

Power Business Intelligence In The Data Science Visualization Process To Forecast CPO Prices

Albara¹, Al-Khowarizmi^{2*}, Riyan Pradesyah¹

¹ Department of Shariah Banking, Universitas Muhammadiyah Sumatera Utara, Indonesia.

² Department of Information Technology,
Universitas Muhammadiyah Sumatera Utara, Indonesia.

*Corresponding Author:

Email: alkhowarizmi@umsu.ac.id

Abstract.

Forecasting is one of the techniques in data mining by utilizing the data available in the data warehouse. With the development of science, forecasting techniques have also entered the computational field where the forecasting technique uses the artificial neural network (ANN) method. Where is the method for simple forecasting using the Time Series method. However, the ability to create data visualizations certainly hinders researchers from maximizing research results. Of course, with the development of the Power BI software, the data science process is more neatly presented in the form of visualization, where the data science process involves various fields so that in this paper the results of forecasting the price of crude palm oil (CPO) are presented for the development of the CPO business with the hope of implementing the Business Process. intelligence (BI) by involving ANN, namely the time series for forecasting. From the final results, accuracy in forecasting with time series involves 2 accuracy techniques, the first using MAPE and getting a result of 0.03214% and the second using MSE to get 962.91 results.

Keywords: Power BI, Forecasting, CPO and Time series.

I. INTRODUCTION

Data Science is an amalgamation of several sciences involving mathematics, statistics and computation which aim together to perform dataset analysis of big data sets [1][2][3]. The application of data science is applied with several certain algorithms to obtain data patterns and perform accurate forecasting so that it is useful to support decision making from business processes [4][5]. Where forecasting is one of the techniques in data mining that is useful for making knowledge of data to be able to predict future data with learning from previous datasets or data [6]. Like what Lubis [7] did forecasting big data using the nearest neighbour method, in his research, he got the optimal time for the 40000 credit transaction data in order to provide credit and get optimal results with 16 minutes to conduct training and gain new knowledge. In addition, Rahmat [8] conducted research using big data to predict gold prices using the multilayer perceptron evolving method, which is a development of the multilayer perceptron method by adding virtual nodes to avoid epoch repetitions and obtaining MAPE in forecasting accuracy of 0.769%. The forecasting process also aims to create a smart system by implementing machine learning in its final results [9]. Machine learning techniques are not much different from data mining techniques [10]. So that the application of forecasting involves some optimal algorithms in learning previous data [11]. In accordance with its application, the data science process is useful for simplifying business processes by predicting data based on previous data and past times so that time series algorithms can be used as the application of data mining as a predictor of business process sustainability [12].

Time series is an observation in an event, symptom or change that occurs from time to time with the benefit of obtaining patterns, forecasting and predictions in analyzing datasets in the past to obtain data in the future [13]. In business processes using time series has been tested in various business processes, such as Lubis [14] doing forecasting with time series based on changes in gold prices but to get accuracy combined with false alarm rates to get 0.466 results. In addition, Gao & Duru [15] in their research applied a mathematical modeling to ensure the value of the fuzzy model in the time series algorithm by obtaining out-of-sample accuracy values at a certain level in the fuzzy time series model which in conclusion explains hyper-parameters such as time lag and partitions consisting of the number of fuzzy sets, partition types and

membership functions with the fuzzy time series modeling approach are very influential using genetic algorithms (GA) as their implementation. However, there are so many business processes that can be applied, one of the suitable data developed in the advancement of data science is crude palm oil (CPO) [16]. Where CPO is a business that is available in the commodity market or in scientific natural data sources that are allowed to be processed into derivative products in the form of consumed goods and raw materials for industry. Agriculture, especially CPO, has been implemented in many computational techniques [17].

As was done by Al-Khowarizmi [18], forecasting CPO prices using the Simple Evolving Connectionist System method with the aim of changing existing business processes in plantations by applying computational techniques where the CPO sales process uses an agent in tendering CPO prices, so that operational costs are too high and implemented an online tender process with a forecasting value of CPO prices for the next 2 weeks because the CPO harvest process is about 2 weeks. However, the forecasting process gets an accuracy of 0.035% which is calculated using the MAPE formula. In addition, to get an accurate business process on CPO, Al-Khowarizmi [19] developed again by testing the accuracy of forecasting CPO prices on a different method, namely KNN by obtaining MAPE sensitivity which involves detection rates and obtaining more optimal results of 0.000361%. The two studies seem to have the same goal so that the application of business intelligence (BI) can be applied in the agricultural sector. In addition, BI can be relied on in business because BI focuses on systems and technology that can collect data from several sources that can be processed into a form of information for business needs [20]. So that reports from BI to be easily accepted by stakeholders are in the form of visualization that is used for data presentation with structural candidates with graphics to display information from hidden data [21]. As for its use in BI, software as a reference in this paper uses Power BI provided by Microsoft to make it easier to present business needs [22]. So that in this paper forecasting CPO prices using time series and the results are presented with Power BI to facilitate managerial or stakeholder in assisting decision making in the hope that data processing with data mining techniques becomes a reference in data science.

II. MATERIAL AND METHODS

Dataset

In this paper, forecasting using the time series method requires data for training and testing where the dataset in this paper is obtained from www.investing.com, namely CPO price data from January 2nd 2020 to December 30th 2020. From this data, 75% training was conducted. and testing by 20% consisting of 242 datasets.

Time Series Forecasting CPO

Time Series was first developed by Song and Chissom in 1993 [23]. Time Series is a data forecasting method that uses fuzzy principles as the basis. Roughly speaking, a fuzzy set can be interpreted as a class of numbers with cryptic limitations [24]. If universe of discourse (U) is the set of universes, $U = [u_1, u_2, \dots, u_p]$, then a fuzzy set A_i of U with the degree of membership is generally stated as follows: $A_i = \mu_{A_i}(u_1) / u_1 + \dots + \mu_{A_i}(u_p) / u_p$ where $\mu_{A_i}(u_i)$ is the degree of membership from u_i to A_i , where $\mu_{A_i}(u_i) \in [0, 1]$ and $1 \leq i \leq p$. The value of the degree of membership of the $\mu_{A_i}(u_i)$ is defined as follows [25]:

$$\mu_{A_i}(u_i) = \begin{cases} 1, & \text{if } i = j \\ 0.5, & \text{if } i = j - 1 \text{ or } j + 1 \\ 0, & \text{etc} \end{cases} \quad (1)$$

This can be illustrated by the following rules [26]:

Rule 1st : If the actual data X_t is included in u_i , then the degree of membership for u_i is 1, and $u_i + 1$ is 0.5 and if $u_i + 1$ are not, it means zero.

Rule 2nd : If the actual data X_t falls under u_i , $1 \leq i \leq p$ then the degree of membership for u_i is 1, for $u_i - 1$ and $u_i + 1$ is 0.5 and if it is not u_i , $u_i - 1$ and $u_i + 1$ are expressed zero.

Rule 3rd : If the actual data X_t is included in u_i , then the degree of membership for u_i is 1, and for $u_i - 1$ is 0.5 and if it is not u_i and $1 - 1$ it means zero.

Then in obtaining forecasting accuracy, it is done with 2 tests, the first uses MAPE (Mean Absolute Percentage Error) and the second uses MSE (Mean Squared Error). The MAPE is calculated by equation (2) below [27].

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{a-b}{a} \right|}{n} \times 100\% \quad (2)$$

Furthermore, it is also tested with MSE in equation (3) below [28].

$$MSE = \frac{1}{n} \sum_{i=1}^n (a - b)^2 \quad (3)$$

Where:

a = actual dataset

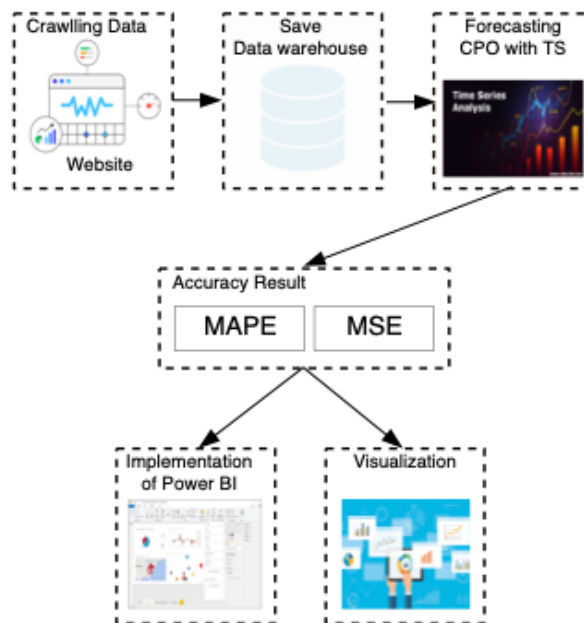
b = result dataset

n = lots of dataset.

General Architecture

This paper is carried out in accordance with the general arc.

Fig 1. General Architecture



Step in Figure 1 as follows:

1. Crawling data from www.investing.com
2. Save data in the data warehouse.
3. Forecasting CPO prices with time series.
4. Obtain accuracy values with MAPE and MSE.
5. Implementation in Power BI.
6. Displaying visualization and combination with python and datascience analysis.

III. RESULT AND DISCUSSION

At this stage, the data crawling process is carried out from www.investing.com in the form of CPO price data that will be forecasted. The data was taken from January 2nd 2020 to December 30th 2020. The data is shown in table 1 below.

Table 1. Dataset CPO Prices

NO	Date	Last One	The Highest	Lowest
1	30/12/20	1.225	1.270	1.215
2	29/12/20	1.245	1.295	1.240
3	28/12/20	1.280	1.335	1.240
4	23/12/20	1.250	1.260	1.125
5	22/12/20	1.175	1.270	1.135
6	21/12/20	1.220	1.300	1.060
7	18/12/20	1.065	1.090	900
8	17/12/20	900	900	850
9	16/12/20	860	885	860
10	15/12/20	860	880	850
11	14/12/20	870	910	870
12	11/12/20	895	975	845
13	10/12/20	855	880	840
14	08/12/20	865	895	855
15	07/12/20	885	940	880
16	04/12/20	875	900	855
...
...
238	08/01/20	800	805	785
239	07/01/20	805	825	805
240	06/01/20	825	835	810
241	03/01/20	830	840	820
242	02/01/20	825	850	810

From table 1 it can be seen that there is a date field, then there is the last price, which is the data on that date, it closes at 23.59 GMT+7. In addition, there is the highest price which is the maximum price on that date and the lowest price which is the minimum price from that date. Table 1 has the highest price, lowest price and closing price because on that date every second and minute the price changes. After crawling the data and making it a dataset, then the dataset is stored in the data warehouse so that forecasting can be done easily. However, the forecasting process in this paper consists of training and testing in which the dataset for training is 75% and testing is 25%. Forecasting is done using a time series algorithm which is implemented using python with a brief syntax as follows:

```
import warnings
warnings.filterwarnings('ignore')
import numpy as nphy
import pandas as pdas
import matplotlib.pyplot as mplt
import seaborn as sns
from dateutil.relativedelta import relativedelta
from scipy.optimize import minimize
import statsmodels.formula.api as smfapi
import statsmodels.tsa.api as smtapi
import statsmodels.api as smapi
import scipy.stats as scst
from itertools import product
from tqdm import tqdm_notebook
%matplotlib inline

ads = pdas.read_csv('../data/CPO.csv', index_col=['Time'], parse_dates=['Time'])
currency = pdas.read_csv('../data/FORECAST.csv', index_col=['Time'], parse_dates=['Time'])

mplt.figure(figsize=(100, 7))
mplt.plot(ads.Ads)
```

```

mplt.title('Dataset Forecasting')
mplt.grid(True)
mplt.show()

mplt.figure(figsize=(100, 7))
mplt.plot(ads.Ads)
mplt.title('Dataset Forecasting')
mplt.grid(True)
mplt.show()

mplt.figure(figsize=(15, 7))
mplt.plot(currency.GEMS_GEMS_SPENT)
mplt.grid(True)
mplt.show()

// MAPE and MSE Accuracy
from sklearn.metrics import r2_score, median_absolute_error, mean_absolute_error
from sklearn.metrics import median_absolute_error, mean_squared_error, mean_squared_log_error
def mean_absolute_percentage_error(y_true, y_pred):
    return nphy.mean(nphy.abs((y_true - y_pred) / y_true)) * 100

```

Above is a program syntax using Python tools with the hope that this syntax is combined with Power BI tools. So that by using this time series get the forecasting results as shown in Table 2 below.

Table 2. Forecasting CPO Price using Time Series

Date	Actual	Forecasting
29/09/20	820	851
30/09/20	800	806
01/10/20	805	811
02/10/20	805	757
05/10/20	805	822
06/10/20	800	835
07/10/20	800	794
08/10/20	805	833
09/10/20	800	850
12/10/20	805	834
13/10/20	835	837
14/10/20	840	811
15/10/20	830	863
16/10/20	835	824
19/10/20	815	801
20/10/20	810	860
21/10/20	800	753
22/10/20	800	840
23/10/20	800	787
26/10/20	810	827
27/10/20	815	795
02/11/20	800	754
03/11/20	815	811
04/11/20	810	825
05/11/20	835	876
06/11/20	865	820
09/11/20	850	824
10/11/20	875	886
11/11/20	875	886
12/11/20	850	809
13/11/20	860	848

16/11/20	860	862
17/11/20	840	890
18/11/20	805	791
19/11/20	810	832
20/11/20	795	809
23/11/20	800	838
24/11/20	770	734
25/11/20	770	733
26/11/20	735	780
27/11/20	700	674
30/11/20	690	650
01/12/20	780	762
02/12/20	780	741
03/12/20	915	927
04/12/20	875	923
07/12/20	885	894
08/12/20	865	840
10/12/20	855	892
11/12/20	895	891
14/12/20	870	830
15/12/20	860	903
16/12/20	860	871
17/12/20	900	871
18/12/20	1.065	1033
21/12/20	1.220	1207
22/12/20	1.175	1136
23/12/20	1.250	1258
28/12/20	1.280	1322
29/12/20	1.245	1281
30/12/20	1.225	1271

From table 2, it can be seen that the results of CPO price forecasting where 25% of the test dataset is based on 75% of the training dataset which is summarized from January 2, 2020 to September 28, 2020 is a training dataset and from September 29 2020 to December 30, 2020 is a test dataset. From the results of these tests, the calculation of accuracy is carried out which in this paper is tested on 2 accuracy techniques, the first using MAPE and the second using MSE. Where in calculating accuracy is used to get the percentage of errors in training. The results of the accuracy using MAPE are as follows:

$$MAPE = \frac{\left| \frac{a-b}{a} \right| + \left| \frac{a-b}{a} \right| + \left| \frac{a-b}{a} \right| + \left| \frac{a-b}{a} \right| + \dots + \left| \frac{a-b}{a} \right|}{60} \times 100\%$$

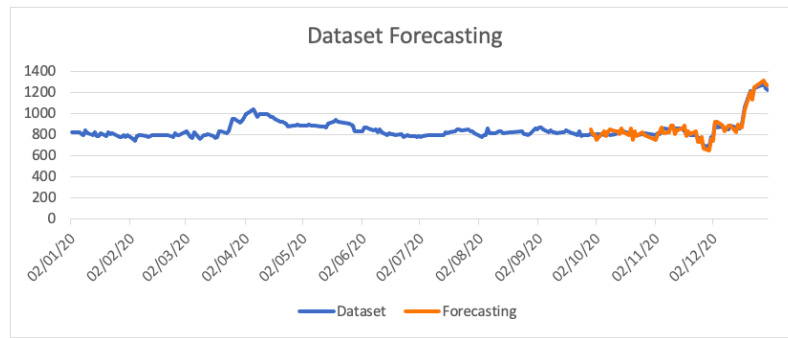
$$MAPE = 0.3214 \%$$

MAPE results in forecasting CPO prices using a time series algorithm of 0.3214%. Then the MAPE value is compared to MSE and get the results.

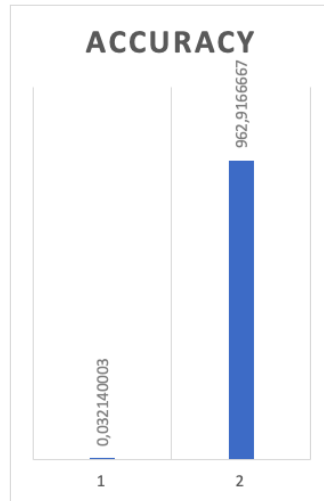
$$MSE = \frac{1}{6} (a+b)^2 + (a+b)^2 + (a+b)^2 + (a+b)^2 + \dots + (a+b)^2$$

$$MSE = 962.91$$

The MSE result is 962.91. so that the accuracy value both using MAPE and MSE is dynamic or can change based on the number of training and testing datasets in forecasting. From these results, data is provided using Power BI tools to get a presentation based on visualization where the forecasting results are shown in Figure 2.



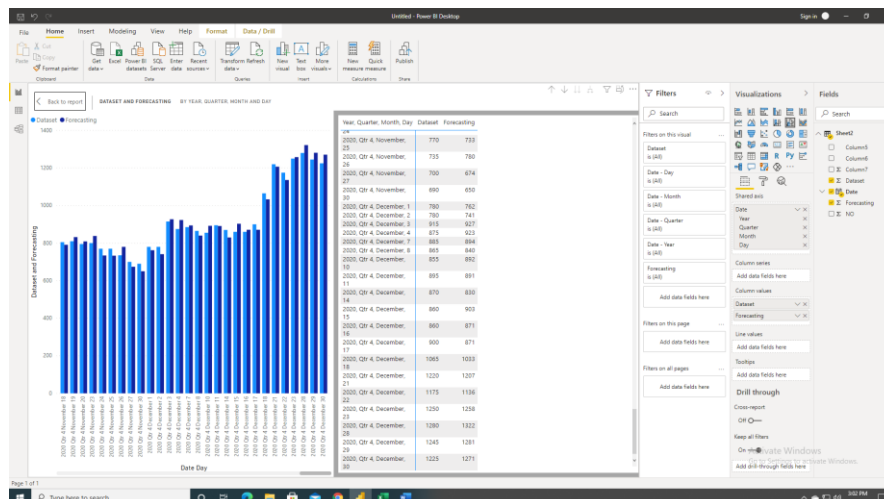
(a)



(b)

Fig 2. (a) Forecasting Visualization. (b) Accuracy Visualization

From Figure 2. It can be seen that the Power BI display is easier to accept for users because the user interface data display is more flexible, the information presented is more effective, can help make decisions more quickly and accurately and make it easy to control real time data on one screen. Thus, based on the final results of this paper, Power BI is an application that is capable of integrating with the Python syntax so that the Python syntax for forecasting with the time series algorithm as above can be stated in Power BI so that decision makers in the business can easily forecast using Power tools. BI only. Figure 3 shows the results of visualization of CPO price forecasting using the time series algorithm.

Fig 3. Python time series forecasting

With the process as shown in Figure 3, it makes it easier for application users, especially business policy makers, to forecast using the time series algorithm. However, in its application, it is necessary to test dynamic datasets such as multiplying the dataset in training so that the accuracy value gets very small errors so that the repetitive process can become a data science process and reference so that the data structure changes dynamically.

IV. CONCLUSIONS

In summary, this paper forecasts CPO prices using the time series algorithm and then contributes to computational techniques by making comparisons in obtaining accuracy values using the MAPE formula and the MSE formula. Where the forecasting consists of 242 datasets consisting of 75% training dataset and 25% testing dataset. From a total of 25% of the test dataset got a MAPE value of 0.03214% and an MSE of 962.61. The accuracy value can change according to the amount of training and testing in forecasting. Forecasting CPO prices using the time series algorithm is implemented with the Python application and in this paper the Python syntax is outlined in Power BI so that the visualization results make it easier for business policy makers to support decision making. With a concept like in this paper, the forecasting process that is carried out repeatedly can create a data science media because it is able to pour business processes from various data and fields..

V. ACKNOWLEDGMENTS

We would like to thank all those who have contributed to this research.

REFERENCES

- [1] M. K. M. Nasution, O. Salim Sitompul, and E. Budhiarti Nababan, "Data Science," *J. Phys. Conf. Ser.*, vol. 1566, no. 1, 2020, doi: 10.1088/1742-6596/1566/1/012034.
- [2] G. Vicario and S. Coleman, "A review of data science in business and industry and a future view," *Appl. Stoch. Model. Bus. Ind.*, vol. 36, no. 1, pp. 6–18, Jan. 2020, doi: <https://doi.org/10.1002/asmb.2488>.
- [3] L. Cao, "Data science: A comprehensive overview," *ACM Comput. Surv.*, vol. 50, no. 3, 2017, doi: 10.1145/3076253.
- [4] M. Molina-Solana, M. Ros, M. D. Ruiz, J. Gómez-Romero, and M. J. Martin-Bautista, "Data science for building energy management: A review," *Renew. Sustain. Energy Rev.*, vol. 70, pp. 598–609, 2017, doi: <https://doi.org/10.1016/j.rser.2016.11.132>.
- [5] B. V. Dhar, "P64-Dhar_Data_Science_Prediction," 2013.
- [6] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Comput. Educ.*, vol. 143, p. 103676, 2020, doi: <https://doi.org/10.1016/j.compedu.2019.103676>.
- [7] A. R. Lubis, M. Lubis, Al-Khowarizmi, and D. Listriani, "Big Data Forecasting Applied Nearest Neighbor Method," 2019, doi: 10.1109/ICSECC.2019.8907010.
- [8] R. F. Rahmat, A. Rizki, A. F. Alharthi, and R. Budiarto, "Big data forecasting using evolving multi-layer perceptron," *2016 4th Saudi Int. Conf. Inf. Technol. (Big Data Anal. KACSTIT 2016)*, 2016, doi: 10.1109/KACSTIT.2016.7756069.
- [9] N. Shabbir, R. AhmadiAhangar, L. Kütt, M. N. Iqbal, and A. Rosin, "Forecasting Short Term Wind Energy Generation using Machine Learning," in *2019 IEEE 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)*, 2019, pp. 1–4, doi: 10.1109/RTUCON48111.2019.8982365.
- [10] A. Boukerche and J. Wang, "Machine Learning-based traffic prediction models for Intelligent Transportation Systems," *Comput. Networks*, vol. 181, p. 107530, 2020, doi: <https://doi.org/10.1016/j.comnet.2020.107530>.
- [11] A. Al-Khowarizmi, O. S. Sitompul, S. Suherman, and E. B. Nababan, "Measuring the Accuracy of Simple Evolving Connectionist System with Varying Distance Formulas," *J. Phys. Conf. Ser.*, vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012004.
- [12] Y. Demchenko *et al.*, "EDISON Data Science Framework: A Foundation for Building Data Science Profession for Research and Industry," in *2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, 2016, pp. 620–626, doi: 10.1109/CloudCom.2016.0107.

- [13] T. Radivilova, L. Kirichenko, and B. Vitalii, "Comparative analysis of machine learning classification of time series with fractal properties," in *2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL)*, 2019, pp. 557–560, doi: 10.1109/CAOL46282.2019.9019416.
- [14] A. R. Lubis, S. Prayudani, and Al-Khowarizmi, "Optimization of MSE Accuracy Value Measurement Applying False Alarm Rate in Forecasting on Fuzzy Time Series based on Percentage Change," in *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, 2020, pp. 1–5, doi: 10.1109/CITSM50537.2020.9268906.
- [15] R. Gao and O. Duru, "Parsimonious fuzzy time series modelling," *Expert Syst. Appl.*, vol. 156, p. 113447, 2020, doi: <https://doi.org/10.1016/j.eswa.2020.113447>.
- [16] A. L. Rucitra and M. A. Pradana, "Demand forecasting for crude palm oil (CPO) using the time series method," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 475, no. 1, 2020, doi: 10.1088/1755-1315/475/1/012059.
- [17] M. Handini, "The effect of sludge on oil palm mills with different doses on the growth and yield of two varieties of soybean plants (*Glycine max* (L.) Merrill)," in Bahasa "Pengaruh pemberian sludge pabrik kelapa sawit dengan dosis yang berbeda terhadap pertumbuhan dan," Doctoral Dissertation, Universitas Islam Negeri Sultan Sarif Kasim Riau, 2014.
- [18] 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," *Bull. Electr. Eng. Informatics*, vol. 9, no. 4, pp. 1604–1611, 2020, doi: 10.11591/eei.v9i4.2388.
- [19] M. E. Al Khowarizmi, Rahmad Syah, Mahyuddin K. M. Nasution, "Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 3, 2021, doi: <http://doi.org/10.11591/ijece.v11i3.pp%25p>.
- [20] L. M. D. Yiu, A. C. L. Yeung, and T. C. E. Cheng, "The impact of business intelligence systems on profitability and risks of firms," *Int. J. Prod. Res.*, pp. 1–24, May 2020, doi: 10.1080/00207543.2020.1756506.
- [21] G. Andrienko *et al.*, "Big data visualization and analytics: Future research challenges and emerging applications," *CEUR Workshop Proc.*, vol. 2578, 2020.
- [22] M. Pearson, B. Knight, D. Knight, and M. Quintana, "Administering Power BI BT - Pro Microsoft Power Platform: Solution Building for the Citizen Developer," M. Pearson, B. Knight, D. Knight, and M. Quintana, Eds. Berkeley, CA: Apress, 2020, pp. 321–330.
- [23] E. Bas, U. Yolcu, and E. Egrioglu, "Intuitionistic fuzzy time series functions approach for time series forecasting," *Granul. Comput.*, no. 2013, 2020, doi: 10.1007/s41066-020-00220-8.
- [24] A. Ramírez-Rojas, L. D. G. Sigalotti, E. L. Flores Márquez, and O. Rendón, "Chapter 2 - WITHDRAWN: Stochastic processes," A. Ramírez-Rojas, L. D. G. Sigalotti, E. L. Flores Márquez, and O. B. T.-T. S. A. in S. Rendón, Eds. Elsevier, 2019, pp. 21–86.
- [25] T. C. Mills, "Chapter 1 - Time Series and Their Features," T. C. B. T.-A. T. S. A. Mills, Ed. Academic Press, 2019, pp. 1–12.
- [26] D. S. G. Pollock, "Chapter 19 - Prediction and Signal Extraction," in *Signal Processing and its Applications*, D. S. G. B. T.-H. of T. S. A. Pollock Signal Processing, and Dynamics, Ed. London: Academic Press, 1999, pp. 575–615.
- [27] 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," *J. Phys. Conf. Ser.*, vol. 1361, no. 1, pp. 0–8, 2019, doi: 10.1088/1742-6596/1361/1/012089.
- [28] A. R. Lubis, M. Lubis, and A.- Khowarizmi, "Optimization of distance formula in K-Nearest Neighbor method," *Bull. Electr. Eng. Informatics*, vol. 9, no. 1, pp. 326–338, 2020, doi: 10.11591/eei.v9i1.1464.