e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

Optimization Of Recurrent Neural Network In Indonesia Stock Exchange Price Prediction Modeling

Dwi Andriyanto[1], Yan Rianto[2*]
Computer Science Study Program[1,2]
Faculty of Information Technology[1,2]
Nusa Mandiri University[1,2]

14002432@nusamandiri.ac.id[1] , yan.yrt@nusamandiri.ac.id[2]

Received: February 12, 2022**. Revised:** March 15, 2022**. Accepted:** March 29, 2022**. Issue Period:** Vol.6 No.1 (2022), Pp. 1-10

Abstrak: Salah satu perdagangan online diantaranya perdagangan saham di bursa, bahkan untuk meningkatkan jumlah investor pemerintah mengajak masyarakat berinvestasi di pasar modal dengan membeli saham secara rutin dan berkala dalam bentuk saham. Dengan bertambahnya calon investor dalam bursa saham tentu dibutuhkan pengetahuan tentang seluk beluk perdagangan saham agar hasil dari pengembalian sesuai yang dikehendaki. Dengan penelitian ini dapat membantu mempermudah masyarakat untuk melakukan prediksi harga saham yang dikehendaki melalui teknologi machine learning meskipun belum memiliki ilmu teknikal tentang saham lebih mendalam. Tiga jenis algoritma RNN,LSTM,GRU digunakan untuk mencari metode terbaik dengan melakukan optimasi parameter sehingga mendapatkan r-square error 0,96 melalui jumlah epoch 100 dan learning rate 0,01.

Kata Kunci: *saham,time series,rnn,lstm,gru, neural network*

Abstract: One of the online trades is stock trading on the stock exchange. To increase the number of investors, the government invites the public to invest in the capital market by buying shares regularly and periodically in the form of shares. With the increase in potential investors in the stock market, deep knowledge about stock trading is needed so that the returns are as desired. This research can help to make it easier for the public to predict the desired stock price through machine learning technology even though they do not have more in-depth technical knowledge about stocks. Three types of algorithms RNN, LSTM, GRU are used to find the best method by optimizing parameters so as to get an r-square error of 0.96 through the number of epochs of 100 and a learning rate of 0.01.

Keywords: *stock, time series, rnn, lstm, gru, neural network*

I. PRELIMINARY

1.1 Background

Almost all countries are having a pandemic known as Covid-19, including Indonesia. The effects of the pandemic is not only occurs to the health aspect but also on people's economic activities. Changes in the pattern of economic transactions, which are usually carried out directly through trading activities in the market, are now starting to adjust through online trading following the government's policy of restricting community activities or known as PPKM.

One of the online trades is stock trading on the stock exchange, even to increase the number of investors, the government invites the public to invest in the capital market by buying shares regularly and periodically in

DOI: 10.52362/jisicom.v6i1.744

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

the form of let's save shares.[1]. With the increase in potential investors in the stock market, knowledge about the ins and outs of stock trading is needed so that the returns are as desired.

The frequency of stock trading will certainly affect the number of outstanding shares, it will attract investors to buy or sell. Frequency is the number of times a share sale and purchase transaction occurs which is recorded in volume form. Trading volume describes the condition of securities and is a description of the liquidity of a security. Volume is the total value of the transaction of buying and selling shares by investors and is often used as a motivational parameter for buying/selling tendencies in order to obtain capital gains.

The market price is determined by the price at the opening (Open) and the price at closing (Close). In one trading day, the highest price (High) and the lowest price (Low) for the stock that occurred on that day are also recorded. The price multiplied by the number of shares traded is the market capitalization value. Stocks with high capitalization/market value are long-term targets for investors to describe the company's fundamentals so that they have a low level of risk. The greater the number of shares traded on a daily basis indicates the level of liquidity of these shares and attracts investors to make trading transactions.

1.2 Related Research

Research conducted by Polash Dey and his colleagues in 2021 with the title "Comparative Analysis of Recurrent Neural Networks (RNN) in Stock Price Prediction for Different Frequency Domains"[2]discusses a comparative analysis of three commonly used RNNs, namely simple RNN, Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU) and analyzes their efficiency for stocks that have different stock trends and different price ranges and for different time frequencies. Using a data set of three companies from June 30, 2000 to July 21, 2020 and stock price fluctuations with prices ranging from $30, $50, and $290 during that period. Honda Motor Company (HMC) as slightly fluctuating data whose share price is between USD 15 to USD 45, Oracle Corporation (ORCL) as moderately fluctuating data with a price range of USD 10 to USD 60, and Intuit Incorporation (INTU) as data that highly volatile with a low of USD 20 and a high of USD 310. The average error performance metric in the form of R-Square is above 96% for stock price fluctuations with 1 and 3 day intervals. As for the 5-day interval, especially Oracle's stock is still around 90%.

Stock price prediction through eight trigrams feature engineering accompanied by ensemble machine learning, especially prediction of daily stock patterns by combining traditional candlestick charts with artificial intelligence methods. The method can provide a machine learning prediction method that is suitable for each pattern based on the results trained using data from 2000 to 2017 the Chinese stock market resulting in a prediction accuracy of more than 60% for some trend patterns. Additional indicators that can improve accuracy such as momentum, increase in the amount of data, standardization of features, and elimination of anomalous data can effectively address data disruption[3].

The use of the Logistics Regression model provides a maximum average accuracy of 68.622% for Apple's stock price prediction. Learning models are compared and analyzed by understanding various technical terms related to finance and capital markets in general in the stock market (Stock Market Index, Outstanding Stock, Market Capitalization, S&P500) very helpful in preprocessing datasets to achieve the best results[4].

The model architecture for stock price prediction is proposed by utilizing two algorithms, namely Long Short Term Memory (LSTM) and BiDirectional Long Short Term Memory (BI-LSTM) model. By setting the right hyper-parameters, the research can predict future stock trends with high accuracy. RMSE for the LSTM and BI-LSTM models is measured by varying the number of epochs, hidden layers, solid layers, and different units used in the hidden layers to find a better model that can be used to accurately forecast future stock prices.[5].

The solution to solve the problem is using a hybrid mechanism of the RNN-LSTM network integrated with metaheuristic optimization techniques, obtained with hyper-adjusted parametric values leading to a more precise learning process with minimized error rates and increased accuracy. Moreover, the comparative results evaluated over six different exchange data sets reflect the efficacy of the optimized RNN-LSTM network by achieving maximum forecasting accuracy of about 4-6% increments.[6].

Analysis look-back period parameter with iterative neural network compares the stock price prediction performance of three deep learning models namely Vanilla RNN, LSTM, and GRU to predict stock prices of the

DOI: 10.52362/jisicom.v6i1.744

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

two most popular and strongest models of commercial banks listed on the Nepal Stock Exchange (NEPSE). From the experiments conducted, it was found that the GRU was the most successful in predicting stock prices. Moreover, research work has suggested suitable values of the look-back period that can be used with LSTM and GRU for better stock price prediction performance.[7].

1.3 Objectives and Scope

The research was conducted with the aim of building a Deep Learning model with Hyperparameter Tuning to improve model performance. In addition, building a model with other methods to find a method that matches the related dataset based on the performance of the model, and produces a good model in predicting stock prices. For related parties, this research can help the public to be more interested in investing through stock exchange trading without having to master stock technical knowledge more deeply.

To make it easier to do limitations, the scope that will be stated in this study includes, the dataset is limited to Oracle Corporation[8]as moderately fluctuating data (price range USD 10 to USD 60). The best method from this research is then applied to stocks listed on the Indonesia Stock Exchange using random stock data having a market capitalization of above Rp500 billion, taken based on the price fraction category according to Rule Number II-A concerning Trading in Equity Securities Number KEP-00061/BEI/07-2021 dated July 23, 2021[9].

This research is expected to help make it easier for the public to predict the desired stock price with the help of machine learning even though they do not have more in-depth technical knowledge about stocks. This research offers novelty through epoch optimization and learning rate on stock price predictions that have a medium fluctuating value with an R-Square error performance metric of 0.96 which is better than previous studies.[2]which reached 0.90.

II. METHODS AND MATERIALS

Different types of repetitive neural networks *Recurrent Neural Network* (RNN) is used in time series analysis, especially in stock price prediction. However, because not all stock prices follow the same trend, one model cannot be used to predict movements of all types of stock prices. Therefore, in this study a comparative analysis was carried out on three commonly used RNNs, namely simple RNN, Long Short Term Memor (LSTM), and Gated Recurrent Unit (GRU).

2.1 Recurrent Neural Network (RNN)

This algorithm has an architecture with few layers with backward connection flow so it is considered complicated. With this backward connection, it can handle sequential data such as historical data (time series), process text structures such as sentences, sub-words or letters, but can also handle moving image (video) data sets. That's why because it can handle data with backward connections, RNN is considered a Deep Learning category because it can solve the vanishing gradient problem[10]. In accordance with the characteristics of the RNN which is a neural network layer, although simple, it requires repetition because it will be used several times with successive times. So that the learning process in RNN is bound to a time series, therefore this algorithm is also known as Back Propagation Through Time (BPPT) learning.

The output of the hidden layer in the t-time step is used the formula:

$$h_t = f_w(h_{t-1}, x_t) \tag{1}$$

$$h_t = \tanh(w_{hh}\, h_{t-1} + w_{sh}\, x_t + b_h) \tag{2}$$

The output for the t-th time step is used the formula:

$$o_t = f_0(w_{ho}\, h_t + b_0) \tag{3}$$

For the loss function, the cross entropy formula is used:

$$E_t(o_t,o_t) = -\, O_t \log(\hat{O}_t) \tag{4}$$

$$E = \sum_t E_t(O_t,\hat{O}_t) = \sum_t -\, O_t \log(\hat{O}_t) \tag{5}$$

DOI: 10.52362/jisicom.v6i1.744

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

where E is the total loss function in one input sequence, $E_t(o_t,o_t)$ is the loss function for one time step t and t is the actual value of the output in the t-time step.

2.2 Long Short Term Memory (LSTM)

The first step in the LSTM is to determine the information to be generated from Ct-1 using the sigmoid function (σ) or what is known as the forget gate. The function of the sigmoid is to determine whether the information will be forwarded with a value of 1 or the information will be terminated because it is worth 0. The formulation is as follows:

$$f_t = \sigma (w_f [S_{t-1},x_t] + b_f) \quad (6)$$

Next, determine the information to be added to Ct which is a combination of the previous hidden state St-1 and the current information xt by using the sigmoid function as an input gate and the tanh function as an intermediate gate. The result of the information is multiplication which will be added to the cell state in the current state of Ct. The formula used is:

$$i_t = \sigma (w_i [S_{t-1},x_t] + b_i) \quad (7)$$

$$\hat{c}_t = \tanh (w_c [S_{t-1},x_t] + b_c) \quad (8)$$

$$\hat{c}_t = f_t * C_{t-1} + i_t * \hat{c}_t \quad (9)$$

The last step is to determine the LSTM output, it is necessary to first calculate the sigmoid from the combination of St-1 and xt which is the output gate. This output gate determines the resulting cell state value at St. Furthermore, the calculated value of the tanh function of Ct is multiplied by the output gate value and the result becomes the output of the LSTM:

$$_{Ot} = \sigma (w_0 [S_{t-1},x_t] + b_0) \quad (10)$$

$$S_t = \tanh * (C_t) \quad (11)$$

2.3 Gated Recurrent Units (GRU)

The development of LSTM for several variations such as peephole connection, which combines input and forgotten gates, is known as the Gate Reccurent Unit (GRU) which was popularized by Cho and Chung in 2014. The principle is to simplify the computational process but have equal performance. In the GRU the forgotten gate is combined with the input gate into one, namely the update gate. GRU also combines cell state into hidden state[11].

GRU is also similar to RNN in making improvements in order to eliminate gradient problems. The difference is in the GRU algorithm using update gates and reset gates where this is an important thing in deciding which information to pass to the output. More specifically, the GRU can store information without deleting it over time and can delete information that is not relevant to what was predicted in the data.[2]. The update gate helps the model determine how much past information to pass into the future. This has a strong impact as the model can retain information from the past and eliminate the risk of missing gradient problems. The equations used are:

$$z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1} + b^{(z)}) \quad (12)$$

The reset gate is used to decide how much past information should be forgotten and is calculated by the equation:

$$r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1} + b^{(r)}) \quad (13)$$

In the end the cell output will be calculated by the equation:

$$r_t = z_t \odot h_{t-1} + (1-z_t) \odot \hat{h}t \quad (14)$$

where is the candidate activation vector

$$\hat{h}t = \tanh (Wx_t + r_t \odot Uh_{t-1} + b^{(h)}) \quad (15)$$

2.4 Performance Metrics Error

To ensure that the model made is working as expected, a clear measure is needed. The trick is to compare the actual value (Y) and the predicted value (Ẏ) for a number of (n) data. The smaller the error value, the better the results. This research is included in the category of regression so that the metrics used include[12]:

DOI: 10.52362/jisicom.v6i1.744

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

- R Square or Coefficient of Determination

R Square or the coefficient of determination can be interpreted as the proportion of variance in the dependent variable that can be predicted from the independent variable. The closer to 1 the result is the best[13]. The formula used is:

$$R2 = 1 - \frac{\sum_{i=1}^{n} (\dot{Y}i - Yi)^2}{\sum_{i=1}^{n} (Yi - \dot{Y}i)^2} \quad (16)$$

- *Mean Absolute Error*(MAE)

The absolute value of the difference between the predicted value and the actual value. The goal of absolute operation is to eliminate positive and negative errors. The formula used is:

$$MAE = \sum_{i=1}^{n} |Yi - \dot{Y}i| \quad (17)$$

- *Mean Absolute Percentage Error*(MAPE)

Has a very intuitive interpretation of relative error making it more suitable for handling errors with relative variation rather than absolute variation. However, it has a number of drawbacks to data that are strictly positive by definition and bias towards low estimates. Not suitable for predictive models where large errors are expected[13]. The formula used is:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Yi - \dot{Y}i}{Yi} \right| \quad (18)$$

- *Root Mean Squared Error*(RMSE)

The squared value of the difference between the predicted value and the actual value. The purpose of the quadratic operation is to eliminate negative and positive errors so that they have the same contribution. The formula used is:

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n} (Yi - \dot{Y}i)^2}}{n} \quad (19)$$

2.5 Method

Devices in processing data using Google Colab Pro[14] equipped with a Graphics Processing Unit (GPU) accelerator with 8GB of memory. As an initial step in the experiment, we present the stages of research starting from the dataset to the performance evaluation model to make the basis for making decisions on the best model performance.

Source: Research Results (2021)
Figure 1. Research Stages

This research is focused on stock trading with 5 day interval. Therefore, an initial activity is needed in the form of adjusting the trading interval from 1 day to 5 days. The steps taken are started by determining the features (date) of the first day of the 5-day period. Value (open) is inputted with the opening price of the first day of the 5 day timeframe. Value (high) is filled with the highest price within a period of 5 days. Value (low) is filled with the lowest price within 5 days. The value (close) is filled with the closing price of the last day of the 5-day period. As for the (volume) traded, it is filled with the total trading volume within a period of 5 days.

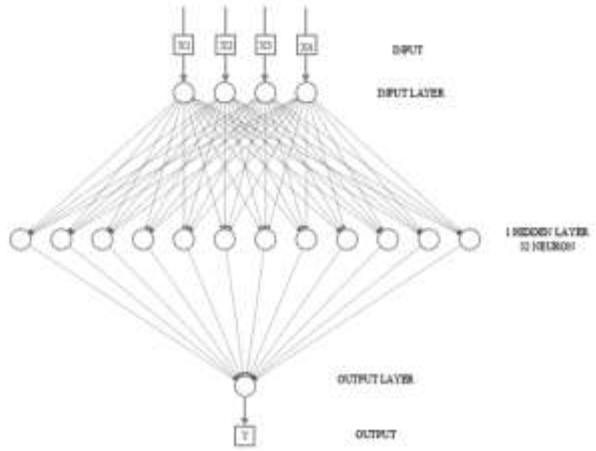
One of the obstacles in researching data with time series categories is the presence of data with empty values. If there is data with empty values, it will be able to disrupt the pattern of price movements, so that if it is not handled properly it will have an influence on stock price predictions. In general, handling empty data can be filled with a certain value such as using the previous stock price. This method will be used in research to solve the problem of empty data.

Scikit-learn has a scale handling feature, namely StandardScaler which is used to make each feature have an average value of 0 and a variance of 1 MinMaxScaler which functions to change data is in the range 0 to 1.

DOI: 10.52362/jisicom.v6i1.744

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

The model that will be developed for Oracle stock price prediction as shown in Figure 2 uses a neural network model with 4 input layers, 1 hidden layer containing 32 neurons, layer 1 output and the algorithms used are RNN, LSTM, and GRU. The plan uses 81 models trained with training data on 90% of the total data and tested with data testing of 10% of data.



Source: Research Results (2021)
Figure 2. Architectural Model Design

By using neural network and RNN derived algorithm as the first hidden layer. Adam, Adamax, RMSProp were used as the model optimizer, and MSE was used as the loss function. The epochs used are 100, 200, 300 and the learning rate is 0.1, 0.01, 0.001. Various experiments were carried out during the training process where further evaluation of each model would be carried out. The evaluation of the model will use data testing using scikit-learn libraries. The performance metrics used to evaluate the model are R-Square or the coefficient of determination, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE). To ensure that the selected model is not over-fitting, the experiment is also equipped with plotting training and loss history along with mae and mse.

Through model design and performance metrics the error that has the smallest error and the fastest computation time will be selected for the model. The model is then applied to stock price predictions on the Indonesia Stock Exchange with certain criteria.

III. DISCUSSION AND RESULTS

3.1 Discussion

Table 1 shows the dataset has 1466 instances. The lowest open price is $8 and the highest open is $60 with a standard deviation of $14. The first quartile of the open price is $16, the second quartile is $28, and the third quartile is $40. Average open price $29. The minimum high is $9 and the maximum high is $61 with a standard deviation of $14. The first quartile of high prices is $16, the second quartile is $29, and the third quartile is $41. Average price high $29. The minimum low is $7 and the maximum low is $59 with a standard deviation of $14. The first quartile of low prices is $15, the second quartile is $27, and the third quartile is $40. Average low price $28. The minimum closing price is $8 and the maximum closing price is $60 with a standard deviation of $14. The first quartile of the closing price was $15, the second quartile was $28, and the third quartile was $40. Average closing price $29. The minimum volume is 5,690,833 and the maximum volume is 126,909,300 with a standard deviation of 17,139,650. The first quartile of volume is 16,127,140, the second quartile is 28,065,560, and the third quartile is 39,884,180. The average volume is 30,361,360.

DOI: 10.52362/jisicom.v6i1.744

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

Table 1. Statistical Description

	Open	*High*	*Low*	*Close*	*Adj.Close*	*Volume*
count	1,466	1,466	1,466	1,466	1,466	1,466
mean	29	29	28	29	26	30,361,360
std	14	14	14	14	14	17,139,650
min	8	9	7	8	7	5,690,833
0.25	16	16	15	15	13	16,127,140
0.50	28	29	27	28	24	28,065,560
0.75	40	41	40	40	37	39,884,180
max	60	61	59	60	59	126,909,300

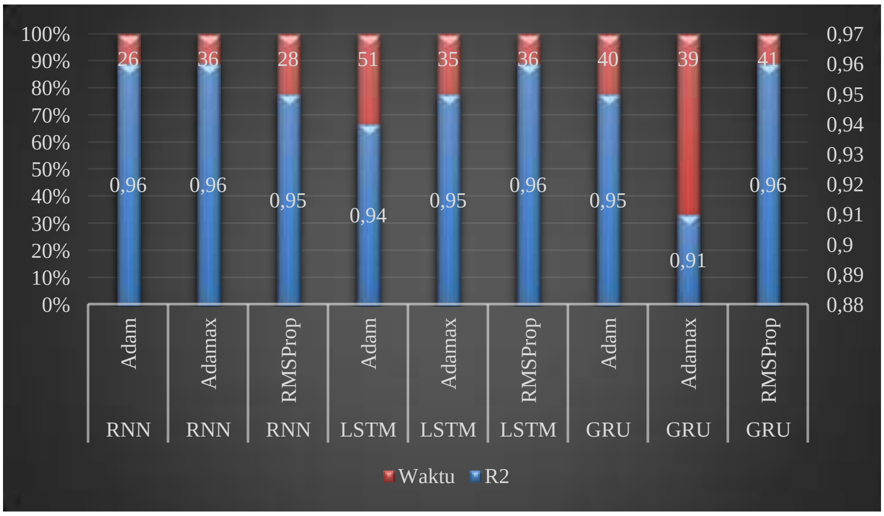
Source: Research Results (2021)

3.2 Results

To facilitate the assessment of model performance, experiment on Oracle stock data by grouping it into the RNN, LSTM, and GRU algorithms. According to the model development plan by adding tuning parameter settings in the form of Adam, Adamax and RMSProp optimizers.

In forecasting stock prices, the most important thing is the smaller the level of error achieved. However, another consideration that is equally important is the computational speed to achieve the best prediction results. For this reason, this study considers in addition to focusing on the minimum error level, also to the time period needed to carry out the model calculation process so that the best results are obtained.

This research has obtained nine best models where each algorithm is represented by the best 1 optimizer. The average coefficient of determination has a value above 0.95, only one has a value of 0.91, namely the GRU algorithm model and the Adamax optimizer. The time required for the calculation of the model is 51 seconds and the fastest is 26 seconds. The speed of this process is proportional to the number of epochs used, namely 100, 200, 300 with the number of batch sizes used in 16 batches.



Source: Research Results (2021)
Figure 3. The Best Model Performance Graph

In this case the research proposal with the best performance model obtained the calculation results with the fastest time of 26 seconds, epoch 100, learning rate 0.01, MAE 0.53, MAPE 0.01, RMSE 0.70 and R-Square

DOI: 10.52362/jisicom.v6i1.744

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

with a value of 0.96. This performance calculation belongs to the highest R-Square category for the RNN algorithm model with the Adam optimizer as shown in Figure 3.

3.3 Implementation

The model with the best performance can be applied for stock prediction on the Indonesia Stock Exchange (IDX), of course, through several experiments, this can be answered. For this reason, a simple application was developed using data that has a market capitalization above Rp500 billion based on price fraction category according to Rule Number II-A concerning Trading of Equity Securities Number KEP-00061/BEI/07-2021 dated 23 July 2021[9]. After getting the issuer code, the dataset is taken from the Yahoo Finance page[15]. The list of shares to be used are:

Table 2. Model Application for Indonesian Stock Exchange Shares

No	*Code*	*Market Cap(Rp)*	*Price(Rp)*	*Price Fraction Category*	*Start Listing*
1	ASII	204.4 T	5050	5	04-04-90
2	BRIS	87.9 T	2160	4	01-01-11
3	UNVR	153.7 T	4030	4	11-01-82
4	KLBF	75 T	1615	3	30-07-91
5	BBKP	16.3 T	500	3	10-07-06
6	MLPL	6.9 T	468	2	06-11-89
7	BABP	11.0 T	436	2	15-07-02
8	BHIT	7.0 T	95	1	24-11-97
9	FREN	38.8 T	126	1	29-11-06
10	ARTO	213.0 T	15525	5	12-01-16
11	BBRI	633.5 T	4110	4	10-11-03

Source:[15](has been reprocessed)

Share price is less than Rp200,- (two hundred rupiah) category 1, share price is in the range of Rp200,- (two hundred rupiah) to less than Rp500,- (five hundred rupiah) category 2, share price is in the range of Rp500,- (five hundred rupiahs) up to less than Rp2,000,- (two thousand rupiah) category 3, share prices are in the range of Rp2,000,- (two thousand rupiahs) to less than Rp5,000,- (five thousand rupiah) rupiah) category 4, share price of Rp. 5,000,- (five thousand rupiah) or more, shall be assigned to category 5. Thus, the classification of share prices is categorized as low in price fraction 1 and highest in price fraction 5.

Table 3. Results of Model Testing for IDX Stocks

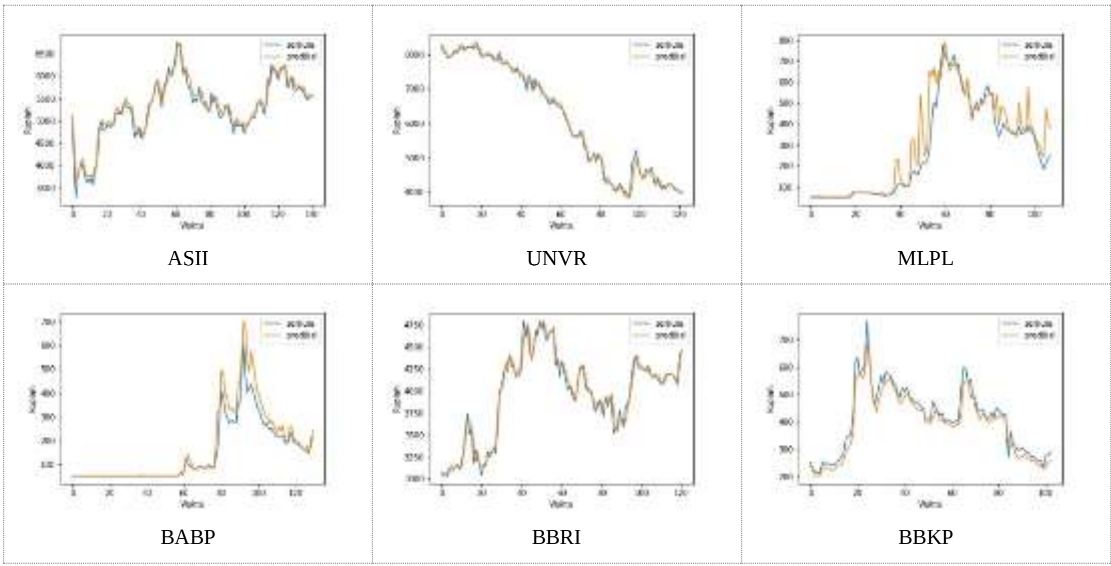
No	*Code*	*Price Fraction Category*	*R2*	*MAE*	*MAPE*	*RMSE*	*Time*
1	ASII	5	0.96	78.63	0.01	96.83	83
2	BRIS	4	0.7	65.33	0.03	69.84	12
3	UNVR	4	0.97	201.02	0.04	247.94	47
4	KLBF	3	-0.7	17.03	0.01	21.76	8
5	BBKP	3	0.94	24.59	0.06	30.98	40
6	MLPL	2	0.86	44.14	0.19	81.96	42
7	BABP	2	0.88	22.84	0.09	43.88	83
8	BHIT	1	-9	20.45	0.23	57.64	84
9	FREN	1	-46	91.99	0.97	151.61	42
10	ARTO	5	-2.4	2174.3	0.14	2249.15	16
11	BBRI	4	0.97	58.65	0.01	79.22	43

Source: Research Results (2021)

DOI: 10.52362/jisicom.v6i1.744

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.6 No.1, June 2022
Journal of Information Systems, Informatics and Computing
Website/URL:http://journal.stmikjayakarta.ac.id/index.php/jisicom
E-mail:jisicom@stmikjayakarta.ac.id,jisicom2017@gmail.com

Based on Table 3 for IDX stock data[16], this model can be used for price fractions 2, 4, and 5, but cannot be used for price fractions with category 3. Stocks that can use this model are ASII (0.96), UNVR (0.97), MLPL (0.86), BABP (0.88) and BBRI (0.97), BBKP (0.94) as shown in Figure 4, while the rest cannot be predicted using this model.



Source: Research Results (2021)
Figure 4. Graph of model application on 6 IDX Stocks

IV. CONCLUSION

The use of neural network algorithms such as RNN, LSTM, GRU is very helpful in solving regression problems related to price predictions, especially data that is sequence/time series such as stock data. Preprocessing data really helps model performance to produce minimal errors such as filling in blank values with the average value of the previous stock price. The selection of the data scale type in the form of a standard scaler can improve performance than the minmax scaler. The need for parameter settings in order to produce model performance in accordance with the characteristics of the stock data being researched. The best error performance metrics reach R Square with a value of 0.96 with the model architecture, namely the neural network model with input layer 4, hidden layer 1 with 32 neurons, output layer 1; model parameters with input layer activation and hidden layer using ReLU, output layer activation using sigmoid, number of epochs 100, and batch size 16; The best optimization uses Adam's optimizer with a learning rate of 0.01. The performance of the model proposed in this study can produce the smallest error performance using the RNN Algorithm compared to previous studies, namely R Square 0.90[2]. The application of the model generated from this research has been simulated on the stock price of the Indonesia Stock Exchange such as with R-Square errors, namely ASII (0.96), UNVR (0.97), MLPL (0.86), BABP (0.88) and BBRI (0.97), BBKP (0.94).

Different levels of stock price fluctuations require further experimentation with various combinations of parameters such as batch size, epoch, unit activation, and other parameters. The use of other algorithms such as the Facebook Prophet and the addition of engineering features related to stock prices such as stock technical, fundamental, and sentiment analysis can be considered for further research.

REFERENCE

[1] "Let's Save Stocks,"*Indonesia stock exchange*, 2017. https://yuknabungsaham.idx.co.id/about-yns

DOI: 10.52362/jisicom.v6i1.744

(accessed Sep. 05, 2021).

[2] P. Dey*et al.*, "Comparative Analysis of Recurrent Neural Networks in Stock Price Prediction for Different Frequency Domains," Algorithms, vol. 14, no. 8, p. 251, Aug. 2021, doi:10.3390/a14080251.

[3] Y. Lin, S. Liu, H. Yang, and H. Wu, "Stock Trend Prediction Using Candlestick Charting and Ensemble Machine Learning Techniques With a Novelty Feature Engineering Scheme,"*IEEE Access*, vol. 9, pp. 101433–101446, 2021, doi:101109/ACCESS.2021.3096825.

[4] S. Ravikumar and P. Saraf, "Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms," in*2020 International Conference for Emerging Technology (INCET)*, Jun. 2020, pp. 1–5, doi:10.109/INCET49848.2020.9154061.

[5] MA Istiake Sunny, MMS Maswood, and AG Alharbi, "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model," in*2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, Oct. 2020, pp. 87–92, doi:10.109/NILES50944.2020.9257950.

[6] K. Kumar and MTU Haider, "Enhanced Prediction of Intra-day Stock Market Using Metaheuristic Optimization on RNN–LSTM Network,"*New Gen. Comput.*, vol. 39, no. 1, pp. 231–272, Apr. 2021, doi:10.1007/s00354-020-00104-0.

[7] AS Saud and S. Shakya, "Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE,"*Procedia Comput. science.*, vol. 167, pp. 788–798, 2020, doi:10.1016/j.procs.2020.03.419.

[8] "Comparative-Analysis-of-RNNs-in-Stock-Price-Prediction." https://github.com/ehfahad/Comparative-Analysis-of-RNNs-in-Stock-Price-Prediction/blob/main/Comparative-Analysis-of-RNNs-in-Stock-Price-Prediction.zip (accessed Sep 05, 2021).

[9] PT. Indonesia Stock Exchange, "Kep-00061/BEI/07-2021," Jakarta, 2021. [Online]. Available: https://www.idx.co.id/media/10022/peraturan_ii_a_perdagangan_ Efek_berisah_ekuitas.pdf.

[10] Suyanto DR.ST.M.Sc, K. Nur Ramadhani ST, MT, and S. Mandala Ph.D,*Machine Learning Modernization For Big Data*, 1st ed. Bandung: Bandung Informatics, 2019.

[11] Y. Wang, W. Liao, and Y. Chang, "Gated Recurrent Unit Network-Based Short-Term Photovoltaic Forecasting,"*Energies*, vol. 11, no. 8, p. 2163, Aug. 2018, doi:10.3390/en11082163.

[12] B. Santosa and A. Umam,*Data Mining and Big Data Analytics*, 2nd ed. Yogyakarta: Pustaka Media Publisher, 2018.

[13] D. Chicco, MJ Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,"*PeerJ Comput. science.*, vol. 7, p. e623, Jul. 2021, doi:10.7717/peerj-cs.623.

[14] "Google Colab." https://colab.research.google.com/ (accessed Sep. 05, 2021).

[15] "Yahoo Finance." https://finance.yahoo.com/.

[16] "IDX Stock Application List." https://github.com/nausya/thesis/tree/main/indosham (accessed Jan. 02, 2021).

DOI: 10.52362/jisicom.v6i1.744