Rice Price Prediction in Province Nusa Tenggara Barat (NTB) Using Comparison of Linear Regression and Random Forest Algorithms

Article history

Received:
10 November 2025

Received in revised form:
15 November 2025

Accepted:
2 December 2025

Published online:
26 December 2025

[1]Ujang Wiharja, [2]Sri Hartanto, [3]Abdul Kodir Albahar

Universitas Krisnadwipayana, Jakarta
ujangwiharja@unkris.ac.id

*Corresponding author
ujangwiharja@unkris.ac.id

Abstract

This research provides transparent prediction accuracy in rice industry management and can be used to more accurately forecast prices in the West Nusa Tenggara region. To determine whether the model provides better prediction, the researchers analyzed and forecasted rice prices using two machine learning algorithms: Linear Regression and Random Forest. Forecasting rice prices is difficult because the elements that support changes in rice prices, such as planted land, production levels, consumption levels, currency (rupiah) volatility, and the volume of rice imports into Indonesia, are interrelated. Based on the study, Random Forest outperformed Linear Regression, with an R^2 value of 0.710, indicating a better model fit. In addition, the Random Forest algorithm shows a lower error rate, which is reflected in the RMSE of 1038,394. The dataset used for this study covers the period 2006 to 2021 and is sourced from various official institutions, including the Central Bureau of Statistics and Bank Indonesia.

Keywords: *Prediction, Rice, Prices, Forest, Regression*

1. Introduction

The agricultural or food industry is one of the issues and it is the one that the government emphasizes the most. In Indonesia, rice is the main staple food [1]. As a result, the government always strives to increase food supply, especially for rice commodities [2]. In addition, the government has never hesitated to import rice and has always strictly maintained the supply of rice commodities [1], [2].

The current price problem is a very complex issue, and it is difficult to predict the price of rice, which is influenced by several factors such as planting area, rice and rice production, rice consumption, rupiah, and the amount of rice imports [3]. Technological advances have made humans more effective and conscientious workers [4]. Appropriate feature selection and reduction methods can improve prediction accuracy and model training efficiency [5]. Decision-making based on more complicated statistical analysis is now more effectively implemented thanks to data mining and improved statistical technology [4]-[8]. Using data mining

* *Corresponding author. ujangwiharja@unkris.ac.id*

techniques, comparison algorithms can be used to find the best performance model to predict rice prices more precisely, and people will increasingly feel its usefulness [6],[9].

The algorithms used in this study are Linear Regression and Random Forest algorithms [4], [10]-[13]. The data used for rice prices are planting area, rice and rice production, rice consumption, rupiah and rice imports. This research data comes from datasets from the Central Bureau of Statistics of NTB Province, the NTB Provincial Food Security Office, Bank Indonesia, the Central Bureau of Statistics Indonesia, and the NTB Provincial Trade Office from 2006 – 2021.

Manuscript must be an original work that has not been published or under consideration for publication elsewhere. The journal welcomes manuscripts clearly written in English and should be typed on A4 size paper format. The margins are 3.3cm on the left side, 3.65cm on the right, 2.03cm on the top, and 3.05cm on the bottom. Paper orientation in all pages should be in portrait style.

2. Technical Work Preparation

2.1. System Requirement

System requirements analysis is carried out to find anticipated needs so that they can be incorporated into an application.

i. Hardware Requirements
   a. Model: MacBook Pro
   b. Processor: 2.4 GHz Intel Core i5
   c. Memory: 4GB 1067 MHz DDR3
ii. Software Requirements
   a. Operation Systems Windows 10
   b. Orange
   c. Microsoft Excel and Other Supporters.

2.2. Variabels and Theirs Measurement

In this study, the following types of variables were measured:

i. The price of rice is listed in Rupiah per kilogram (Rp/Kg) and is the selling rice in Nusa Tenggara Barat Province.
ii. Production is measured in tons, which is the volume of rice grown in NTB Province.
iii. The amount of rice consumed in NTB province.
iv. Current Indonesian rupiah exchange rate.
v. Import rice in tons.

2.3. Data Collection

Data was collected from various sources, including the Central Bureau of statistics of NTB Province, NTB Provincial Food Security Office, Bank Indonesia, and NTB Provincial Trade Office, which was used in this study. Data was collected from 2006 to 2021. Here is the dataset that will be used in Table 1: Development of Planting Area, Production and Rice Productivity in NTB Province in 2010 – 2021, Table 2: Development of Rice Production in NTB in 2006 – 2021, Table 3: Development of Rice Consumption in NTB in 2006 – 2021, Table 4: Development of Rupiah Exchange Rate per USD in 2006 – 2021, Table 5: Development of Indonesia's Rice Imports in 2006 – 2021, Table 6: Development of Rice Prices in NTB Province in 2006 – 2021.

Table 1. Development of Planted Area, Production and Productivity of Rice in NTB Province, 2010-2021.

Year	Planting Area (HA)	Production (TON)	Productivity (TON/HA)
2010	374.284	1.774.499	47,41
2011	418.062	2.067.137	49.45
2012	425.448	2.114.231	49.69
2013	438.057	2.193.698	50.08
2014	433.712	2.116.637	48.80
2015	456.395	2.330.865	51.07
2016	460.662	2.095.118	46.49
2017	471.728	2.323.700	49.26
2018	289.243	1.460.338	50.49
2019	281.668	1.402.182	49.78
2020	273.462	1.317.190	48.17
2021	277.113	1.432.460	51.69

Table 2. Development of Rice Production in NTB in 2006-2021

Year	RICE PRODUCTION (TON)	Production (%)
2006	877.845	-
2007	862.849	(1,71)
2008	990.852	14,83
2009	1.059.380	6,92
2010	995.036	(6,07)
2011	1.031.831	3,70
2012	1.095.082	6,13
2013	1.136.242	3,76
2014	1.190.042	4,37
2015	1.359.136	14,21
2016	1.177.944	(13,33)
2017	1.358.750	15,35
2018	1.409.855	3,76
2019	794.498	(43,64)
2020	746.336	(6,06)
2021	811.769	8,76

Table 3. Development of Rice Consumption in NTB in 2006-2021

Year	Rice Production (TON)	Production (%)
2006	527.480	-
2007	522.396	(0,96)
2008	521.905	(0,09)
2009	520.996	(0,17)
2010	518.874	(0,41)
2011	514.568	(0,83)
2012	509.678	(0,95)
2013	504.240	(1,07)
2014	560.921	11,24
2015	567.213	1,12
2016	573.341	1,08
2017	580.298	1,21
2018	586.601	1,09
2019	625.963	6,71
2020	618.042	(1,26)
2021	637.736	3,19

Table 4. Development of Rupiah Exchange Rate per USD in 2006-2021

Year	Rupiah/USD	Rupiah/USD (%)
2006	9.065	-
2007	9.466	4,23
2008	11.005	13,98
2009	9.447	(16,49)
2010	9.036	(4,54)
2011	9.113	(0,84)
2012	9.718	6,22
2013	12.250	20,66
2014	12.502	2,01
2015	13.864	9,82
2016	13.503	(2,67)
2017	13.616	0,82
2018	14.553	6,43
2019	13.970	(4,17)
2020	14.175	1,45
2021	14.340	1,15

Table 5. Development of Indonesia Rice Import Amount in 2010-2021

Year	Import (TON)	Import (%)
2006	438.109	-
2007	1.406.848	68,85
2008	289.689	(385,64)
2009	250.473	(15,65)
2010	687.582	63,57
2011	2.750.476	75,00
2012	1.810.372	(51,92)
2013	472.665	(283,01)
2014	844.164	44,01
2015	861.601	2,06
2016	1.283.179	32,85
2017	305.279	(320,32)
2018	2.253.825	86,45
2019	444.509	(407,03)
2020	356.286	19,85
2021	252.376	(41,17)

Table 6. Development of Rice Price in NTB Province in 2006-2021

Year	Rice Price (Rp/KG)	Rice Price (%)
2006	4.016	-
2007	4.608	14,74
2008	4.945	7,31
2009	5.286	6,89
2010	6.436	21,75
2011	6.893	7,10
2012	7.231	4,90
2013	7.802	7,89
2014	8.138	4,88
2015	8.598	5,65
2016	9.153	5,45
2017	9.075	(0,85)
2018	9.473	4,38
2019	9.431	(0,44)
2020	9.695	2,80
2021	10.103	4,21

3. Results

3.1 Results of Data Acquisition and Data Statistical Information

The following is an overview of the time-series data from the dataset used, as shown in Figure 1.

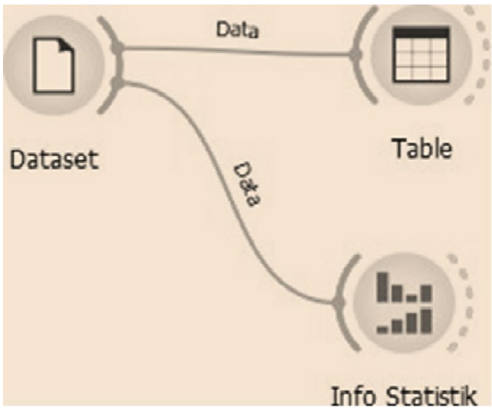


Figure 1. Statistical Information Orange Icon

It aims to display rice prices, planting area, rice and rice production, rice consumption, rupiah, and rice imports using Line Charts and time series to analyze data sorted by time, change over time, and predict future trends.

As Figure 2 shows. Rice Price Chart Dataset 2006-2021 In NTB Province, rice prices are known to have increased between 2006 and 2016. The decline in rice prices only occurred in 2017 and 2019. Up 0.85% in 2017 and 4.38% in 2019 compared to the previous year. The price of rice ranged from Rp.4,016/Kg in 2006 to Rp10,103/Kg in 2021, the most expensive price.

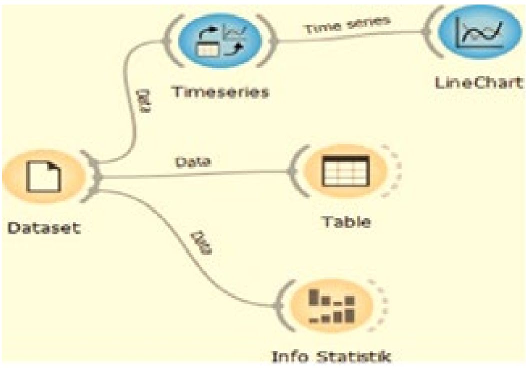


Figure 2. Orange Graphics Icon

3.2 Preprocessing

The following data preprocessing handling uses the Normalize feature with an interval of -1 to 1, shown in Figure 3, Orange Normalization Icon Display.

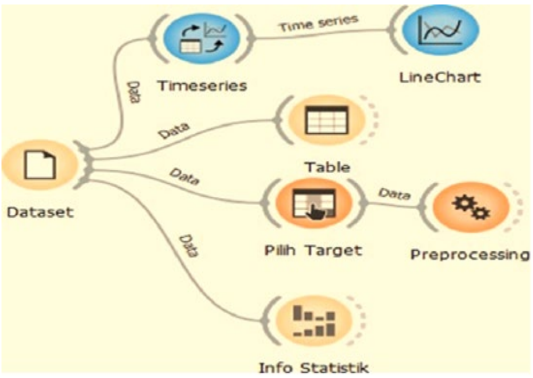


Figure 3. Orange Normalised Icon

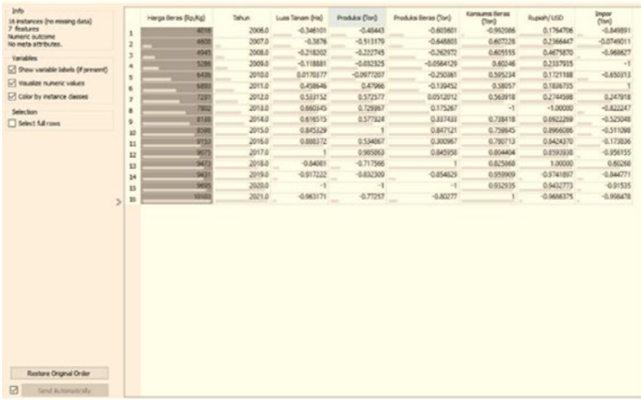


Figure 4. Dataset Normalization Results with Intervals -1 and 1

Figure 4 Display the results of preprocessing handling with normalization to change the scale of values in the dataset so that the data has a uniform range that does not dominate other variables and eliminates significant scale differences between variables, thus preventing one variable from dominating influence in analysis or modeling.

3.3 Prediction Modeling

Two algorithms, Linear Regression and Random Forest, are compared in prediction modeling. This modeling will compare model performance without and with preprocessing treatment. In this modeling, cross-validation is five-fold.

Here's an Orange Icon Display performance comparison of both models. Figure 5 for the process of conducting prediction testing using cross-validation with a comparison of linear regression and random forest algorithms without preprocessing data so that prediction results with different model performances are obtained.

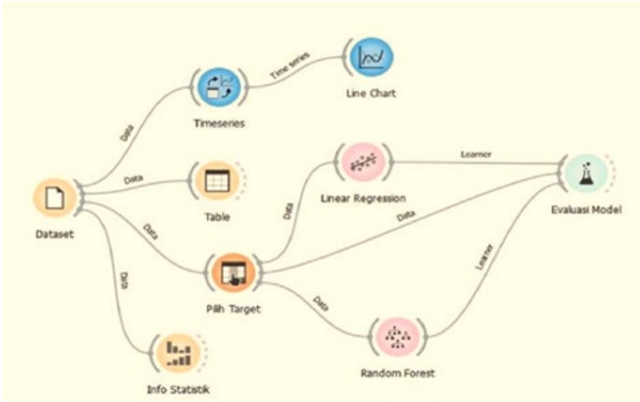


Figure 5. Orange Non-Processing Icon

Figure 6 for the process of performing prediction testing using cross-validation with a comparison of linear regression and random forest algorithms by preprocessing data so that if there are missing values, preprocessing can help overcome the problem by filling in or deleting the missing values appropriately.

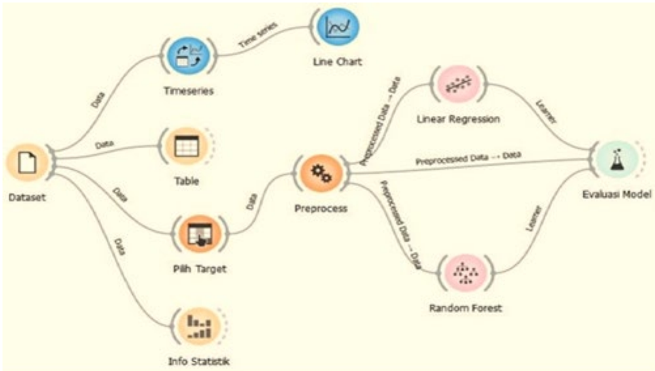
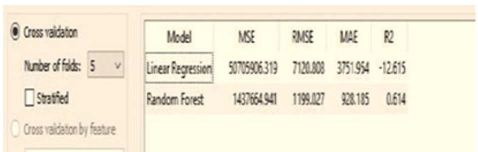


Figure 6. Orange Preprocessing Icon

4. Discussion

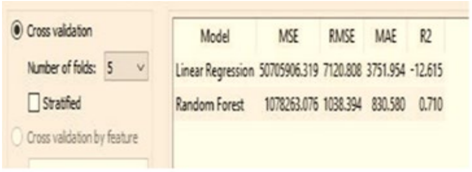
The study applied the cross-validation method to the test procedure, using five random numbers to generate test data according to the built-in prediction model [4], [10], [14]–[17]. The test results obtained by the cross-validation procedure are shown in Figure 7: For the performance of the second prediction model algorithm without preprocessing treatment that has the best performance, namely Random Forest with R2 = 0,614 dan RMSE = 1199,027.



Model	MSE	RMSE	MAE	R2
Linear Regression	50705906.319	7120.808	3751.954	-12.615
Random Forest	1437664.941	1199.027	928.185	0.614

Figure 7. Non-preprocessing Prediction Model

Figure 8 is the performance of the prediction model of the two algorithms with preprocessing treatment with the best performance, namely Random Forest with R2 = 0.710 and RMSE = 1038.394. With the results of the comparison above, the best performance is the Random Forest algorithm with preprocessing treatment. The model is used for prediction with the model and results, as shown in Figure 9.



Model	MSE	RMSE	MAE	R2
Linear Regression	50705906.319	7120.808	3751.954	-12.615
Random Forest	1078263.076	1038.394	830.580	0.710

Figure 8. Prediction Model Preprocessing

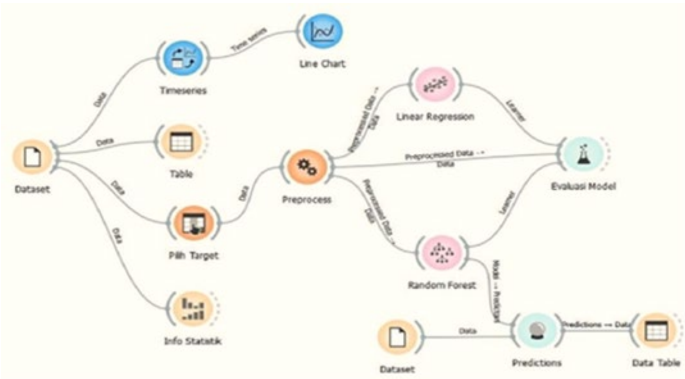
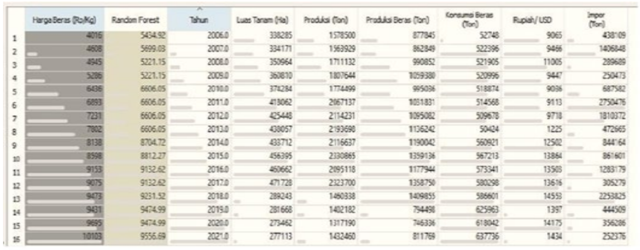


Figure 9. Orange Random Forest Icon

Figure 10 is the data predicted using random forest. Combining historical rice price data and random forest algorithms, prediction results can be optimized for analysis, helping users better understand the market and make more informed judgments.



	Harga Beras (Rp/Kg)	Random Forest	Tahun	Luas Tanam (Ha)	Produksi (Ton)	Produksi Beras (Ton)	Konsumsi Beras (Ton)	Rupiah/ USD	Impor (Ton)
1	4016	5434.92	2006.0	338285	1578500	877845	52748	9065	438109
2	4608	5699.03	2007.0	334171	1563929	862849	522396	9466	1406848
3	4945	5221.15	2008.0	350964	1711132	990852	521905	11005	289689
4	5286	5221.15	2009.0	360810	1807644	1059380	520996	9447	250473
5	6436	6606.05	2010.0	374284	1774499	995036	518874	9036	687582
6	6893	6606.05	2011.0	413062	2067137	1031831	514568	9113	2750476
7	7231	6606.05	2012.0	425448	2114231	1095082	509678	9718	1810372
8	7802	6606.05	2013.0	438057	2193698	1136242	50424	1225	472665
9	8138	8704.72	2014.0	433712	2116637	1190042	560921	12502	844164
10	8598	8812.27	2015.0	456395	2330865	1359136	567213	13864	861601
11	9153	9132.62	2016.0	460662	2095118	1177944	573341	13503	1283179
12	9075	9132.62	2017.0	471728	2323700	1358750	580298	13616	305279
13	9473	9231.52	2018.0	289243	1460338	1409855	586601	14553	2253825
14	9431	9474.99	2019.0	281668	1402182	794498	625963	1397	444509
15	9695	9474.99	2020.0	273462	1317190	746336	618042	14175	356286
16	10103	9556.69	2021.0	277113	1432460	811769	637736	1434	252376

Figure 10. Random Forest Prediction

The next analysis process determines the relationship between the most influential variables on the target data using correlations, as shown in Figure 11.

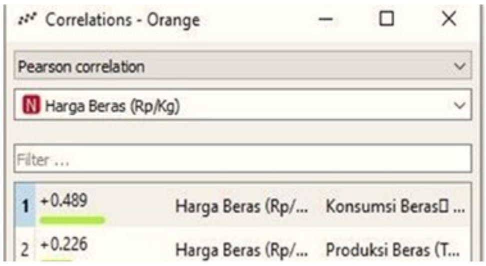


Figure 11. Correlations Results

This is a correlation display that is useful to see the magnitude of the influence on each variable rice price in NTB and provides good information in making prediction models in machine learning [6], [14], [17]–[21]. Based on the value of correlations, it can be seen the sequence of variables that affect the most, namely rice consumption and rice production, in the sense that the plus sign (+) indicates that when the value of one variable increases, the value of other variables also tends to increase while the minus sign (-) indicates that when the value of one variable increases, the value of other variables tends to decrease. In other words, the correlation sign indicates the direction of the relationship between these variables.

5. Conclusion

Based on the results of comparative research of Linear Regression and Random Forest Algorithms to predict rice prices where the accuracy results of both model performances obtained the best results with the preprocessing treatment, the Random Forest Algorithm with R2 = 0.710 means that because R2 is close to number 1, the accuracy rate is better and RMSE = 1038.394 means that the error rate is lower than the linear regression algorithm. The most influential factors on rice prices in NTB Province are rice consumption and production, as shown by correlation testing in the orange data mining application.

6. Acknowledgments

The authors wish to express their gratitude to all the individuals and organizations that contributed to the completion of this study. Special thanks are due to the Central Bureau of Statistics Province NTB, the Food Security Office of NTB Province, Bank Indonesia, the Central Bureau of Statistics Indonesia, and the Trade Office of NTB Province for providing the necessary data for this research. We also extend our appreciation to the researchers whose work has significantly informed and inspired this study, particularly W. G. M. Louhenapessy, L. Hasanah, and R. K. Paul et al, whose insights into agricultural economics and machine learning applications have been foundational. Additionally, we acknowledge the contributions of various scholars in the field of machine learning and agricultural forecasting, including M. Yağci, E. H. Bouras et al., and İ. Ercanlı, whose pioneering work has greatly enhanced the direction of this research. Their findings have enriched our understanding of the complex dynamics at play in rice price forecasting and agricultural data analysis. This research would not have been possible without their collective contributions.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

References

[1] W. G. M. Louhenapessy, “Beras Sebagai Barang Subtitusi dan Komoditi Pangan Utama,” EcceS (Economics, Soc. Dev. Stud., vol. 5, no. 1, p. 120, 2018, doi: 10.24252/ecc.v5i1.5240.

[2] L. Hasanah, “Analisis Faktor-Faktor Pengaruh Terjadinya Impor Beras di Indonesia Setelah Swasembada Pangan,” Growth J. Ilm. Ekon. Pembang., vol. 1, no. 2, pp. 57–72, 2022.

[3] R. K. Paul et al., "Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India," PLoS One, vol. 17, no. 7 July, pp. 1–17, 2022, doi: 10.1371/journal.pone.0270553.

[4] F. Moradi, S. M. M. Sadeghi, H. B. Heidarlou, A. Deljouei, E. Boshkar, and S. A. Borz, "Above-ground biomass estimation in a Mediterranean sparse coppice oak forest using Sentinel-2 data," Ann. For. Res., vol. 65, no. 1, pp. 165–182, 2022, doi: 10.15287/afr.2022.2390.

[5] E. H. Houssein, A. Hammad, and A. A. Ali, Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review, vol. 34, no. 15. Springer London, 2022. doi: 10.1007/s00521-022-07292-4.

[6] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," Smart Learn. Environ., vol. 9, no. 1, 2022, doi: 10.1186/s40561-022- 00192-z.

[7] J. Savoy, "Machine Learning Models," Mach. Learn. Methods Stylometry, pp. 109–151, 2020, doi: 10.1007/978-3-030-53360-1_6.

[8] E. H. Bouras et al., "Cereal yield forecasting with satellite drought- based indices, weather data and regional climate indices using machine learning in morocco," Remote Sens., vol. 13, no. 16, 2021, doi: 10.3390/rs13163101.

[9] C. Chen et al., "Forecast of rainfall distribution based on fixed sliding window long short-term memory," Eng. Appl. Comput. Fluid Mech., vol. 16, no. 1, pp. 248–261, 2022, doi: 10.1080/19942060.2021.2009374.

[10] N. Raj and J. Brown, "An eemd-bilstm algorithm integrated with boruta random forest optimiser for significant wave height forecasting along coastal areas of queensland, australia," Remote

Sens., vol. 13, no. 8, 2021, doi: 10.3390/rs13081456.

[11] İ. Ercanlı, "Innovative deep learning artificial intelligence applications for predicting relationships between individual tree height and diameter at breast height," For. Ecosyst., vol. 7, no. 1, 2020, doi: 10.1186/s40663-020-00226-3.

[12] T. D. Pham et al., "Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, Vietnam," Remote Sens., vol. 12, no. 8, pp. 1–24, 2020, doi: 10.3390/RS12081334.

[13] H. İnaç, Y. E. Ayözen, A. Atalan, and C. Ç. Dönmez, "Estimation of Postal Service Delivery Time and Energy Cost with E-Scooter by Machine Learning Algorithms," Appl. Sci., vol. 12, no. 23, 2022, doi: 10.3390/app122312266.

[14] M. Ali et al., "Coupled online sequential extreme learning machine model with ant colony optimization algorithm for wheat yield prediction," Sci. Rep., vol. 12, no. 1, pp. 1–23, 2022, doi: 10.1038/s41598-022-09482-5.

[15] M. Islam and F. Shehzad, "A Prediction Model Optimization Critiques through Centroid Clustering by Reducing the Sample Size, Integrating Statistical and Machine Learning Techniques for Wheat Productivity," Scientifica (Cairo)., vol. 2022, 2022, doi: 10.1155/2022/7271293.

[16] M. D'este, M. Elia, V. Giannico, G. Spano, R. Lafortezza, and G. Sanesi, "Machine learning techniques for fine dead fuel load estimation using multi-source remote sensing data," Remote Sens., vol. 13, no. 9, pp. 1–17, 2021, doi: 10.3390/rs13091658.

[17] M. Moeini, A. Shojaeizadeh, and M. Geza, "Supervised machine learning for estimation of total suspended solids in urban watersheds," Water (Switzerland), vol. 13, no. 2, 2021, doi: 10.3390/w13020147.

[18] A. J. DeSalvio, A. Adak, S. C. Murray, S. C. Wilde, and T. Isakeit, "Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms," Sci. Rep., vol. 12, no. 1, pp. 1– 14, 2022, doi: 10.1038/s41598-022-11591-0.

[19] P. Bannigan et al., "Machine learning models to accelerate the design of polymeric long-acting injectables," Nat. Commun., vol. 14, no. 1, pp. 1–12, 2023, doi: 10.1038/s41467-022-35343-w.

[20] E. Experimental, D. Using, and M. Learning, "Employing Experimental Data Using Machine Learning," 2022.

[21] M. C. V. David, "Rice Yield Modeling Using Machine Learning Algorithms Based on Environmental and Agronomic Data of Pampanga River Basin, Philippines," Univers. J. Agric. Res., vol. 11, no. 5, pp. 836–848, 2023, doi: 10.13189/ujar.2023.110509.