DOI: 10.15849/IJASCA.230720.15

Int. J. Advance Soft Compu. Appl, Vol. 15, No. 2, July 2023
Print ISSN: 2710-1274, Online ISSN: 2074-8523

A Comparison of Two Deep Learning Models on The Stock Exchange Predictions

Dyah E Herwindiati, Janson Hendryli, and Nor Haniza Sarmin

Department of Informatics, Faculty of Information Technology, Universitas Tarumanagara, Jakarta, Indonesia
e-mail: dyahh@fti.untar.ac.id
Department of Informatics, Faculty of Information Technology, Universitas Tarumanagara, Jakarta, Indonesia
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

Abstract

In this study, we introduce the deep learning approach for the time series forecasting model, particularly for the stock price prediction, using two popular deep learning methods: the long short-term memory (LSTM) networks and the gated recurrent unit (GRU) networks. The data are collected from companies in the LQ45 index of the Indonesian Stock Exchange and the deep learning models are implemented using the Python programming language and the TensorFlow library. The results are evaluated using root mean squared error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination or R2. From the experiment, we demonstrate that the LSTM model achieves RMSE 0.005844, MAPE 0.01427, R^2 0.99898, and the GRU model achieves RMSE 0.005601, MAPE 0.001594, R^2 0.99907 in the training and validation phase where we utilize data up to Dec 31st, 2022. Furthermore, we test the model using unseen data from PT Adaro Energy Indonesia Tbk and find the GRU model achieves better performance with R^2 0.66885 compared to the LSTM with R^2 0.38756. From the experiment, we find that the deep learning approach can be considered a good forecasting model.

Keywords: *deep learning, GRU, LSTM, stock price prediction, time series forecasting models*

1 Introduction

The time series data is the data that is observed successfully over certain period of time [1]. The main goal of time series analysis is to develop a suitable model to describe patterns or trends in the data more accurately. However, forecasting time series data predict future outcomes based on the past. The performance of a time series model can be interpreted based on its error. There are four general types of patterns for time series data: horizontal patterns, trends, seasons, and cycles. An essential step in choosing the correct method for time series data is to consider various data patterns.

Forecasting time series models, including economics, business, and finance, are always used daily. A time series is a sequence of observations on a variable measured at

Received 4 May 2023; Accepted 27 June 2023

successive points or over consecutive periods. Measurements can be taken hourly, daily, weekly, monthly, annually, or regularly. The patterns in the data are essential in understanding how time series have behaved in the past.

Time series forecasting models are a particular class of predictive modeling used to forecast future events. The autoregressive integrated moving average (ARIMA) models have been explored in the literature for forecasting time series before machine learning and deep learning methods were introduced. The ARIMA model is introduced by [2]. The Box-Jenkins methodology consists of activities to identify, estimate, and diagnose the ARIMA model with time series data. ARIMA is a forecasting technique with a time series approach that uses correlation techniques. The ARIMA model has demonstrated an efficient ability to generate short-term forecasts and ARIMA is very well used in short and medium-term forecasting. ARIMA has several essential stages: stationarity, significance test model parameters, and model diagnostic tests. The weakness of the model estimation in the traditional ARIMA is that the model fitting depends on the skill and experience of the researcher.

Deep learning is a subset of machine learning, which is a field that has gained a lot of popularity recently. According to [4], Arthur L. Samuel, one of the pioneers of artificial intelligence, defines machine learning as "the field of study that gives a computer the ability to learn without being explicitly programmed". In general, we prefer to define machine learning as the combination of powerful techniques in computer science, mathematics, and statistics that are being used to discover hidden patterns in complex datasets. Both machine learning and deep learning are within the domain of artificial intelligence and are closely related. Deep learning focuses on artificial neural networks with deep interconnected layers which can learn hidden representations of data, extract complex features, and make predictions that are highly accurate.

This paper aims to introduce the time series forecasting models with deep learning approaches to social field researchers. The methods discussed in this paper are the long short-term memory (LSTM) model and the gated recurrent unit (GRU). Both are part of a deep learning model architecture called recurrent neural networks (RNN) that are most suitable for sequential and time series data. The LSTM model is developed to improve the limitation of RNNs for long sequential data. The GRU is successively an improvement of LSTM by simplifying the unit computations.

2 Related Work

In machine learning terminology, the stock price prediction is a regression problem with a real or continuous target value. Several studies have shown that the performance of regression model can be better compared to the time series methods [5]. Furthermore, the recent trend in quantitative finance incorporates not only traditional forecasting techniques, such as ARIMA but also machine learning and deep learning model. For example, [6] proposes the combination of LSTM networks and machine learning models to predict Bitcoin prices. The study from [6] incorporates LSTM to extract structured financial information from news and apply the information to a machine learning model.

Meanwhile, in the field of stock price forecasting, [7] applies LSTM and then combines and compares it to the ARIMA model. [8] introduces a deep convolutional generative adversarial network (DCGAN) architecture to forecast the closing price of stocks, although the DCGAN itself is not initially a forecasting model, a deep learning model introduced by [9] for image classification tasks. In [10], the authors employ simple three-layer LSTM to predict stock prices from the LQ45 indices with a mean absolute percentage error of 18.6135.

3 Methodology

The steps to employ deep learning models for time series forecasting start from data preprocessing, model training, evaluation, and testing. In the data preprocessing step, the raw data are prepared and transformed into a format that can be easily and effectively processed by the learning model. Some of the techniques in preprocessing the raw data include data cleaning, transformation, normalization, feature extraction, and so on. After the preprocessing steps, the data are usually split into 3 subsets, namely the training, validation, and test sets. As explained above, deep learning models can predict future data by learning the patterns from the training set. The validation set is used to evaluate and fine-tune the trained model weights or parameters to achieve better prediction accuracy. Finally, the performance of the deep learning model is evaluated on the unseen test set. The practice of splitting the data is mainly to prevent overfitting the model to the training data. By testing the model on unseen data, we ensure that the model not only performs well only on the training data but can be generalized to the data it has never seen.

3.1 Data

This study uses the closing prices of companies in the LQ45 stock market index from the Indonesia Stock Exchange (IDX). The LQ45 consists of 45 companies with "high liquidity and huge market share" [10] from several sectors, such as banking, real estate, mining, telecommunication, etc. The stock price data is collected daily from their initial public offering (IPO) date to December 31, 2022.

The historical stock price data are then preprocessed to remove rows with missing values. Furthermore, the data are scaled to a given range, i.e., between minus one and one. We obtain 182,582 rows of data after these preprocessing steps. Fig. 1 shows the example of the price chart for PT Telkom Indonesia Tbk (TLKM.JK), PT Bank Rakyat Indonesia Tbk (BBRI.JK), and PT Adaro Energy Indonesia Tbk (ADRO.JK) after the preprocessing steps. Note that the length of data varies because of the difference in their IPO date.

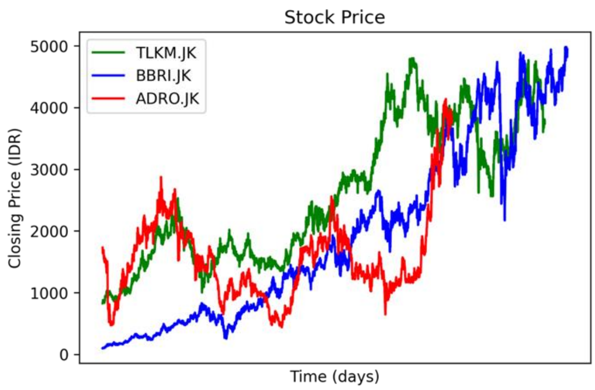

Fig. 1. Visualization of the historical stock price of TLKM, BBRI, and ADRO

3.2 Preprocessing

The next step is to restructure the time series stock price data as input for the deep learning models. The method used for transforming the dataset is the sliding window method [11] as illustrated in Fig. 2. Each row d_i in the transformed data is defined as $(x_{k-\delta}, \dots, x_{k-2}, x_{k-1}, y)$ where y denotes the closing price on day k, x_{k-i} denotes the closing price on the previous i days, and δ denotes the maximum time lag or window width [12]. Finally, the transformed data is split into train and test sets with a 70%-30% proportion, respectively.

Date	Close Price
2020-01-01	8095
2020-01-02	8050
2020-01-03	8055
2020-01-04	8075
2020-01-05	8100
2020-01-06	8125
2020-01-07	8105
2020-01-08	8200
2020-01-09	8225
2020-01-10	8250

Sliding Window

X0	X1	X2	Y
8095	8050	8055	8075
8050	8055	8075	8100
8055	8075	8100	8125
8075	8100	8125	8105
8100	8125	8105	8200
8125	8105	8200	8225
8105	8200	8225	8250

Fig. 2. The reorganization of historical stock price data as inputs for deep learning models with time lag = 3

3.3. Deep Learning

Deep learning is a field in machine learning that studies the neural networks with multiple hidden layers. The key advantage of the multiple hidden layers is the ability to learn input features directly by extracting complex features from the input data. Meanwhile, the feature engineering in traditional machine learning approaches is manually done by human experts who select relevant features from the data.

The deep learning approach has recently demonstrated remarkable success in various tasks, including in the field of natural language processing and computer vision. Yet, the deep learning dates to as early as the 1940s by different name, such as cybernetics, connectionism, and artificial neural networks [13]. The earliest deep learning models were designed to take a set of n input values $x_1, x_2, \dots, x_n$ and returned an output y by computing $y = f(x, w) = x_1 w_1 + x_2 w_2 + \cdots + x_n w_n$ with $w_1, w_2, \dots, w_n$ denotes the weights or model parameter. For example, the McCulloch-Pitts neuron, introduced in 1943, could differentiate two different categories by checking whether y is positive or negative. The weights in the McCulloch-Pitts neuron could be set by human operators. In the 1950s, the perceptron model was introduced. The perceptron was famously known to have a limitation where it cannot learn the XOR function. The problem was solved by introducing hidden layers on the perceptron that can learn non-linear representations. This solution gave birth to the multilayer perceptron model which, consequently, developed to the neural networks.

Another advance in the deep learning field occurs in the 1980s and 1990s with the introduction of key concepts such as the convolutional neural network for image classification and the backpropagation algorithm which remain the dominant approach to train deep learning models. Finally, the wave of deep neural networks research in the 2000s popularized the use of the term "deep learning" [13].

3.4 Recurrent Neural Network

In this study, we employ two deep learning methods to predict the stock price: the long short-term memory (LSTM) networks and the gated recurrent unit (GRU) networks. Both are variants of a family of neural networks, the recurrent neural networks or RNNs, for processing sequential data [13]. The RNNs have shown impressive performance in various tasks that involve sequential data, such as the language translation [14], text summarization [15], audio transcription [16], and voice authentication [17].

The key feature of an RNN is its ability to capture and utilize the temporal dependencies present in sequential data. The RNN has connection that allow information to persist and be shared across time steps. The basic building block of an RNN is a memory cell that maintains a hidden state. At each time step, the memory cell takes as input the current input data and the previous hidden state. It performs computations using learned weights and produces both an output and an updated hidden state. The process is repeated for each time step, allowing the network to process the entire sequence.

Training an RNN involves labeled sequential data and the backpropagation through time (BPTT) method. One of the challenges of learning long-term dependencies in RNN arises from the exponentially smaller weights from the multiplications, i.e., the gradients propagated over many time steps tend to vanish. The problem is called the vanishing gradient problem.

3.5 Long Short-Term Memory Networks and Gated Recurrent Unit

Initially introduced by [18], the original LSTM aims to solve the vanishing gradient problem in the standard RNN by adding self-loops to produce paths where the gradient can flow for long durations [13]. The LSTM networks improved the RNN by introducing the LSTM cells which have a structure called gates where information flows and is carefully regulated [19]. The gates, which are essentially sigmoid neural networks and a pointwise multiplication operation, are called input gate, output gate, and forget gate.

In the term of a stock price prediction model, the LSTM model consists of a sequence of LSTM cells where the length of the sequence constitutes the time lags or time window in calendar days. For example, to predict the stock price of one eminent on June 15th, 2022, with time lags of 5 days, the LSTM sequence consists of 5 LSTM cells. Each cell corresponds to the stock price on June 14th, June 13th, June 12th, June 11th, and June 10th in the year 2022.

Each LSTM cell receives two inputs: x_{t} and $h_{\mathrm{t}-1}$, which denotes the price at day t and a value from the previous LSTM cell at day $(t-1)$, respectively. The two inputs are fed into a "forget gate layer" which outputs a number between 0 and 1 for each number in a cell state $c_{\mathrm{t}-1}$. The number denotes the probability that the model will keep the information or not. The next step constitutes the mechanism where the model decides what new information will be stored in the cell state c_{t}. The new information in c_{t} is then used to update the old state $c_{\mathrm{t}-1}$. Finally, the last step of the computation is to decide what information to output or pass along to the next LSTM cell. This step uses the new information in the cell state c_{t} and the output from the forget gate layer to decide which information to keep and discard.

The LSTM model can predict the stock price using the above computations by fine-tuning or learning the pattern from our historical stock price datasets. The learning step is the model that tries to update its parameters to match the intended output when given specific inputs. However, the technique to update the parameter is not arbitrary, or rather, it uses an optimization strategy called backpropagation [20].

There are some other variants of LSTM such as the gated recurrent unit or GRU. It is a simpler variant of LSTM introduced by [21] which combines, among some other changes, the forget and input gates into a single "update gate" [19]. See Fig. 3 which illustrates the differences between LSTM and GRU. In the diagram of LSTM, i, f, and o are the input, forget, and output gates, respectively. Meanwhile, the memory cell and the new memory cell content are denoted by c and $\tilde{c}$, respectively. The r, z, h, and $\tilde{h}$ in the GRU are the reset gate, update gate, activation, and candidate activation, respectively. In [22] and [23], the authors demonstrate that there are no other variants that can improve upon both LSTM and GRU across a wide range of tasks [13].

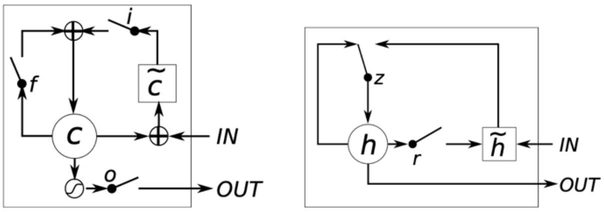

Fig. 3. Illustration of LSTM (left) and GRU (right) from [24]

3.6 Evaluation Metrics

To evaluate the performance of the deep learning models on the stock price prediction task, we use the root mean squared error (RMSE), the mean absolute percentage error (MAPE), and the coefficient of determination as the evaluation metrics. The RMSE, as defined in Equation (1), is the square root of mean squared error (MSE) which is the average squared difference between the actual value (Y) and the predicted value ($\hat{Y}_i$) [25]. The MSE is also used mainly as the loss function when we train deep learning models.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)} \quad (1)$$

Meanwhile, the MAPE is also another commonly used evaluation metric for regression problems. It measures the accuracy of the model by calculating the average absolute percentage error for each period as in Equation (2).

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \quad (2)$$

The coefficient of determination R^2, is a measure of the goodness of fit of a model or how well the model approximates the real data points [26]. The formula to calculate the R^2 can be seen in Equation (3).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)}{\sum_{i=1}^{n}(Y_i - \bar{Y})} \quad (3)$$

4 Finding and Discussion

In this study, the deep learning models are implemented using Tensorflow library in Python programming language. Meanwhile, the data preprocessing and model evaluation are computed using the sklearn library. From 10 experiments, we find that training the model with Adam [27], learning rate 0.01, and time lags of 10 days gives the best performance. We use 20% of the training set as the validation set for early-stop training as well.

We employ simple networks using only one LSTM and GRU layer with 8 hidden units. Consecutively, the output of the LSTM and GRU layers are fed into a fully connected layer as the model output. With these relatively small networks, the model already returns good performance only in 10 iterations, as can be seen from the learning curves in Fig. 4 and Fig. 5.

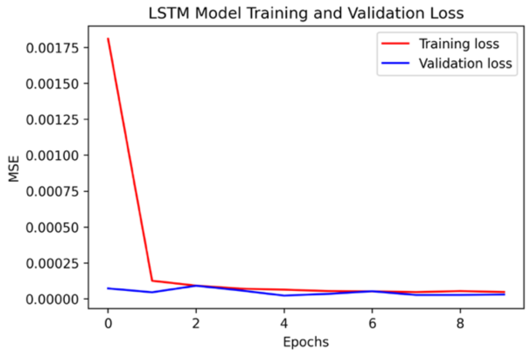

Fig. 4. The training result of the LSTM model

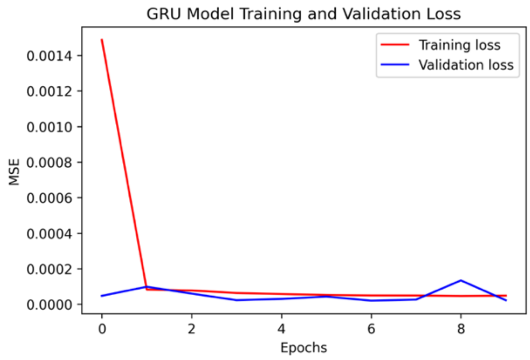

Fig. 5. The training result of the GRU model

After obtaining the best LSTM and GRU model from the training step, i.e., choosing the model with the smallest MSE, we evaluate the model by calculating the RMSE, MAPE, and R^2 from the test data. As can be seen in Table 1, the GRU model has the highest R^2 score of 0.99907 and the lowest RMSE of 0.005601. Meanwhile, the LSTM model has the lowest MAPE of 0.001427.

Table 1. Performance comparison of the LSTM and GRU model using the test data

	LSTM	GRU
RMSE	5.844×10^{-3}	5.601×10^{-3}
MAPE	1.427×10^{-2}	1.594×10^{-2}
R^2	0.99898	0.99907

Additionally, we evaluate the model further by testing it with unseen data, which is the stock prices from January 1st, 2023, to March 31st, 2023, of PT Adaro Energy Indonesia Tbk (ADRO.JK). Fig. 6 and Fig. 7 show the comparison between the real stock prices and the model predictions. Despite the good performance in the training, the GRU only achieve an R^2 of 0.66885 and the LSTM achieved an R^2 of 0.38756 of these unseen data. These results still show that GRU can predict stock prices more accurately than LSTM.

Fig. 6. Comparison of the ground truth and the stock price prediction using LSTM

Fig. 7. Comparison of the ground truth and the stock price prediction using GRU

5 Conclusion

This study introduces the usage of deep learning models to forecast the stock prices in the LQ45 of the Indonesia Stock Exchange. The LQ45 consists of 45 companies with high transactional values, high market capitalization, and good financial conditions. The deep learning models discussed in this study are the long short-term memory (LSTM) model and the gated recurrent unit (GRU). Using the historical stock prices up until 2022, we train and validate the LSTM and GRU models. In addition to that, we evaluate the model further using data up until March 31st, 2023. From the experiments, we find that, although the performance of both models is not too different, the GRU model is mostly better than the LSTM model. Having said that, the experiments can be further be generalized in the future studies to other companies in the Indonesian Stock Exchange with varying financial conditions.

References

[1] Hanke, J. E., & Wichern, D. W. (2005). *Business forecasting*. Pearson Educación: Chicago.

[2] Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Fransisco: Holden-Day.

[3] Kenny, G., Meyler, A., & Quinn, T. (1998). *Forecasting Irish inflation using ARIMA*

models (No. 3/RT/98). Central Bank of Ireland.

[4] Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric reviews*, *29*(5-6), 594-621.

[5] Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data*, *4*(1), 15.

[6] Jakubik, J., Nazemi, A., Geyer-Schulz, A., & Fabozzi, F. (2022). Incorporating financial news for forecasting Bitcoin prices based on long-short term memory networks. *Quantitative Finance*, *23*(2), 335-349.

[7] Xiao, D. & Su, J. (2022). Research on stock price time series prediction based on deep learning and autoregressive integrated moving average. *Scientific Programming, 2022*. https://doi.org/10.1155/2022/4758698

[8] Staffini, A. (2022). Stock price forecasting by a deep convolutional generative adversarial network. *Frontiers in Artificial Intelligence, 5*. https://doi.org/10.3389/frai.2022.837596

[9] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv*. 1511.06434.

[10] Hansun, S. & Young, J. C. (2021). Predicting LQ45 financial sector indices using RNN-LSTM. *Journal of Big Data, 8*(104). https://doi.org/10.1186/s40537-021-00495-x

[11] Yu, Y., Zhu, Y., Li, S., & Wan, D. (2014). Time Series Outlier Detection Based on Sliding Window Prediction. *Mathematical Problems in Engineering, 2014*. https://doi.org/10.1155/2014/879736

[12] Brownlee, J. (2020, August 15). Time series forecasting a supervised learning. *Machine Learning Mastery*. https://machinelearningmastery.com/time-series-forecasting-supervised-learning/

[13] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[14] Auli, M., Galley, M., Quirk, C., & Zweig, G. (2013, October). Joint language and translation modeling with recurrent neural networks. *In Proc. of EMNLP*.

[15] Shini, R. S., & Kumar, V. A. (2021, January). Recurrent Neural Network based Text Summarization Techniques by Word Sequence Generation. *In 2021 6th International Conference on Inventive Computation Technologies (ICICT),* (pp. 1224-1229). IEEE.

[16] Jansen, Herwindiati, D. E., Hendryli, J. (2020). Chinese audio transcription using connectionist temporal classification. *In 2020 International Conference on Computer and Communications Management*, (pp. 92-96).

[17] Bella, Hendryli, J., Herwindiati, D. E. (2020). Voice authentication model for one-time password using deep learning models. *In 2020 International Conference on Big Data Engineering and Technology*, (pp. 35-39).

[18] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780.

[19] Olah, C. (2015, August 27). *Understanding LSTM networks*. Colah's Blog. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[20] Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, *21*(6), 335-346.

[21] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the

properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv*. 1409.1259.
[22] Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015, June). An empirical exploration of recurrent network architectures. *In International conference on machine learning*, (pp. 2342-2350). PMLR.
[23] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28*(10), 2222-2232.
[24] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv*. 1412.3555.
[25] Bickel, P. J., & Doksum, K. A. (2015). Mathematical statistics: basic ideas and selected topics, volumes I-II package. *Chapman and Hall/CRC*.
[26] Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.
[27] Zhang, Z. (2018, June). Improved adam optimizer for deep neural networks. *In 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS),* (pp. 1-2). IEEE.

Notes on contributors

Dyah Erny Herwindiati is Professor at Universitas Tarumanagara, Jakarta, Indonesia. The minimum Vector Variance is a robust estimator she proposed in 2006. Her current research is focused on applying robust algorithms for land cover mapping of remote sensing data, data mining, and machine learning.

Janson Hendryli is lecturer at the Department of Informatics, Faculty of Information Technology, Universitas Tarumanagara of Jakarta, Indonesia. His main teaching and research interests include software development, machine learning, and deep learning.

Nor Haniza Sarmin is a Professor of Mathematics at the Faculty of Science, Universiti Teknologi Malaysia (UTM) Johor Bahru, Johor, Malaysia. Her main teaching and research interests include group theory, graph theory, and splicing systems. She has published more than 500 research articles in national and international journals and proceedings