Application of Sentiment Analysis," *International Journal of Research in Marketing*, no. xxxx, pp. 1–13, 2022. [Online]. Available: https://doi.org/10.1016/j.ijresmar.2022.05.005

[32] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. Suzhou, China: Association for Computational Linguistics, Dec. 2020, pp. 843–857. [Online]. Available: https://aclanthology.org/2020.aacl-main.85

[33] J. Zhao, D. Zeng, S. Liang, H. Kang, and Q. Liu, "Prediction model for stock price trend based on recurrent neural network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 745–753, 2021. [Online]. Available: https://doi.org/10.1007/s12652-020-02057-0

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3298445

[34] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, no. September, pp. 111–132, 2022. [Online]. Available: https://doi.org/10.1016/j.aiopen.2022.10.001

[35] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case," *CoRR*, vol. abs/2001.0, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2001.08317

[36] 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," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, may 2021, pp. 11 106–11 115. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/17325

[37] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 757–770. [Online]. Available: https://aclanthology.org/2020.coling-main.66

[38] D. L. Padial, "Technical Analysis Library in Python," 2022. [Online]. Available: https://github.com/bukosabino/ta

[39] D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, "A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables," *Applied Energy*, vol. 140, pp. 385–394, 2015. [Online]. Available: https://doi.org/10.1016/j.apenergy.2014.12.011

AGUS TRI HARYONO was born in Balikpapan, Kalimantan Timur, Indonesia in 1988. He received his Bachelor's degree in computer science from Mulawarman University, Samarinda, Indonesia, in 2012. He is pursuing a Master's degree in informatics at the Institut Teknologi Sepuluh Nopember (ITS), Indonesia. His research interests include financial time series, text mining, and machine learning.

RIYANARTO SARNO (Senior Member, IEEE) received the Ph.D. degree, in 1992. He is a Professor at the Informatics Department, Institut Teknologi Sepuluh Nopember (ITS). He was the author of more than five books and over 300 scientific articles, incorporating him in the top 2% world ranking scientist by Standford University in 2020. He has researched process mining for five years. His research interests include machine learning, the Internet of Things, knowledge engineering, enterprise computing, and information management.

KELLY ROSSA SUNGKONO was born in Surabaya, in June 1994. She received a master's degree in computer engineering in 2016. She is a Lecturer at the Institut Teknologi Sepuluh Nopember (ITS). Her research interests include databases, information systems, and machine learning.

• • •