❑   2696

Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor

Al-Khowarizmi[1], Rahmad Syah[2], Mahyuddin K. M. Nasution[3], Marischa Elveny[4]
[1]Department of Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
[2,3,4]Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

Article Info

ABSTRACT

Forecasting involves all areas in predicting future events. Many problems can be solved by using a forecasting approach to become a study in the field of data science. Forecasting that learns through data in the light age is able to solve problems with large-scale data or big data. With the big data, the performance of the k-nearest neighbor (k-NN) method can be tested with several accuracy measurements. Generally, accuracy measurement uses MAPE so it is necessary to conduct sensitivity on MAPE by combining it with the detection rate which is the difference technique. In addition, the k-NN process has been developed for the sake of running sensitivity by performing normalized distance using normalized Euclidean distance so that in this paper using the crude palm oil (CPO) price dataset, it is able to forecast and become a future model and apply it to Business Intelligence and analysis. In the final stage of this paper, the accuracy value in doing big data forecasting on CPO prices with MAPE is 0.013526% and MAPE sensitivity combined with a detection rate of 0.000361% so that future processes using different methods need to involve detection rates.
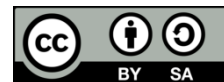
Corresponding Author:

Marischa Elveny
Department of Information Technology
Universitas Sumatera Utara,
Dr. Masyur Road No. 9 Medan, Indonesia
Email: marischaelveny@usu.ac.id

1. INTRODUCTION

Forecasting is a technique of data mining combined with machine learning that is used to analyze and calculate future events by using reference to past data with qualitative and quantitative approaches [1-3]. Forecasting has the objective of estimating prospects for economic progress as well as business activities and the environmental impact on these prospects [4, 5]. So that forecasting for every company or business organization enters the most important part in every management decision making and forecasting itself can be the basis for management to plan activities and decisions in the short, medium and long term [6, 7]. In conducting the forecasting process, a method is needed in its application, namely the K-nearest neighbor method [8, 9].

k-NN is a method for classifying and forecasting objects based on learning data that is closest to the data object carried out by training data [10-12]. k-NN is very often found in papers that do forecasting such as reference [13] doing big data forecasting research, which is to predict customers who will be in arrears in credit payments to a bank and the result is a time measurement when the program carries out the forecasting process. In doing forecasting, of course, a learning technique is needed to train the past [14]. Learning past data can be trained using a method [15, 16]. One of the methods of forecasting is k-NN [17, 18]. In carrying

out training data, you must have large data because of the forecasting process so that large data is needed in forecasting to get as few errors as possible in it [19].

As in Al-Khowarizmi [20] predicting the price of crude palm oil (CPO) and turning it into business intelligence in order to cut processing operating costs are so high that the method used by reference in making predictions is the Simple Evolving Connectionist System method. CPO is a product that is transacted in the commodity market or natural resource product which is commonly processed into various derivative products, both in the form of consumer goods and industrial raw materials [21-23]. So that it is very suitable that the CPO data collected is used as big data to be forecasted.

In obtaining a small error value, of course, a technique of measuring accuracy must be carried out as is done by Lubis [24] to analyse bank customer data and measure the accuracy value looking for the smallest error with MAPE and MSE so that in his research the smallest error value was achieved using MAPE. In addition, reference [25] also applies the smallest error value accuracy in the Simple Evolving Connectionist System method using MAPE but is influenced by the normalized distance formula.

However, Prayudani [26] compared the smallest error value with MAPE, MAE and MSE where the MAPE and MAE results have the same value except that the difference is in units using percent (%). In addition, by getting the value of accuracy and the smallest error value, it will increase productivity in business intelligence [20]. So that the application of data science has been used in various fields of science such as in this paper to measure accuracy to get the smallest error value that will be used to do big data forecasting on CPO prices using the KNN method so that it can be used as a model in a business intelligence and analysis (BIA).

2.   RESEARCH METHOD
2.1. Dataset
The dataset in this paper uses a big data pattern, namely getting CPO price, the dataset is obtained from www.investing.com based on time and world CPO prices from January 1st 2010 to August 25th 2020 in the forecasting process involving training data and data testing. Where is the training data from January 1, 2010 to December 30th 2019 and testing data from January 1st 2020 to August 25th 2020.

2.2. General architecture
In this paper, so that the research does not expand, a general architecture is described in this paper as shown in Figure 1.
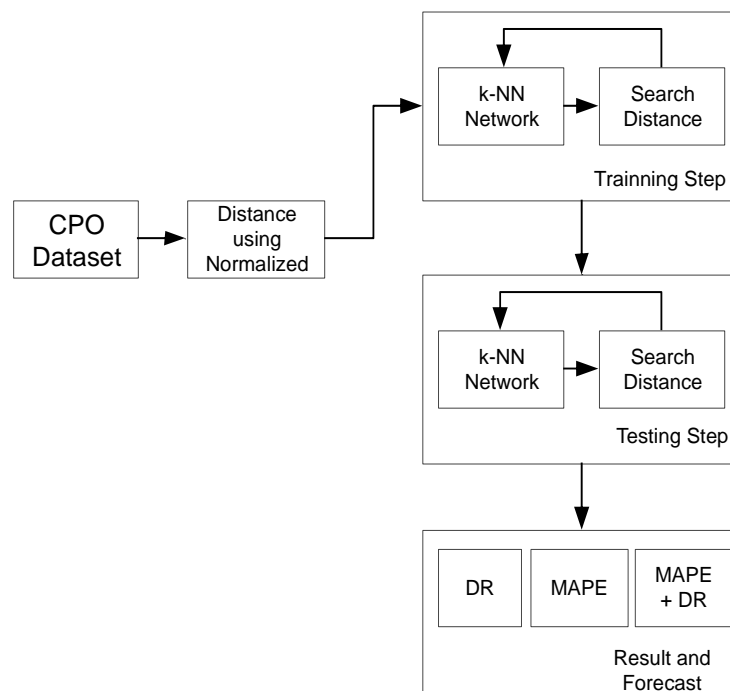

Figure 1. General architecture

In general architecture, it is described in the steps:
a. Putting the dataset on storage.
b. Perform calculations with the normalized Euclidean formula
c. Training data
   - Input data into the k-NN network by calculating the formula
   - Looking for neighbours with the formula assignment.
d. Testing Data
   - Input data into the k-NN network by calculating the formula
   - Looking for neighbours with the formula assignment.
e. Analyse the results by obtaining sensitivity in MAPE with the combination of Detection Rate to obtain the results of big data forecasting with CPO price dataset.

2.3. Development of k-NN distance formulas

K-NN is a machine learning method because it involves past data to predict future data. The k-NN process is not only for forecasting but for predicting classification and association in finding patterns [27]. K-NN is the result of advanced improvisation Nearest Neighbour classification techniques. This concludes because each new data can be trained by many k neighbours, where k is a positive integer, with a small amount of data [28]. Forecasting uses k-NN at its stage of conducting training data and testing data based on categories in the sample or on past data and in accordance with the training sample k which is the closest neighbour to the sample test, then entering into the category that has the largest probability category [29]. Near or far the distance of a point from its neighbour can be calculated using the Euclidean distance [30]. Euclidean distance is represented in (1):

$$D_n = \sqrt{\sum_i^K \frac{|I_i - W_i|^2}{k}} \qquad (1)$$

where:
K   is number of input nodes
I   is initial node value
W   is value of the destination node

D_n is the distance between the starting point where it will move which is a point that is known to its class and new point to be addressed. The distance between the starting point and the training point is calculated and the closest k point is taken and if the new point or the target point is predicted to enter the class with the most forecasting of these points [31].

2.4. Sensitivity accuracy

Calculation of accuracy is one of the important things in pattern recognition. This process is carried out as a measure of evaluation in a system. Measuring the level of accuracy can use various ways, one of which is using the Detection Rate. The detection rate (DR) equation is represented in (2) [32]:

$$DR = \frac{a}{a+b} \qquad (2)$$

where:
a is actual data
b is result data

Then in this paper also measured by mean absolute percent error (MAPE). The (3) of MAPE is as [25, 33]:

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{a-b}{a} \right|}{n} \times 100\% \qquad (3)$$

where:
a = actual data
b = result data
n = lots of data

In addition to the Detection rate and MAPE, accuracy measurements are carried out by combining MAPE and Detection rate to get a good sensitivity value with (4).

$$Sensitivity\ MAPE = \frac{\sum_{t=1}^{n}|DR|}{n} \times 100\% \qquad (4)$$

where the difference value in getting the sensitivity MAPE value is combined with the detection rate.

3. RESULTS AND DISCUSSION

In this stage, forecasting is carried out because the forecasting process is needed to plan future activities, therefore it must consider and decide what forecasting is actually needed. Forecasting is a process that is not difficult, but the mistake of deciding the goal or what is desired will result in different results so that the accuracy of forecasting will be doubted. Each process also needs to determine every detail of the forecast in terms of weeks, months or years. In forecasting, identify what data is needed and what data is available. Identifying this data will have an impact on the choice of forecasting methods later.

After getting the data to be evaluated like a dataset, the next step is to choose and determine the right model or forecasting method. Generally, the method chosen for forecasting is a method that has considered factors such as cost and ease of use. In addition, one of the most important factors is the accuracy of forecasting. The most common way is to find the best two or three methods and then test them on historical data to see which forecasting method or model is the most accurate.

In this paper, big data forecasting is carried out on CPO prices using the k-NN method and obtaining an accurate sensitivity value on the accuracy measurement technique with MAPE in combination with the detection rate. Before doing data training and testing, first determine what data will do the training and data testing. The dataset used from January 1, 2010 to August 25, 2020 in the forecasting process involves training data and data testing. Where is the training data from January 1, 2010 to December 30, 2019 and testing data from January 1, 2020 to August 25. The dataset is summarized in Table 1.

Table 1. Dataset

No	Date	Price	No	Date	Price
1	Aug 24 2020	755.5	23	Jul 23 2020	739.3
2	Aug 21 2020	758.4	24	Jul 22 2020	730.1
3	Aug 20 2020	752.6	--	....	....
4	Aug 19 2020	762.8	--	....	....
5	Aug 18 2020	757.1	--	....	....
6	Aug 17 2020	754.9	2349	Dec 31 2011	546.5
7	Aug 14 2020	745.4	2350	Dec 29 2011	541.6
8	Aug 13 2020	741.9	2351	Dec 28 2011	542
9	Aug 12 2020	741.4	2352	Dec 27 2011	536.6
10	Aug 11 2020	727.9	2353	Dec 26 2011	537.9
11	Aug 10 2020	741.8	2354	Dec 24 2011	532.5
12	Aug 07 2020	739.7	2355	Dec 23 2011	529.2
13	Aug 06 2020	752.1	2356	Dec 22 2011	516
14	Aug 05 2020	740.9	2357	Dec 21 2011	513.4
15	Aug 04 2020	743.4	2358	Dec 20 2011	511.4
16	Aug 03 2020	740.8	2359	Dec 19 2011	512.7
17	Jul 31 2020	739.7	2360	Dec 17 2011	511
18	Jul 30 2020	739	2361	Dec 16 2011	505.8
19	Jul 29 2020	736.9	2362	Dec 15 2011	509.1
20	Jul 28 2020	732.3	2363	Dec 14 2011	518.5
21	Jul 27 2020	740.4	2364	Dec 13 2011	511.8
22	Jul 24 2020	743.7	2365	Dec 12 2011	507.3

Table 1 is a summarized dataset, data of 2365 rows and as a whole can be seen in the graph available in Figure 2. Figure 2 is the entire dataset. Next, do the forecasting process with the k-NN method where the first process of forecasting determines the value of k using the normalized Euclidean distance as in (1). The forecasting process is carried out based on 165 testing data and 2,200 training data with detailed training from 1 January 2010 to 30 December 2019 and testing data from 1 January 2020 to 25 August. Forecasting is an ongoing process at this stage. After making the forecast, one must record what actually happened (actual) and then use that information to measure the accuracy of the forecast and get the sensitivity of the accuracy. It should be noted that the best forecasting method in the past is not necessarily the best result for the future. Therefore, it is necessary to change the forecasting method along with changes in the data to be forecast. The forecasting results in this paper go well and can be seen in Table 2.

Figure 2. Dataset of CPO price

Table 2. Forecasting result and sensitivity

Date	Price	Forecasting	Error	DR
Aug 24 2020	755.5	757.325	0.0024156	0.0003301
Aug 21 2020	758.4	757.725	0.00089	0.0003299
Aug 20 2020	752.6	756.85	0.0056471	0.0003303
Aug 19 2020	762.8	755.05	0.0101599	0.0003311
Aug 18 2020	757.1	749.825	0.009609	0.0003334
Aug 17 2020	754.9	745.9	0.0119221	0.0003352
Aug 14 2020	745.4	739.15	0.0083848	0.0003382
Aug 13 2020	741.9	738.25	0.0049198	0.0003386
Aug 12 2020	741.4	737.7	0.0049906	0.0003389
Aug 11 2020	727.9	740.375	0.0171383	0.0003376
Aug 10 2020	741.8	743.625	0.0024602	0.0003362
Aug 07 2020	739.7	744.025	0.005847	0.000336
Aug 06 2020	752.1	744.3	0.010371	0.0003359
Aug 05 2020	740.9	741.2	0.0004049	0.0003373
Aug 04 2020	743.4	740.725	0.0035983	0.0003375
Aug 03 2020	740.8	739.1	0.0022948	0.0003382
Jul 31 2020	739.7	736.975	0.0036839	0.0003392
Jul 30 2020	739	737.15	0.0025034	0.0003391
Jul 29 2020	736.9	738.325	0.0019338	0.0003386
Jul 28 2020	732.3	738.925	0.0090468	0.0003383
Jul 27 2020	740.4	738.375	0.002735	0.0003386
Jul 24 2020	743.7	734.275	0.0126731	0.0003405
Jul 23 2020	739.3	726.8	0.0169079	0.0003439
Jul 22 2020	730.1	718.85	0.0154088	0.0003478
Jul 21 2020	724	711.35	0.0174724	0.0003514
Jul 20 2020	713.8	705	0.0123284	0.0003546
Jul 17 2020	707.5	700.1	0.0104594	0.0003571
Jul 16 2020	700.1	693.575	0.0093201	0.0003604
Jul 15 2020	698.6	689.125	0.0135628	0.0003628
Jul 14 2020	694.2	682.7	0.0165658	0.0003662
Jul 13 2020	681.4	677.25	0.0060904	0.0003691
Jul 10 2020	682.3	673.725	0.0125678	0.0003711
Jul 09 2020	672.9	671.5	0.0020805	0.0003723
Jul 08 2020	672.4	668.875	0.0052424	0.0003738
Jul 07 2020	667,3	666.9	0.0005994	0.0003749
...	...	...	...	...
...	...	...	...	...
Jan 15 2020	814	826.625	0.0155098	0.0003024
Jan 14 2020	823.6	830.65	0.00856	0.000301
Jan 13 2020	831.3	830.85	0.0005413	0.0003009
Jan 10 2020	837.6	825.65	0.014267	0.0003028
Jan 09 2020	830.1	815.575	0.0174979	0.0003065
Jan 08 2020	824.4	811.45	0.0157084	0.0003081
Jan 07 2020	810.5	811.275	0.0009562	0.0003082
Jan 06 2020	797.3	811	0.017183	0.0003082
Jan 03 2020	813.6	812.5	0.001352	0.0003077
Jan 02 2020	823.7	810.625	0.0158735	0.0003084
Jan 01 2020	809.4	805.7	0.0045713	0.0003103

Table 2 shows the summary results of forecasting using k-NN where the process of determining the neighbor's distance uses the normalized Euclidean distance and the calculation of the error value and the detection rate value can be seen. The detection rate value is smaller than the error value commonly used in the calculation of accuracy. For more details, the summary of Table 2 can be seen in the graph provided in Figure 3.
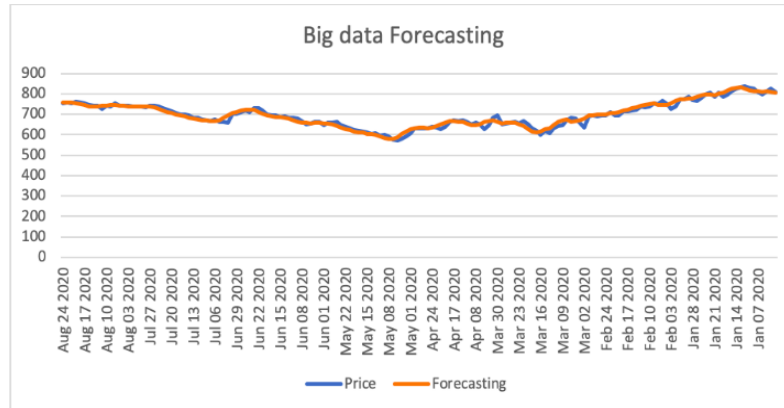

Figure 3. CPO price forecasting

Figure 3 shows a complete graph of the forecasting results where the blue line is the actual data and the orange line is the forecasting result. Forecasting results run perfectly due to the influence of the normalized Euclidean distance. So that it is in line with the objectives of this paper in obtaining sensitivity to accuracy measurements using MAPE and using MAPE in combination with detection rates. Where the MAPE results can be calculated based on (3).

$$MAPE = \frac{2.231757}{165} \times 100\% = 0.013526\ \%$$

MAPE results of 0.013526% show good accuracy results, but testing is also carried out to get the MAPE Sensitivity value combined with the detection rate. Where the detection rate calculation uses (2) and the sensitivity results use (4).

$$Sensitivity\ MAPE = \frac{0.059544}{165} \times 100\% = 0.000361\ \%$$

The result of MAPE sensitivity looks smaller at 0.000361%. This further shows that the MAPE sensitivity combined with the detection rate obtain a very good accuracy value. So that MAPE sensitivity results have been found and can be tested on other methods.

4. CONCLUSION

In the end, this paper has found a sensitivity value in the accuracy measurement technique using MAPE combined with the detection rate and has been tested to perform big data forecasting on the k-NN method. The forecasting process is carried out with 2365 datasets with training data of 2200 and testing data of 165. The results of forecasting are going well so that this paper can develop towards data science as data processing to get very good accuracy values. As in the aim of this paper, to get a good sensitivity value in accuracy using MAPE and accuracy using MAPE combined with detection rates where the results received in forecasting CPO prices at k-NN, the MAPE value is 0.013526% while MAPE is combined with detection. rate of 0.000361% so that the best sensitivity value in forecasting using MAPE combined with detection rate.

REFERENCES

[1] A. Ji, Y. Qiao, and H. Qiu, "Combination Forecasting of Fuzzy Forecast," *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, Tianjin, 2009, pp. 262-266.

[2] H. Shafizadeh-Moghadam, R. Valavi, H. Shahabi, K. Chapi, and A. Shirzadi, "Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping," *Journal of Environmental Management*, vol. 217, pp. 1-11, 2018.

[3] C. Voyant et al., "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569-582, 2017.

[4] C. Zhang, Y.-X. Tian, Z.-P. Fan, Y. Liu, and L.-W. Fan, "Product sales forecasting using macroeconomic indicators and online reviews: a method combining prospect theory and sentiment analysis," *Soft Computing*, vol. 24, no. 9, pp. 6213-6226, 2020.

[5] M. H. Park, J. S. Lee, and I. C. Doo, "A Study of the Demand Forecasting Model for Publishing Business using Business Analysis," *International Journal of Computing and Digital Systems*, vol. 9, no. 5, pp. 801-812, 2020.

[6] P. Savioz and M. Blum, "Strategic forecast tool for SMEs: how the opportunity landscape interacts with business strategy to anticipate technological trends," *Technovation*, vol. 22, no. 2, pp. 91-100, 2002.

[7] N. Hemayatkar, K. Khalili-Damghani, H. Didehkhani, and R. Samiee, "Developing a fuzzy inference system to devise proper business strategies: a study on carpet industry," *Journal of Industrial Engineering International*, vol. 15, no. 3, pp. 529-544, 2019.

[8] M. Kim, Y. Kim, H. Kim, W. Piao, and C. Kim, "Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant," *Frontiers of Environmental Science & Engineering*, vol. 10, no. 2, pp. 299-310, 2016.

[9] M. Hammad, A. Alqaddoumi, H. Al-Obaidy, and K. Almseidein, "Predicting software faults based on k-nearest neighbors classification," *International Journal of Computing and Digital Systems*, vol. 8, no. 5, pp. 461-467, 2019.

[10] Y. K. Zamil, S. A. Ali, and M. A. Naser, "Spam image email filtering using K-NN and SVM," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 1, pp. 245-254, 2019.

[11] S. B. Jadhav, V. R. Udupi, and S. B. Patil, "Soybean leaf disease detection and severity measurement using multiclass SVM and KNN classifier," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 4092-4098, 2019.

[12] E. G. Nihad, E. N. El Mokhtar, Z. Abdelhamid, and A. A. Mohammed, "Hybrid approach of the fuzzy C-means and the K-nearest neighbors methods during the retrieve phase of dynamic case based reasoning for personalized follow-up of learners in real time," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, pp. 4939-4950, 2019.

[13] A. R. Lubis, M. Lubis, Al-Khowarizmi, and D. Listriani, "Big Data Forecasting Applied Nearest Neighbor Method," *2019 International Conference on Sustainable Engineering and Creative Computing (ICSECC)*, Bandung, Indonesia, 2019, pp. 116-120.

[14] G. P. Zhang, B. E. Patuwo, and M. Y. Hu, "A simulation study of artificial neural networks for nonlinear time-series forecasting," *Computers & Operations Research*, vol. 28, no. 4, pp. 381-396, 2001

[15] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 133-142.

[16] J. Wu, S. Guo, J. Li, and D. Zeng, "Big Data Meet Green Challenges: Greening Big Data," *IEEE Systems Journal*, vol. 10, no. 3, pp. 873-887, 2016.

[17] F. H. Al-Qahtani and S. F. Crone, "Multivariate k-nearest neighbour regression for time series data-A novel algorithm for forecasting UK electricity demand," *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, 2013, pp. 1-8.

[18] S. Taneja, C. Gupta, S. Aggarwal, and V. Jindal, "MFZ-KNN-A modified fuzzy based K nearest neighbor algorithm," *2015 International Conference on Cognitive Computing and Information Processing (CCIP)*, Noida, 2015, pp. 1-5.

[19] N. Huyghues-Beaufond, S. Tindemans, P. Falugi, M. Sun, and G. Strbac, "Robust and automatic data cleansing method for short-term load forecasting of distribution feeders," *Applied Energy*, vol. 261, p. 114405, 2020.

[20] Al-Khowarizmi, I. R. Nasution, M. Lubis, and A. R. Lubis, "The effect of a secos in crude palm oil forecasting to improve business intelligence," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 4, pp. 1604-1611, 2020.

[21] I. Goh and Y. Shen, "Forecasting Crude Palm Oil Price Movement through Dynamic Sampling Band based on Historical Price Distribution," Swinburne University of Technology, 2017.

[22] A. A. Karia and I. Bujang, "Progress accuracy of CPO price prediction: Evidence from ARMA family and artificial neural network approach," *International Research Journal of Finance and Economics*, vol. 64, pp. 66-79, 2011.

[23] Swati Yeole, "Magna Carta : A 800 Years Journey," *From Unfettering Chains to Free Flight*, pp. 1-378, 2016.

[24] A. R. Lubis, M. Lubis, and A. Khowarizmi, "Optimization of distance formula in K-Nearest Neighbor method," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 1, pp. 326-338, 2020.

[25] A. Al-Khowarizmi, O. S. Sitompul, S. Suherman, and E. B. Nababan, "Measuring the Accuracy of Simple Evolving Connectionist System with Varying Distance Formulas," in *Journal of Physics: Conference Series*, vol. 930, no. 1, 2017.

[26] S. Prayudani, A. Hizriadi, Y. Y. Lase, Y. Fatmi, and Al-Khowarizmi, "Analysis Accuracy of Forecasting Measurement Technique on Random K-Nearest Neighbor (RKNN) Using MAPE and MSE," *Journal of Physics: Conference Series*, vol. 1361, no. 1, pp. 0-8, 2019.

[27] J. S. Raikwal and K. Saxena, "Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set," *International Journal of Computer Applications*, vol. 50, no. 14, pp. 35-39, 2012.

[28] C. K. Jaya, R. Sunitha, and A. T. Mathew, "Security prediction of high voltage transmission system (HVTS) based on k-nearest neighbor (k - NN) classifier technique," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 6, pp. 5773-5782, 2019.

[29] R. S. V Teegavarapu, "Missing precipitation data estimation using optimal proximity metric-based imputation, nearest-neighbour classification and cluster-based interpolation methods," *Hydrological Sciences Journal*, vol. 59, no. 11, pp. 2009-2026, 2014.

[30] J. Amores, N. Sebe, and P. Radeva, "Boosting the distance estimation: Application to the K-Nearest Neighbor Classifier," *Pattern Recognition Letters*, vol. 27, no. 3, pp. 201-209, 2006.

[31] Al-Khowarizmi, "Modified of Simple Evolving Connectionist System using Distance Formulas," Universitas Sumatera Utara, 2017.

[32] M. Ballester, et al., "Detection rate and diagnostic accuracy of sentinel-node biopsy in early stage endometrial cancer: a prospective multicentre study (SENTI-ENDO)," *The Lancet Oncology*, vol. 12, no. 5, pp. 469-476, 2011.

[33] R. F. Rahmat, A. Rizki, A. F. Alharthi, and R. Budiarto, "Big data forecasting using evolving multi-layer perceptron," *2016 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT)*, Riyadh, 2016, pp. 1-5.

BIOGRAPHIES OF AUTHORS

Al-Khowarizmi was born in Medan, Indonesia, in 1992. He is a lecturer in Department of Information Technology-Faculty of Computer Science and Information Technology at Universitas Muhammadiyah Sumatera Utara (UMSU), Medan Indonesia. He got master from University of Sumatera Utara (USU), Indonesia in 2017 and graduate from Universitas Harapan (UnHar) Medan Indonesia in 2014, both information system. His main research interest is data science, big data, machine learning, neural network, artificial intelligence and business intelligence.

Rahmad Syah from Doctoral program in Universitas Sumatera Utara, Medan Indonesia. M.Kom, Informatics, University of Sumatera Utara (USU) Medan. Researcher Field of study in Modelling and Computational System, Information Systems, Business Intelligence and Analytics (BIA), Artificial Intelligence, Data Science.

Mahyuddin K. M. Nasution, Professor from Universitas Sumatera Utara, Medan Indonesia. Mahyuddin K. M. Nasution was born in the village of Teluk Pulai Dalam, Labuhan Batu Regency, North Sumatera Province. Worked as a Lecturer at the University of Sumatera Utara, fields: Mathematics, Computer and Information Technology. Education: Drs. Mathematics (USU Medan, 1992); MIT, Computers and Information Technology (UKM Malaysia, 2003); PhD in Information Science (Malaysian UKM).

Marischa Elveny, from Universitas Sumatera Utara, Medan Indonesia. M.Kom, Computers Science (USU). Worked as a Lecturer at the Faculty of Computer Science and Information Technology. Research focus: Artificial Intelligence, Social Network, Information Technology, Business Intelligence and Analytics (BIA).