International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Application of Ensemble Transformer-RNNs on Stock Price Prediction of Bank Central Asia

Muhammad Rizki Nur Majiid *[1], Renaldy Fredyan [2], Gede Putra Kusuma [3]

Submitted: 10/11/2022 **Accepted**: 12/02/2023

Abstract: Breaking news information about the stock market is gathered from numerous finance websites. Internet portals offer free financial information about businesses. The impact of significant financial reforms, the environment, natural disasters, and news events on the stock market is minimal. The online financial platform generates many time series data. Additionally, we include data from the USD, CNY, Gold, and Oil unrelated to the stock share but relevant. With the use of various machine learning algorithms, market reforms are projected. Those datasets are collected from yahoo finance and investing.com, along with other stock market aspects. These models are trained to utilize an extended dataset, including open price, close price, low price, high price, and volume, from Bank Central Asia's stock price. The Ensemble Transformer LSTM (ET-LSTM) and Ensemble Transformer GRU (ET-GRU) architecture forecast the stock price for the following day. The data set is improved using a variety of deep learning approaches to get more accurate findings. Both suggested approaches use ensemble architecture to deliver 9% of MAPE. Market movements are perfectly aligned in terms of high and low stock prices. Algorithms for high-frequency trading can further enhance the outcomes.

Keywords: Stock Prediction, Deep Learning, Transformer Model, Multi-Head Attention Mechanism, Ensemble Model

1. Introduction

Along with advances in technology, it is easier for the general public to be moved to invest as early as possible. One of the popular investments made is stock trading. The role of artificial intelligence is needed in the decision-making process to make it easier for ordinary people to trade stocks [1]. Because it can be challenging to know when and how to allocate budgets, investors have utilized technical and quantitative approaches to estimate asset price movement. These techniques involve seeing a valuable pattern in market data and figuring out when to invest.

Companies may also benefit from stock price predictions in addition to stock traders [2]. Making forecasts about stock prices can therefore serve as an indirect predictor of the firm's future health since stock prices can serve as a proxy for the state of the company. A company's decision-makers can prepare for upcoming business issues by forecasting stock movements. The financial news offers crucial details to buy and sell a specific company's shares [3]. Investors and traders make decisions based on the dynamic shift in news about a company. This study modifies the deep learning architecture to anticipate stock market closing values. The data is gathered from reputable sources like Google News and Yahoo Finance [4]. Budget announcements, demonetization, as well as unexpected occurrences like terrorist attacks and natural catastrophes, have a significant impact on the stock market. It's quite uncommon to predict stock prices with the requisite 100% accuracy.

For predicting the stock market, linear models like AR, ARMA, and ARIMA have been utilized [5]. These models are only effective for a specific time series of data; hence they are not effective for other time series of data. Stock market forecasting involves a greater degree of risk than other industries because of the stock market's ambiguous and unpredictable character. It is among the most significant causes of the difficulty in making accurate stock market predictions. This is where deep learning models for financial forecasting come into play [6]. When given a number or quantity that is nearly but not quite accurate, deep learning models can learn from their mistakes and become more efficient. The following traits make the practical application of deep learning to forecasting problems particularly successful: Even if the data set is exceedingly complex, they can still investigate the relationship between input and output. Even if new test samples were not used during network training, they can nevertheless recognize them.

The ability to anticipate the price of stocks based on their previous day's price is not well studied. The proposed algorithm significantly outperforms prediction-based techniques based on supervised learning theories like logistic regression, Gaussian Discriminant Analysis (GDA), quadratic discriminant analysis (QDA), and SVM [7], according to an experimental study using real-time data from the IDX Stock Exchange. On the plus side, the rarity of studies lends this one an air of novelty and gives other researchers a chance to fill in the blanks[8].

1 Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jl. Raya Kebon Jeruk No. 27, Jakarta Barat, 11530, INDONESIA
ORCID ID : 0000-0002-0198-510X
2 Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jl. Raya Kebon Jeruk No. 27, Jakarta Barat, 11530, INDONESIA
ORCID ID : 0000-0002-4722-6408
3 Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jl. Raya Kebon Jeruk No. 27, Jakarta Barat, 11530, INDONESIA
ORCID ID : 0000-0003-4241-997X
** Corresponding Author Email: muhammad.majiid@binus.ac.id*

The performance of prediction algorithm strategies is impacted by dataset properties [9]. This paper is organized into five sections. Section 2 is related works of previous research that has been done by using machine learning algorithm. Section 3, focus on develop proposed method, dataset description, technical indicators, and feature extraction. Section 4 explain about experiment results and analysis model. Lastly, conclude and future work present the paper in section 5.

2. Related Works

The use of neural networks in stock market prediction has been the subject of extensive research in the past. In order to study the momentum impact, Takeuchi & Lee made predictions about which stocks would have better or lower monthly returns. The feed-forward neural network classifier uses an auto-encoder with stacked RBMs to extract features from stock prices [10]. The number of hidden units is drastically decreased to provide dimensionality reduction in the encoder's final layer. Before training, the dataset is split up into smaller mini bunches. The RBMs are unrolled using them to create an encoder-decoder [11]. The back propagation algorithm supports it.

According to Feiyu Wang research, some S&P 500 equities will outperform the market on any given day. The design consists of an MLP and a 3-layer Deep Belief Network [12]. Back-propagation is used to pre-train and fine-tune the DBN module. The Z-score for each time period is used to normalize the characteristics. The number of neurons, layers, and regularization parameters are all determined using the validation data. The model outperforms regularized logistic regression and MLP. A deep learning model is trained by Liang et al. to forecast German stock returns [13]. A parametric forecasting model for Value-at-Risk or VaR based on the normal inverse Gaussian distribution (Hereinafter NIG-DCS-VaR), which daily VaR forecast can be creative incorporates intraday information [14].

Sharang and Rao use a Deep Belief Network (DBN) made up of two stacked RBMs to trade a portfolio of US Treasury note futures. Three separate classifiers—regularized logistic regression, support vector machines, and a neural network with two hidden layers—use the DBN's hidden characteristics as input. An approach called contrastive divergence is used to train DBN. Using PCA, the portfolio is built to be unaffected by the first principal component. In comparison to a random predictor, the results are 5–10% more accurate. To forecast daily S&P 500 movements, structured information collected from headlines is used [15]. Using Open IE, headlines are processed to produce structured event representations of actor, action, object, and time. Short-term and long-term consequences of events are combined using CNN (Convolutional Neural Network). Results of organized events produce more accurate stock market prediction features than words [16].

For 43 commodities and currency futures, Price-based classification models employ a deep neural network to forecast the direction of the price change over the next 5 minutes [17]. They employ traditional backpropagation with stochastic gradient descent to represent lagged price differences and co-movements between contracts with 9896 neurons in their input layer. Overall accuracy for three-class classification is 42%. The oscillation box hypothesis predicts that a stock's price will change over the course of a specific series. Oscillation box theory is used by DBNs to make trade decisions [18]. When a stock price leaves a range, it goes into a new box. Stacked RBMs plus a final back-propagation layer combine to create a DBM. Each layer is unsupervised trained from low to high using Block Gibbs sampling. They then refine the entire model by supervising the training of the back-propagation layer [19].

Using news stories, text-based classification models forecast bank trouble. To minimize dimensionality, semantic pre-training is done using neural networks. The classification function of the second neural network. According to feed-forward topology, once the learning of connection weights is finished, the middle projection layer supplies the semantic vectors. The input layer for banks is made up of 716 thousand sentences that were taken from Reuters News articles that were written during and after the crisis. An individual Usefulness metric provides a classification model [20].

The open, high, low, and close prices are used to measure the daily volatility of the S&P 500. A single LSTM (Long Short-Term Memory) block makes up a single LSTM hidden layer. Using mean absolute percent error (MAPE) as the objective loss function, the "Adam" approach is applied to batches of 32 samples. The LSTM approach performs better than the GARCH [21], Ridge, and LASSO approaches [22]. Big Data refers to technological advancements that enable the storage of significant amounts of both structured and unstructured data. The country's economy is reliant on a wide range of sales and financial outcomes [23]. Another research has been conducted based on human attention mechanism including the dual-stage two-phase (DSTP) model and the influence mechanism of target information and non-target information, the experimental results demonstrate that the present work can be successfully used to develop expert or intelligent systems for a wide range of applications [24].

Several recent studies have used neural network methods to predict stock prices. New research, such as that conducted by Lu [25] and Kamara [26], utilizes a combination of Convolutional Neural Networks (CNN) with Recurrent Neural Networks (RNN), and Attention Mechanism. However, these two pieces of research are still focused on model development. They have not yet reached the development of data varieties. Therefore, it is essential to adopt Dai's innovation [27] by adding features of external commodity price movements such as exchange rates, gold prices, and crude oil prices. So, it is necessary to develop CNN, RNN, and Attention Mechanism models using a combined dataset between stock prices and other related commodities.

Some models fared well in terms of having strong prediction accuracy in financial time series data. There is still room for development, though. First, because the Deep Neural Network (DNN) model is fully connected, one of its main flaws is that it is unable to identify any underlying patterns in the feature space. Second, the training phase of DNN encounters the disappearing and exploding gradient problem. In this study, we investigate whether stock price time series can contain traits that can be used to accurately forecast future returns. Even skilled investors find it challenging to anticipate stock returns using publicly accessible data because of the large level of noise in stock price movements, even though deep learning research takes tasks that are simple for humans to complete into account. Any patterns that do exist are additionally susceptible to alteration as traders strive for profits and gain experience over time.

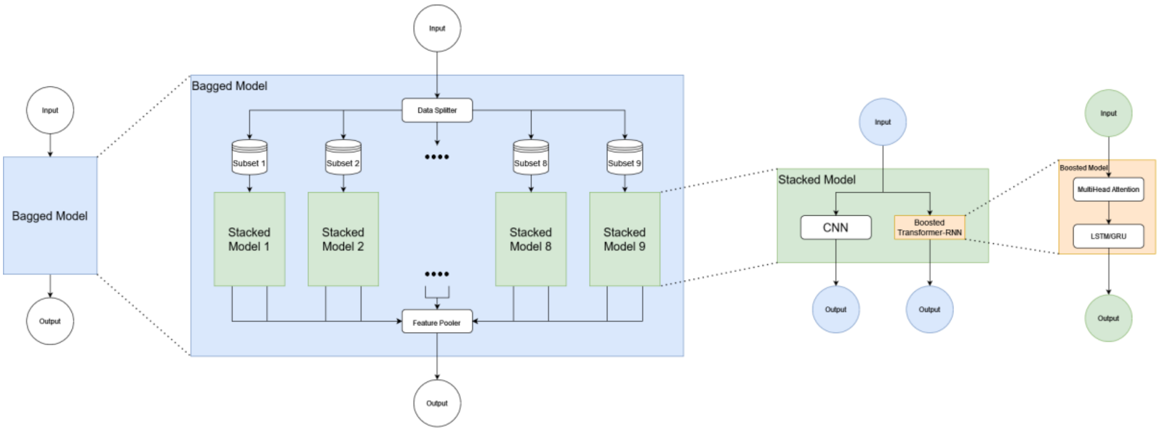

Fig. 1 The architecture of ET-RNN combined with CNN using ensemble architecture.

3. Method

3.1. Ensemble Learning

Ensemble learning is a machine learning method that combines multiple models to produce better predictive results. Three ensemble methods will be used in this research: Bagging, Stacking, and Boosting.

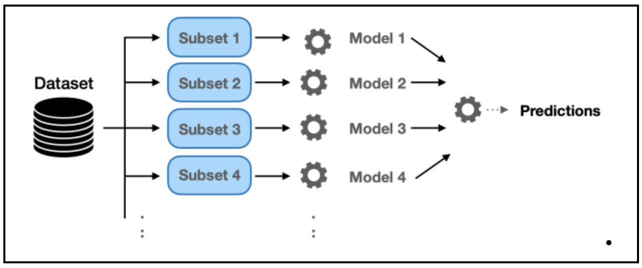

Fig. 2. Bagging Ensemble by combining multiple models to multiple subsets of training dataset to make final predictions [28].

The concept of bagging is that each model learns from error using different subsets of the training dataset, as shown in Fig. 2. In this case, the subset represents each feature of the dataset. Because the data processed is time series, splitting the data based on the index is not recommended.

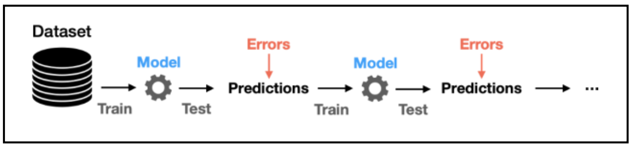

Fig. 3. Boosting Ensemble is used to produce stronger learned model [28].

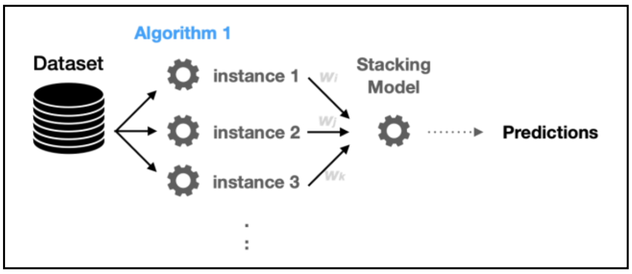

Fig. 4. Stacking Ensemble trains several models using the same data [28].

Boosting is an ensemble method by sequentially building a model on top of other models, as shown in Fig. 3. This method relies on the model's ability to process the previous model's prediction results, resulting in better prediction results. This technique has often proven successful in classification cases using machine learning.

Stacking is like a combination of Boosting and Bagging. This architecture trains multiple models at once using the same data then aggregates the prediction results of each model and is processed by a model before generating the final prediction. The visualization of this method can be seen in Fig. 4.

3.2. Proposed Model

This study attempts to propose a state-of-the-art model for sequential data, Transformers, and arrange them in such a way using an ensemble architecture. The whole Ensemble Transformer-RNN (ET-RNN) model is illustrated in Fig. 1.

The central part of this architecture is the Multihead Attention layer arranged in series with the Recurrent Neural Network (RNN) layer. The Multihead Attention layer gives weight to each data point in the sequence. The sequences given the weight are processed by the RNN layer, a particular unit for processing sequential data. In this case, the RNN unit used is either LSTM or GRU. This choice is made because the two models are developments from ordinary RNNs. From now on, the combination of these two layers will be referred to as the Transformer-RNN.

The Transformer-RNN unit is then arranged in parallel with the Convolutional Neural Network (CNN). The choice of CNN as a stacking pair of Transformer-RNN is based on its success in the image data processing. Because the data used is sequential, the CNN unit used is Convolution 1D, a CNN unit designed to process sequential data.

The stacking model of Transformer-RNN and CNN was replicated nine times (according to the number of features from the dataset) and arranged in parallel following the Bagging Ensemble principle. Each stacking model processes different features and produces two outputs collected and passed as one.

The hyperparameter configuration can be seen in Table 1. Keep in mind that there are 9 stack models, so this configuration applies to all layers. Can be seen in Fig. 1 that the dimensions of the model are wide enough, so to reduce the possibility of overfitting, the number of parameters of each layer is kept simple. A little side note, in this model, dropouts are embedded in Mutihead Attention, RNN, and after CNN.

Table 1. Proposed Model's Hyperparameters Configuration

Hyperparameter	Value
Attention head	3
Attention key dimension	3
RNN unit	3
CNN filter	3
CNN kernel size	3
Dense unit	3
Dropouts' probability	0.5
Time-lag length	260

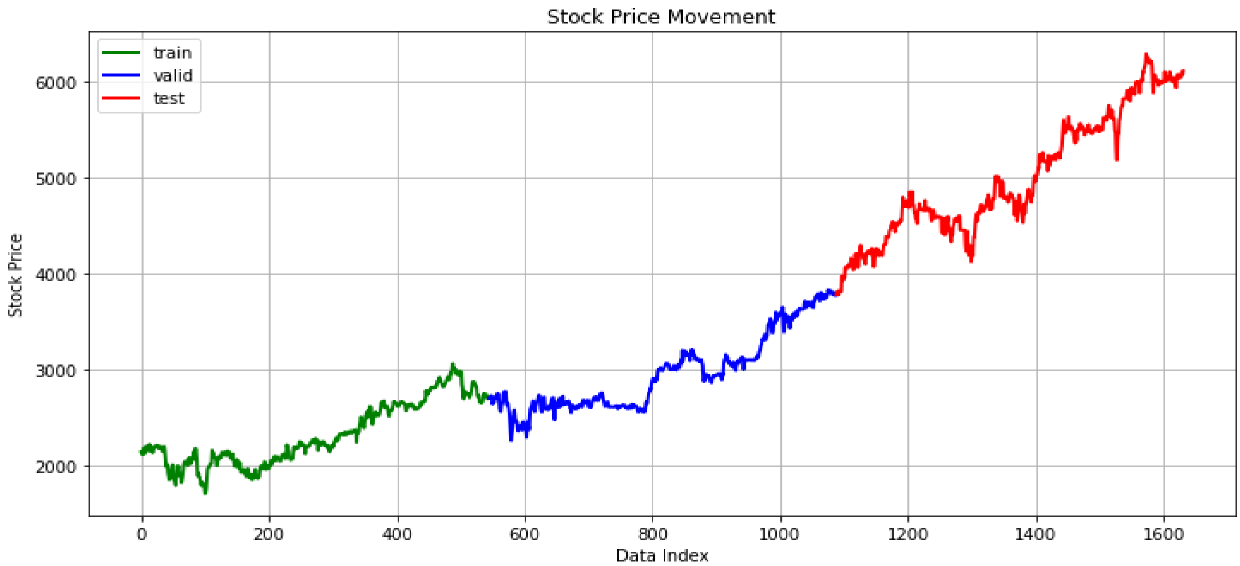

Fig. 5. BBCA stock value movement across the dataset.

3.3. Data Preparation

The two primary data sources are Yahoo Finance (finance.yahoo.com) for Bank Central Asia stock (BBCA) prices and investing.com for ZG gold futures, Brent international crude prices, and CNY-to-IDR and USD-to-IDR exchange rates. All data refer to the working days of the Indonesian stock market from March 15, 2012, to October 12, 2018. It has been verified that each dataset comprises 11 columns and 2979 rows. Date, goal value, high, low, open, close, volume, and five columns of external elements, including the USD to IDR exchange rate, the CNY to IDR exchange rate, the price of ZG gold, and the Brent international oil prices, are included in each column. Remember that the target column represents the stock's closing price the following day. The processed data is divided into three: training sets, validation sets, and testing sets. The proportion of this data division has a ratio of 1:1:1, so that each set has 544 rows of data.

4. Results

4.1. Training and Validation Results

ET-RNN methods will be compared with various combinations of other ensemble models and two models from the previous paper to prove the superiority of the proposed method. Two comparison models from the previous paper are An Ensemble of a Boosted Hybrid of Deep Learning Models and Technical Analysis for Forecasting Stock Prices (EHTS) [26] and CNN-Bidirectional LSTM-Attention Mechanism (CNN-BiLSTM-AM) [25]. A summary of all observations from training and validation results can be seen in Table 2.

Starting from the model developed from previous research, there are two models: EHTS and CNN-BiLSTM-AM. EHTS produces outstanding training performance, and the Root Mean Squared Error (RMSE) of the training set has been achieved to below 30. These results show that EHTS can capture features very well. The validation error is relatively low, with an RMSE of around 330.

Table 2. Ensemble Models' Training and Validation Results

Model	Learning Rate	Batch Size	Epoch	Training RMSE	Training MAPE	Validation RMSE	Validation MAPE
EHTS	0.01	16	10	**25.8758**	**1%**	330.4818	9%
CNN-BiLSTM-AM	0.001	64	10	74.7625	3%	409.5666	11%
Bagging MLP	0.001	16	10	97.7613	3%	464.9484	12%
Bagging CNN	0.001	16	10	97.9076	3%	470.1574	11%
Bagging LSTM	0.01	16	10	365.3106	11%	917.7173	27%
Bagging GRU	0.01	16	10	339.6797	11%	847.8549	24%
Boosting CNN-MLP	0.001	16	10	112.1587	4%	327.0522	9%
Booosting LSTM-MLP	0.01	16	10	351.6848	11%	868.2741	25%
Boosting GRU-MLP	0.01	16	10	49.4592	2%	600.1826	13%
Boosting MHA-MLP	0.001	16	10	120.5957	4%	1861.3034	26%
Stacking CNN-RNNs	0.01	16	10	108.0282	4%	795.7303	14%
ET-LSTM	0.01	16	10	79.7424	3%	**301.0534**	**8%**
ET-GRU	0.01	16	10	79.9422	3%	303.6424	**8%**

For CNN-BiLSTM-AM, the error rate value of the training set is quite good, with an RMSE below 100. However, the validation error is less impressive because it exceeds the RMSE value of 400.

Four bagging ensemble models are compared: Bagging-MLP, Bagging-CNN, Bagging-LSTM, and Bagging-GRU. The training results of these four models are pretty good for Bagging-MLP and Bagging-CNN with RMSE below one hundred. However, unsatisfactory results were obtained by both Bagging-RNNs models, and it is possible that further fine-tuning is needed to improve their performance. As for the validation results, all bagging models still have not succeeded in producing RMSE below 400.

For the boosting ensemble, the four models compared are CNN-MLP, LSTM-MLP, GRU-MLP, and Multi-head Attention-MLP (MHA-MLP). For training, GRU-MLP scored satisfactory results with RMSE below 50. CNN-MLP and MHA-MLP were good at producing Training RMSE of 112 and 120. In contrast, the validation results were only CNN-MLP, producing a reasonably small error rate. Only one model was tested for stacking, namely the stacking of CNN-RNNs. Although there is only one, this model already represents CNN, LSTM, GRU, MLP, and even involves MHA for feature weighting. The training results from this stacking model are pretty good, with RMSE in the range of 100. However, the validation results are unsatisfactory because they produce RMSE above 700.

The two proposed models produce RMSE below 80, and this is entirely satisfactory because it means that the model can extract features quite well. Even so, this value is still surpassed by the two models from previous studies, namely EHTS and CNN-BiLSTM-AM. For validation, Transformer-LSTM and Transformer-GRU outperformed all the comparison models with an RMSE of around 300. Due to their excellent performance on the validation set, the model will continue to the evaluation stage with EHTS and CNN-BiLSTM-AM.

movement of the validation set throughout the training process. The loss function used during training is Mean Squared Error (MSE), and the model is trained for ten epochs. The training loss decreased effectively for only two epochs, then decreased slowly until the fourth epoch, and almost converged from the fifth epoch to the end. As for the loss from validation, it drops significantly in one epoch only, then decreases slowly until the third epoch, and then converges to the end. The loss graph shows that the two proposed models do not require many epochs to be able to adapt to new data.

4.2. Evaluation Results

Table 3. Ensemble Models' Evaluation Result

Model	Testing RMSE	Testing RMSE	ComputationTime
EHTS	1782.8283	32%	24ms
CNN-BiLSTM-AM	2224.1882	42%	9ms
ET-LSTM	490.3815	9%	41ms
ET-GRU	493.7659	9%	38ms

Table 3. shows that the evaluation results of the two models from the previous paper are interesting. Even though EHTS is considered the best when conducting training, the RMSE value has increased to above 1000 in the evaluation process. Likewise, with CNN-BiLSTM-AM, the evaluation results are disappointing; even the RMSE can break above 2000. In contrast to the previous models, Transformer-LSTM and Transformer-GRU maintained the RMSE value below 500 until the evaluation phase.

Regardless of the model used, this BBCA data seems quite challenging to predict. A high numerical value result in a significant nominal RMSE score. Nevertheless, this is not the case with Mean Average Percentage Error (MAPE). Both Transformer-RNNs models have MAPE values below 10% in training, validation, and evaluation data. Apart from the sizeable nominal problem, Tables 2. and 3. prove that all models suffer

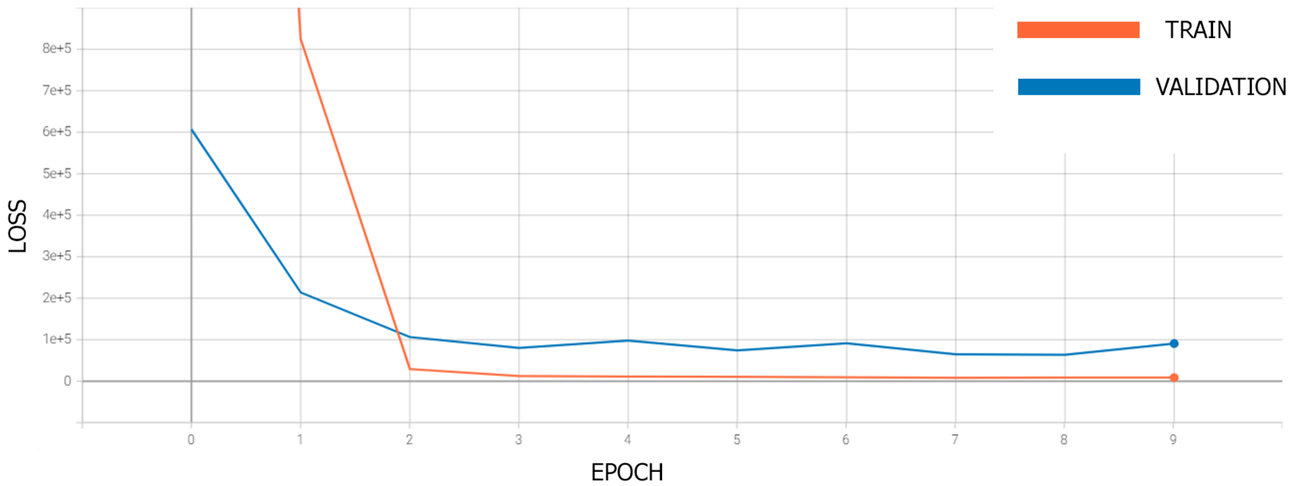

Fig. 6. Transformer-RNN's training loss history through epochs.

The visualization of the Transformer-LSTM training process can be seen in Fig. 6. Please note that the graph of Transformer-GRU is also like Transformer-LSTM. Therefore, this graph will be considered a general historical training graph of Transformer-RNNs. The orange line represents the loss movement of the training set. In contrast, the blue line represents the loss

from overfitting. This problem proves that the movement pattern of the BBCA is quite different between the training, validation, and testing sets.

Even though it has the best performance if we measure it from computing effectiveness, Transformer-RNNs still seem to have a computational time issue. Compared to the previous models, Transformer-LSTM and Transformer-GRU require more time to process the same data. This flaw is because the model's dimensions are broad, causing the model to be quite heavy to run. Further development is needed to reduce the model's dimensions and maintain the resulting performance. Likewise, with the problem of overfitting, all models need to be prepared to deal with contrasting value movement patterns in time series data.

5. Conclusion

This paper proposes two models of ET-RNNs called ET-LSTM and ET-GRU. This model uses the opening price, highest price, lowest price, closing price, and volume, then added with external variables such as exchange rates, gold prices, and crude oil prices as datasets. With the composition of data like this, the model can predict stock prices by considering the country's economic conditions represented by external variables. This architecture utilizes multi-head attention to weighing the sequence. It is followed by the LSTM or GRU layer to extract features from the sequence. It is paralleled with 1D Convolution with similar functions. The series of models are arranged in such a way using three ensemble concepts: bagging, boosting, and stacking to produce a more robust model.

The experimental results prove that ET-LSTM and ET-GRU have the lowest validation and evaluation error values compared to EHTS and CNN-BiLSTM-AM. These results show that ET-RNNs are more resistant to overfitting than other models, so they are suitable as a reference for investors to maximize profits. This research from ET-RNNs also provides technical insight into developing models for processing time series data.

Further development can be done by fine-tuning the ET-LSTM and ET-GRU models to produce lower errors and be more resistant to overfitting. In addition to performance, it is necessary to develop these two models of ET-RNNs to be lighter and faster in computation time. In addition to predicting stocks, this model can be used for other sequence data cases, such as air quality prediction, rainfall prediction, and text processing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Y. Baek and H. Y. Kim, "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module," *Expert Syst. Appl.*, vol. 113, pp. 457–480, 2018, doi: 10.1016/j.eswa.2018.07.019.

[2] D. T. N. Huy, B. T. T. Loan, and P. T. Anh, "Impact of selected factors on stock price: A case study of Vietcombank in Vietnam," *Entrep. Sustain. Issues*, vol. 7, no. 4, pp. 2715–2730, 2020, doi: 10.9770/jesi.2020.7.4(10).

[3] S. Smith and A. O'Hare, "Comparing traditional news and social media with stock price movements; which comes first, the news or the price change?," *J. Big Data*, vol. 9, no. 1, p. 47, 2022, doi: 10.1186/s40537-022-00591-6.

[4] V. Ingle and S. Deshmukh, "Ensemble deep learning framework for stock market data prediction (EDLF-DP)," *Glob. Transitions Proc.*, vol. 2, no. 1, pp. 47–66, 2021, doi: 10.1016/j.gltp.2021.01.008.

[5] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction," *J. Appl. Math.*, vol. 2014, p. 614342, 2014, doi: 10.1155/2014/614342.

[6] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," *Appl. Stoch. Model. Bus. Ind.*, vol. 33, no. 1, pp. 3–12, Jan. 2017, doi: https://doi.org/10.1002/asmb.2209.

[7] H. Yu, R. Chen, and G. Zhang, "A SVM Stock Selection Model within PCA," *Procedia Comput. Sci.*, vol. 31, pp. 406–412, 2014, doi: https://doi.org/10.1016/j.procs.2014.05.284.

[8] D. A. Puspitasari and Z. Rustam, "Application of SVM-KNN using SVR as feature selection on stock analysis for Indonesia stock exchange," *AIP Conf. Proc.*, vol. 2023, no. October 2018, 2018, doi: 10.1063/1.5064204.

[9] D. Oreski, S. Oreski, and B. Klicek, "Effects of dataset characteristics on the performance of feature selection techniques," *Appl. Soft Comput.*, vol. 52, pp. 109–119, 2017, doi: https://doi.org/10.1016/j.asoc.2016.12.023.

[10] L. Takeuchi and Y. Lee, "Applying Deep Learning to Enhance Momentum Trading Strategies in Stocks," *Cs229.Stanford.Edu*, no. December 1989, pp. 1–5, 2013.

[11] Y. Chen, W. Lin, and J. Z. Wang, "A Dual-Attention-Based Stock Price Trend Prediction Model With Dual Features," *IEEE Access*, vol. 7, pp. 148047–148058, 2019, doi: 10.1109/ACCESS.2019.2946223.

[12] F. Wang, "Predicting S&P 500 Market Price by Deep Neural Network and Enemble Model," *E3S Web Conf.*, vol. 214, p. 2040, Jan. 2020, doi: 10.1051/e3sconf/202021402040.

[13] C. Liang, Y. Zhang, and Y. Zhang, "Forecasting the volatility of the German stock market: New evidence," *Appl. Econ.*, vol. 54, no. 9, pp. 1055–1070, Feb. 2022, doi: 10.1080/00036846.2021.1975027.

[14] S. Song and H. Li, "Predicting VaR for China's stock market: A score-driven model based on normal inverse Gaussian distribution," *Int. Rev. Financ. Anal.*, vol. 82, p. 102180, 2022, doi: https://doi.org/10.1016/j.irfa.2022.102180.

[15] A. Sharang and C. Rao, "Using machine learning for medium frequency derivative portfolio trading," 2015.

[16] X. Teng, X. Zhang, and Z. Luo, "Neurocomputing Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction," *Neurocomputing*, vol. 505, pp. 92–100, 2022, doi: 10.1016/j.neucom.2022.07.016.

[17] M. Dixon, D. Klabjan, and J. H. Bang, "Classification-based financial markets prediction using deep neural networks," *Algorithmic Financ.*, vol. 6, no. 3–4, pp. 67–77, 2017, doi: 10.3233/AF-170176.

[18] A. M. El-Masry, M. F. Ghaly, M. A. Khalafallah, and Y. A. El-Fayed, "Deep Learning for Event-Driven Stock Prediction Xiao," *J. Sci. Ind. Res. (India).*, vol. 61, no. 9, pp. 719–725, 2002.

[19] C. A. S. Assis, A. C. M. Pereira, E. G. Carrano, R. Ramos, and W. Dias, "Restricted Boltzmann Machines

for the Prediction of Trends in Financial Time Series," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8. doi: 10.1109/IJCNN.2018.8489163.

[20] S. Rönnqvist and P. Sarlin, "Bank distress in the news: Describing events through deep learning," *Neurocomputing*, vol. 264, pp. 57–70, 2017, doi: https://doi.org/10.1016/j.neucom.2016.12.110.

[21] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Syst. Appl.*, vol. 103, pp. 25–37, 2018, doi: https://doi.org/10.1016/j.eswa.2018.03.002.

[22] S. S. Roy, D. Mittal, A. Basu, and A. Abraham, "Stock Market Forecasting Using LASSO Linear Regression Model BT - Afro-European Conference for Industrial Advancement," 2015, pp. 371–381.

[23] Y. Yan and D. Yang, "A Stock Trend Forecast Algorithm Based on Deep Neural Networks," *Sci. Program.*, vol. 2021, p. 7510641, 2021, doi: 10.1155/2021/7510641.

[24] Y. Liu, C. Gong, L. Yang, and Y. Chen, "DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction," *Expert Syst. Appl.*, vol. 143, p. 113082, 2020, doi: 10.1016/j.eswa.2019.113082.

[25] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Comput. Appl.*, vol. 0123456789, 2020, doi: 10.1007/s00521-020-05532-z.

[26] A. F. Kamara, E. Chen, and Z. Pan, "An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices," *Inf. Sci. (Ny).*, vol. 594, pp. 1–19, 2022, doi: 10.1016/j.ins.2022.02.015.

[27] Z. Dai, H. Zhou, and X. Dong, "Forecasting stock market volatility: The role of gold and exchange rate," *AIMS Math.*, vol. 5, no. 5, pp. 5094–5105, 2020, doi: 10.3934/math.2020327.

[28] M. Alhamid, "Ensemble Models: A guide to learning ensemble techniques in a simple walkthrough," *Towards Data Science*, 2021. https://towardsdatascience.com/ensemble-models-5a62d4f4cb0c