SJER Scientific Journal of Engineering Research

PTTI PENELITI TEKNOLOGI TEKNIK INDONESIA

Date of publication January 30, 2025, date of current version January 30, 2025.
Digital Object Identifier ***10.59247/sjer.v1i1.11***

Article

Implementation of Data Mining Using Simple Linear Regression Algorithm to Predict Export Values

Aldi Bastiatul Fawait[1,*], Sitti Rahmah[1], Apolonia Diana Sherly da Costa[2], Nazaruddin Insyroh[3], Asno Azzawagama Firdaus[4]

[1] Department of Computer Science, Universitas Widya Gama Mahakam, Samarinda, Indonesia; aldi.bas.fawait@uwgm.ac.id; sitti.rahmah@uwgm.ac.id
[2] Guest Scientist of Interdisciplinary Natural Disaster Risk Management, Social Geography and Social Ecology at the Institute for Landscape Ecology, Faculty of Geo Science, University of Münster, Germany; sherly.dacosta@uni-muenster.de
[3] Department of Law and Digital Technologies, Leiden University, Leiden, Netherlands; n.insyroh@umail.leidenuniv.nl
[4] Department of Computer Science, Universitas Qamarul Huda Badaruddin Bagu, Central Lombok, 83371, Indonesia; asnofirdaus@uniqhba.ac.id
* Correspondence

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Abstract: This study aims to analyze the trends in export value in East Kalimantan. The research utilizes secondary data sourced directly from the Central Statistics Agency of East Kalimantan Province. A simple linear regression algorithm for data mining is employed as the analytical method. The findings indicate a decline in East Kalimantan's export value from January 2022 to April 2024, as well as in the forecasted export value from May 2024 to December 2024. The prediction model achieved a Root Mean Square Error (RMSE) value of 3.182%, demonstrating a high level of accuracy in estimating export values. This research is expected to serve as a valuable reference for stakeholders in formulating strategies to enhance East Kalimantan's export performance and contribute to the region's future economic development.

Keywords: Data mining; Prediction; RapidMiner; RMSE; Simple Linear Regression Algorithm

1. Introduction

International trade is a commercial activity involving residents of one country and those of another, based on mutual agreements [1, 2]. The participants may include individuals, individuals and governments, or governments of two or more different countries [3]. International trade encompasses the exchange of goods and services between economic entities from various nations [4, 5]. Essentially, international business activities are driven by the same motive as business in general—the pursuit of benefits from these activities. Exports refer to the act of selling one's products abroad in compliance with government regulations [6], typically requiring payment in foreign currency and communication in a foreign language. The proceeds from export activities are expressed in the form of foreign currency, also known as foreign exchange, which constitutes one of the state's sources of revenue [7, 8].

International trade consists of the shipment of goods (exports) and the receipt of goods (imports) from abroad [9, 10]. In the era of globalization, the activity of exporting goods has become increasingly crucial, as exports serve as one of the driving forces of a country's economy [11, 12]. According to David Ricardo's classical theory, a country can gain benefits from international trade if it specializes in producing and exporting goods that it can produce relatively more efficiently while importing goods that it produces less efficiently [13, 14].

International trade is an economic sector that plays a vital role in supporting the economic development of Indonesia, particularly in East Kalimantan [15, 16]. Export activities generate foreign exchange and capital for development, while imports provide the raw materials and capital needed for progress.

In a previous study it was revealed that exports and exchange rates collectively have a significant impact on

foreign exchange reserves. Separately, exports were found to have a significant effect on foreign exchange reserves. Subsequent research states that the results of simultaneous testing conclude that inflation, the rupiah exchange rate, interest rates, and exports influence foreign exchange reserves. Based on the simultaneous testing, it can be concluded that inflation, the rupiah exchange rate, interest rates, and exports significantly impact Indonesia's foreign exchange reserves. Meanwhile, partial testing results show that export values have a positive and significant effect on foreign exchange reserves.

East Kalimantan contributes to national revenue through international trade, categorized into two main sectors: oil and gas (O&G) and non-oil and gas (non-O&G) exports [17]. The export of oil and gas has shown a declining trend, while the value of non-O&G exports has surpassed that of O&G exports. The export value of East Kalimantan from 2022 to 2024 has experienced a year-on-year decline. In 2022, the export value reached USD 36,058.24 million, decreasing to USD 26,840.63 million in 2023, and further dropping to USD 8,041.06 million in the first quarter of 2024.

To assess the extent of export value development in East Kalimantan, a predictive method is required to estimate future export values in the region. Advances in techniques for collecting and storing data have enabled organizations to manage large volumes of data, resulting in the creation of datasets. Data mining involves a series of processes applied to a group of data to uncover knowledge that is not easily discernible through manual methods. Data mining employs various algorithms to extract valuable information. Simple linear regression is a least-squares method designed to test the causal relationship between an independent variable (X) and a dependent variable (Y). An effective plan to prepare for future scenarios involves making accurate predictions or forecasts about these conditions. Prediction refers to forecasting an event that will occur in the future. It can be applied not only for time-series forecasting but also for classification processes, as it identifies categories based on existing attributes. The root mean square error (RMSE) is the square root of the mean squared error (MSE) obtained from algorithmic calculations. In this study, the author predicts the export value in East Kalimantan using data mining methods, specifically the simple linear regression algorithm, implemented through the RapidMiner application.

2. Method

Descriptive analysis is a method used to analyze data within a data set. What makes it unique is that it can be used to write or provide explanations about a subject by using data or samples that have been collected and analyzed to make decisions that are applicable to everyone. This research employs a data mining methodology using a simple linear regression algorithm. Several stages are carried out, which can be seen in the flowchart in Figure 1.

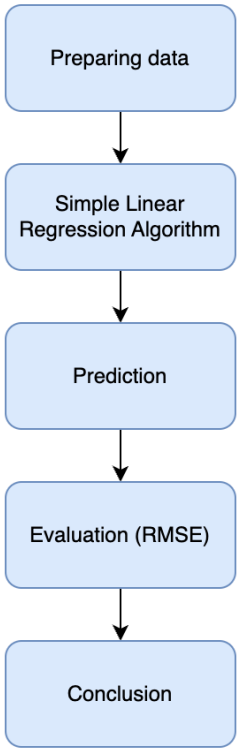

Figure 1. Research flow

2.1. Data Mining

Experts state that data mining is a form of analysis for large-scale observational data to uncover unknown relationships and new methods for interpreting data to make it easier to understand. Data mining is one of the steps in Knowledge Discovery in Databases (KDD) [18-20]. Data mining is defined as the process of identifying meaningful relationships, patterns, and trends by analyzing large volumes of stored data, using pattern recognition techniques that involve statistical and mathematical methods.

2.2. Simple Linear Regression Algorithm

This algorithm is a prediction technique that uses a single independent variable [21, 22]. This method is used to predict the dependent variable Y. The analysis is based on the causal relationship between the independent variable and the dependent variable. The regression coefficient is used to determine the effect of the independent variable's value on the dependent variable's value. The determination of the regression equation for values of a and b is given by Equation 1-3.

$$a = \frac{(\sum x^2)(\sum y) - (\sum xy)(\sum x)}{n(\sum x^2) - (\sum x)} \quad (1)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (2)$$

With its equation: $Y = a + bX$ (3)

Where : X = independent variable
Y = dependent variable
a = constant
b = slope

2.3. Prediction Accuracy Measurement

The accuracy of a prediction is determined by the magnitude of the difference or error between the predicted data and the actual data. At this stage, it is checked whether the performance value reflects the error value on the RMSE, indicating whether the performance is good or poor [23, 24].

$$RMSE = \sqrt{\frac{\sum_{t-1}^{n}(A_t - F_t)^2}{n}} \quad (4)$$

In Equation 4, we know that At represents the actual data value, Ft is the predicted value, n is the number of data points, and Σ is the total of all values. The Root Mean Square Error (RMSE), which uses a gradient-based method, indicates that the lower the RMSE, the more accurate the predictions [25].

2.4. Rapidminer Application

RapidMiner is an application or software tool for data mining knowledge [26]. The company that developed this platform is dedicated to applications that involve large data sets in fields such as trade, research, education, training, and learning. RapidMiner is an independent software that functions to analyze data and serves as a data mining tool, which can be easily integrated with various programming languages.

2.5. Data Analysis

The dataset to be used is obtained from the website of the Central Statistics Agency of East Kalimantan at https://kaltim.bps.go.id/publication.html, which provides various statistical data for East Kalimantan. The specific data on export values for oil and gas (migas) and non-oil and gas (non-migas), as well as the total export value, is available at https://kaltim.bps.go.id/indicator/8/36/1/nilai-ekspor---impor-bulanan-.html. This data, which spans from January 2022 to April 2024, will be used as the test data for historical export values in East Kalimantan (Table 1). Here, X represents the time period from January 2022 to April 2024, and Y represents the total export value of East Kalimantan in millions of US dollars.

3. Results and Discussion

The RapidMiner application can generate information that is typically not noticed by users, allowing the results to serve as a basis for decision-making. The researchers use a linear regression algorithm to predict the export values from May to December 2024. The following steps are taken:

3.1. The equation model design for the linear regression algorithm in RapidMiner.

In this stage, the equation model is created using the linear regression model, as shown in Figure 2. The application of the Simple Linear Regression algorithm using RapidMiner on the testing data resulted in a linear equation. The calculations provided by the software yielded the following regression equation: Y = 3123.836 – 40.708 X.

Table 1. Historical export values in East Kalimantan (January 2022 to April 2024).

Month	X (Period)	Y (Export Value)
Jan-22	1	1084.45
Feb-22	2	2018.40
Mar-22	3	3044.22
Apr-22	4	3329.79
Mei-22	5	2855.18
Jun-22	6	3675.45
Jul-22	7	3596.81
Agu-22	8	3706.87
Sep-22	9	3319.92
Okt-22	10	3254.49
Nov-22	11	3120.19
Des-22	12	3052.47
Jan-23	13	2449.65
Feb-23	14	2650.89
Mar-23	15	2850.50
Apr-23	16	2405.40
Mei-23	17	2245.52
Jun-23	18	1950.80
Jul-23	19	1929.82
Agu-23	20	2023.22
Sep-23	21	1910.92
Okt-23	22	2077.21
Nov-23	23	2186.71
Des-23	24	2159.99
Jan-24	25	1946.46
Feb-24	26	1766.64
Mar-24	27	2215.00
Apr-24	28	2112.96

Source: The Central Statistics Agency of East Kalimantan Province

Table 2. Significance level in RapidMiner.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value
X	-40.708	13.879	-0.499	?	-2.933	0.007
(Interception)	3123.835	230.373	?	?	13.560	0.000

Figure 2. Design of Simple Linear Regression Algorithm in RapidMiner

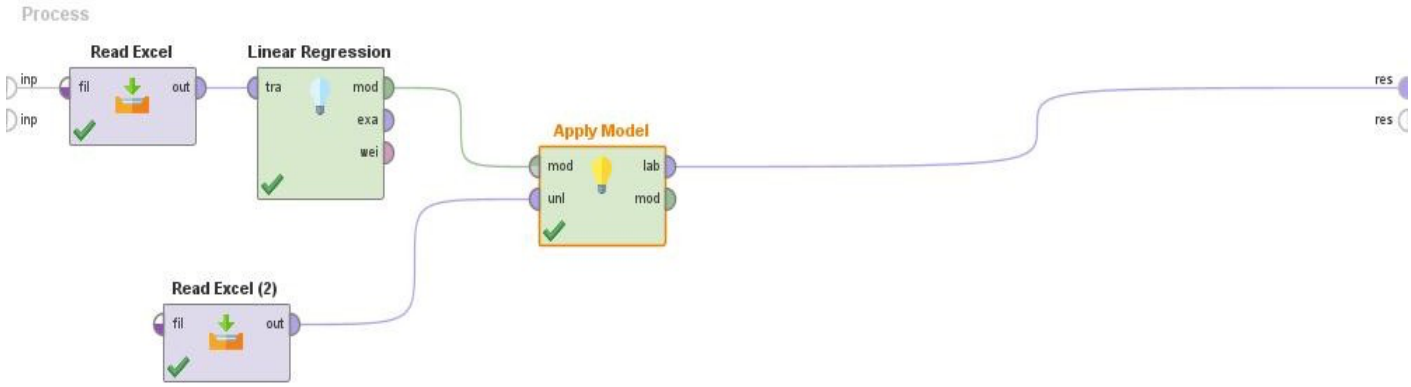

Figure 3. The prediction design in the RapidMiner application

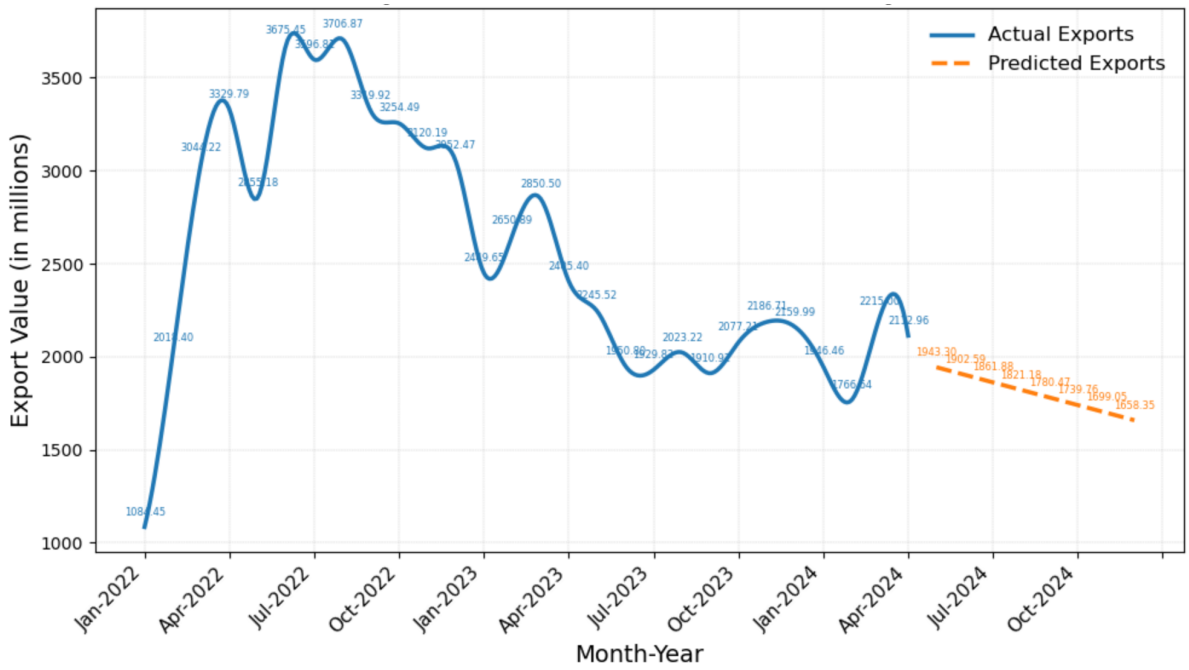

Figure 4. Prediction Results in RapidMiner Application

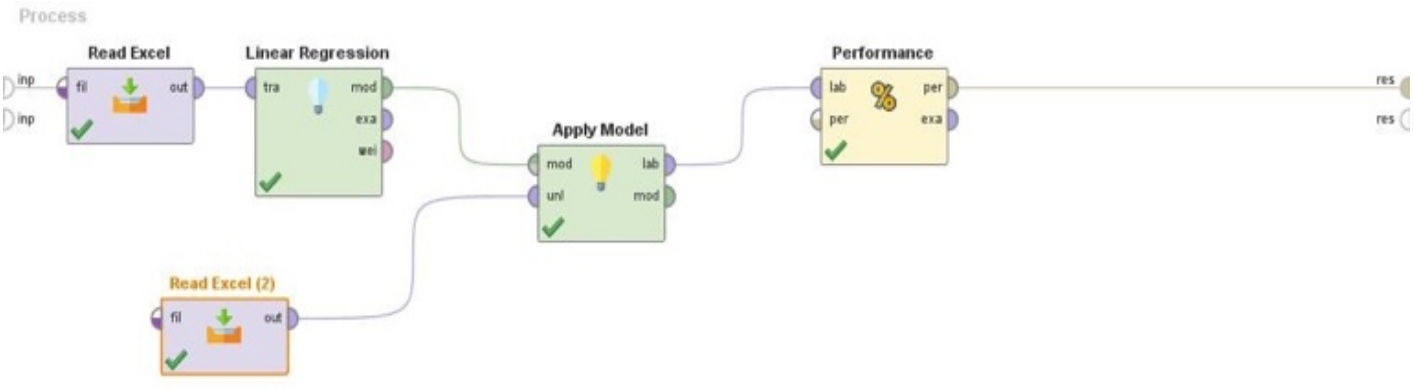

Figure 5. Evaluation Design of Accuracy Measures

Based on Table 2, the hypothesis is obtained:

Ho: There is no relationship between X (period) and Y (export value in East Kalimantan)

H1: There is a relationship between X (period) and Y (export value in East Kalimantan)

Hypothesis Acceptance Criteria: If the P-value is < 0.05, then Ho is rejected, and vice versa.

Since the P-value is 0.007 < 0.05, Ho is rejected, and H1 is accepted. This means there is a relationship between X (period) and Y (East Kalimantan's export value). Therefore, this model equation can be used in subsequent calculations.

3.2. Create a prediction model in RapidMiner

At this stage, the prediction model is created using a linear regression model, as shown in Figure 3. The prediction results generated by the RapidMiner application for the available testing data are shown in Figure 4. Based on the calculations obtained from Rapidminer in Figure 4, the predicted export values for East Kalimantan (in million US$) from May to December 2024 are as follows: 1943.302, 1902.593, 1861.885, 1821.177, 1780.469, 1739.761, 1699.053, 1658.345.

3.3. Create an RMSE evaluation model in RapidMiner

At this stage, the model for evaluating prediction accuracy is created using a linear regression model, as shown in Figure 5.

The accuracy evaluation of the predictions made by the RapidMiner application on the testing data shows an RMSE value of 3.182%. This value indicates that the predicted values are very close to the observed values, with a smaller RMSE signifying higher accuracy. In this case, the predicted export value for East Kalimantan is considered to be highly accurate.

4. Conclusion

This study utilizes a data mining method with a Simple Linear Regression algorithm, based on the export values in East Kalimantan from January 2022 to April 2024. The research results yield the following Simple Linear Regression algorithm equation: Y = 3123.836 – 40.708 X. The predicted export values for East Kalimantan (in million US$) from May to December 2024 are as follows: 1943.302, 1902.593, 1861.885, 1821.177, 1780.469, 1739.761, 1699.053, 1658.345. The RMSE value is 3.182%, indicating that the predicted export values for East Kalimantan are classified as highly accurate.

Based on the results of previous studies, it was concluded that export value significantly impacts the country's foreign exchange reserves. According to the findings of this study, the predicted export value of East Kalimantan from May 2024 to December 2024 is expected to decline. Therefore, based on the conclusions of both previous studies and this research, it is recommended that relevant parties develop effective strategies to increase the export value of East Kalimantan in the future.

5. Conflicts of Interest

The authors declare no conflicts of interest.

6. References

[1] R. Grosse, J. Gamso, and R. C. Nelson, "China's Rise, World Order, and the Implications for International Business," *Management International Review*, vol. 61, no. 1, pp. 1–26, Mar. 2021, doi: 10.1007/s11575-020-00433-8.

[2] M. A. Witt, A. Y. Lewin, P. P. Li, and A. Gaur, "Decoupling in international business: Evidence, drivers, impact, and implications for IB research," *Journal of World Business*, vol. 58, no. 1, p. 101399, Jan. 2023, doi: 10.1016/j.jwb.2022.101399.

[3] V. Z. Eichenauer, A. Fuchs, and L. Brückner, "The effects of trade, aid, and investment on China's image in Latin America," *J Comp Econ*, vol. 49, no. 2, pp. 483–498, Jun. 2021, doi: 10.1016/j.jce.2020.08.005.

[4] F. Sudirjo, "Marketing Strategy in Improving Product Competitiveness in the Global Market," *Journal of Contemporary Administration and Management (ADMAN)*, vol. 1, no. 2, pp. 63–69, Aug. 2023, doi: 10.61100/adman.v1i2.24.

[5] L. Tajoli, F. Airoldi, and C. Piccardi, "The network of international trade in services," *Appl Netw Sci*, vol. 6, no. 1, p. 68, Dec. 2021, doi: 10.1007/s41109-021-00407-1.

[6] T. A. Hassan and S. H. Bhatti, "International Business Law and Regulations," *Journal of Social Sciences Review*, vol. 3, no. 1, pp. 422–432, Mar. 2023, doi: 10.54183/jssr.v3i1.122.

[7] S. Choi, Md. A. Salam, and Y. Kim, "Foreign currency derivative usage and firm value in Bangladesh: comparative analysis between exporters and non-exporters under exchange rate

movements," *International Journal of Emerging Markets*, vol. 16, no. 8, pp. 2070–20921, Oct. 2021, doi: 10.1108/IJOEM-08-2019-0641.

[8] M. A. Wildan, I. Imron, E. Siswati, and S. Rosyafah, "Macroeconomic Factors Affecting Natural Gas Export Management," *International Journal of Energy Economics and Policy*, vol. 11, no. 1, pp. 639–644, Dec. 2020, doi: 10.32479/ijeep.9911.

[9] L. Baranauskaitė and D. Jurevičienė, "Import Risks of Agricultural Products in Foreign Trade," *Economies*, vol. 9, no. 3, p. 102, Jul. 2021, doi: 10.3390/economies9030102.

[10] J. Verschuur, E. E. Koks, and J. W. Hall, "Ports' criticality in international trade and global supply-chains," *Nat Commun*, vol. 13, no. 1, p. 4351, Jul. 2022, doi: 10.1038/s41467-022-32070-0.

[11] M. Jiang and P. Jia, "Does the level of digitalized service drive the global export of digital service trade? Evidence from global perspective," *Telematics and Informatics*, vol. 72, p. 101853, Aug. 2022, doi: 10.1016/j.tele.2022.101853.

[12] M. Mokthtari Moughari and T. U. Daim, "Developing a model of technological innovation for export development in developing countries," *Technol Soc*, vol. 75, p. 102338, Nov. 2023, doi: 10.1016/j.techsoc.2023.102338.

[13] N. M. Abdwlglil and S. F. S. Taib, "The Development of Modern Classical Trade Theories and Contribution to Understanding Trade Patterns and Trade Exchange Processes in the Global Economy," *International Science and Technology Journal*, vol. 34, no. 1, pp. 1–18, Apr. 2024, doi: 10.62341/nasf1817.

[14] T. P. Bagchi, R. P. Mohanty, and S. Sinha, "A tutorial on optimisation involving the David Ricardo theory on comparative advantage," *International Journal of Industrial and Systems Engineering*, vol. 44, no. 1, p. 34, 2023, doi: 10.1504/IJISE.2023.130918.

[15] A. K. Mahmud, B. Iwang, A. Muh. Kayyum, and N. H. Binti Sa'at, "How Technological Development and E-Commerce Drive Economic Growth in Indonesia," *EcceS (Economics Social and Development Studies)*, vol. 11, no. 1, pp. 73–97, Jun. 2024, doi: 10.24252/ecc.v11i1.45419.

[16] R. W. Kusuma, R. Farhan, S. Budiman, S. Kusumawati, A. S. Hayyati, and R. F. Putri, "Performance of Regional Economic in East Kalimantan: Leading Sector of Export-Import Commodities in 2019-2020," *IOP Conf Ser Earth Environ Sci*, vol. 1233, no. 1, p. 012031, Aug. 2023, doi: 10.1088/1755-1315/1233/1/012031.

[17] D. Wulan Sari and I. Mukhlis, "Internationalization of MSMEs Based on Comparative Advantage (Case Study of Nipah Lidi and Palm Oil Exporters in East Kalimantan)," *Asian Journal of Management Entrepreneurship and Social Sciene*, vol. 2, 2022, Accessed: Jan. 30, 2025. [Online]. Available: https://www.ajmesc.com/index.php/ajmesc/article/view/101/50

[18] U. Fayyad, "Data mining and knowledge discovery in databases: implications for scientific databases," in *Proceedings. Ninth International Conference on Scientific and Statistical Database Management (Cat. No.97TB100150)*, IEEE Comput. Soc, pp. 2–11. doi: 10.1109/SSDM.1997.621141.

[19] R. Rajab Asaad and R. Masoud Abdulhakim, "The Concept of Data Mining and Knowledge Extraction Techniques," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 17–20, Mar. 2021, doi: 10.48161/qaj.v1n2a43.

[20] M. M. Ghazal and A. Hammad, "Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects," *International Journal of Construction Management*, vol. 22, no. 9, pp. 1632–1646, Jul. 2022, doi: 10.1080/15623599.2020.1738205.

[21] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques," *Journal of Building Engineering*, vol. 45, p. 103406, Jan. 2022, doi: 10.1016/j.jobe.2021.103406.

[22] D. Alita, A. D. Putra, and D. Darwis, "Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, p. 295, Jul. 2021, doi: 10.22146/ijccs.65586.

[23] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Inf Sci (N Y)*, vol. 585, pp. 609–629, Mar. 2022, doi: 10.1016/j.ins.2021.11.036.

[24] D. Chicco, M. J. 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 Sci*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[25] S. K. Jauhar, S. M. Jani, S. S. Kamble, S. Pratap, A. Belhadi, and S. Gupta, "How to use no-code artificial intelligence to predict and minimize the inventory distortions for resilient supply chains," *Int J Prod Res*, vol. 62, no. 15, pp. 5510–5534, Aug. 2024, doi: 10.1080/00207543.2023.2166139.

[26] S. A. Siallagan and M. Safii, "Grouping of Toddlers with Malnutrition Based on Provinces in Indonesia Using K-Medoids Algorithm," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 1, no. 1, pp. 47–53, Oct. 2021, doi: 10.59934/jaiea.v1i1.53.