Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13666

e-ISSN : 2541-2019
p-ISSN : 2541-044X

A Comparative Analysis of Machine Learning Algorithms for Predicting Paddy Production

Nanda Aditya[1)*], Ibnu Rasyid Munthe[2)], Volvo Sihombing[3)]
[1)2)3)]Universitas Labuhanbatu, Indonesia
[1)]nandaaditya57@gmail.com, [2)]ibnurasyidmunthe@gmail.com, [3)]volvolumbantoruan@gmail.com

Abstract: For countries with large populations, such as Indonesia, food security is a very important issue. The majority of Indonesia's population depends on rice as their main food, and paddy is one of the most widely cultivated food commodities. The very good and accurate national paddy production prediction results really support decisions regarding national paddy production targets for the coming period. Therefore, to ensure supply and price stability, paddy availability must be predicted. Many studies have used machine learning to predict crop yields. By learning important patterns and relationships from input data, machine learning can combine the advantages of other methods to make better predictions of paddy yields. The aim of this research is to conduct a comparative analysis between three machine learning algorithms, namely, random forest, decision tree, and k-nearest neighbors, in predicting paddy production. To determine which algorithm is the best, a model evaluation is carried out using the coefficient of determination (R2-score), mean absolute error (MAE), and mean squared error (MSE). This research goes through methodological stages, starting from collecting datasets, data preprocessing, training and testing split datasets, applying algorithms, and evaluating the model. From this research, results were obtained for the random forest algorithm with an R2-score of 82.38%, MAE of 261726.20, and MSE of 2.19495E+11. For the decision tree, the R2-score was 79.62%, MAE was 323257.99, and MSE was 2.49304E+11. Meanwhile, k-nearest neighbors obtained an R2-score of 76.25%, MAE of 318433.42, and MSE of 2.90577E+11. The results of this research show that the random forest algorithm is the best for predicting paddy production because it obtains a larger R2-score as well as smaller MAE and MSE results.

Keywords: Decision Tree; K-Nearest Neighbors; Machine Learning; Paddy; Prediction; Random Forest.

INTRODUCTION

Food is the most basic need that must be met by everyone. Therefore, the state is obliged to ensure that every citizen receives the right to food (Nasution, Lubis, & Syaifuddin, 2020). The problem of food security is a very important issue for countries with high population levels, such as Indonesia (Sujarwo, Putra, Setyawan, Teixeira, & Khumairoh, 2022). One of the most widely cultivated food commodities in Indonesia is paddy, which is the main source of rice for the majority of the population (Khusna & Mariana, 2021). Rice is the commodity with the highest consumption participation rate, with 98.35% of households in Indonesia consuming rice as a staple food (Badan Pusat Statistik, 2022). Indonesia's rice consumption reached 35.3 million tons, making it the fourth-largest country in the world in terms of rice consumption (Muhamad, 2023). The high level of rice consumption in Indonesia shows that this commodity is very important, and the aspect of its availability is very vital.

National food sustainability requires controlling rice supplies. Rice commodity data is very strategic; therefore, it needs to be managed wisely to produce knowledge that can support targeted national food policies (Tosida, Wihartiko, Hermadi, Nurhadryani, & Feriadi, 2020). The problem is that paddy plants cannot produce continuously, even though agricultural technology is increasingly sophisticated. paddy production is greatly influenced by climatic conditions (Januarti, Yulian, & Erni, 2022). In addition, paddy production sometimes experiences a downward trend, which has an impact on increasing rice prices (Mardianto et al., 2018). Paddy production development planning is becoming increasingly strategic due to the large population and high levels of rice consumption (Putra & Ulfa Walmi, 2020). The decision regarding the national paddy production target for the coming period is strongly supported by the results of good and accurate national paddy production prediction (Zahra & Cahyadi, 2020). Therefore, paddy availability must be predicted to ensure supply and price stability (Airlangga, Rachmat, & Lapihu, 2019). By prediction rice production, the government can consider adopting policies to overcome obstacles that will continue to be faced by society for years (Wijaya, Furqon, & Marji, 2022).

* Corresponding author

Machine learning has been used in many studies to predict crop yields (Cedric et al., 2022; Morales & Villalobos, 2023; van Klompenburg, Kassahun, & Catal, 2020). Machine learning can combine the advantages of other methods, such as data-driven modeling, to make better crop yield predictions by learning important patterns and relationships from input data. Machine learning algorithms estimate functions that relate features or predictors to labels, such as crop yield (Paudel et al., 2021). Machine learning algorithms and computing tools extract critical information for outcome prediction and produce cost-effective and accurate estimates (Martini et al., 2022; Singh Boori, Choudhary, Paringer, & Kupriyanov, 2023; Zhi et al., 2022).

Various machine learning algorithms have also been applied by previous researchers to predict rice production. By applying an artificial neural network, research results show that the rice production prediction system, which consists of 75 tests in 19 regions in West Sumatra, has an accuracy rate of 88.14% (Putra & Ulfa Walmi, 2020). Applying the linear regression algorithm to predicting rice crop yields produces a forecasting model that is close to accurate, with a reliability level of 94.51% (Herwanto, Widiyaningtyas, & Indriana, 2019). The prediction model applied by (Jiya, Illiyasu, & Akinyemi, 2023) uses five machine learning algorithms, and the results show that the Random Trees algorithm has the highest accuracy compared to other algorithms. Researchers such as (Choudhary, Shi, Dong, & Paringer, 2022) predict crop yields using a combination of various secondary data. The results prove that the random forest algorithm has an accuracy of more than 85%. The random forest approach was also implemented by (Jiya et al., 2023) in predicting rice yields, resulting in a relative root mean square error value of less than 5%. A hybrid approach combining random forest and neural network algorithms for predicting rice yields has a prediction accuracy of 98% (Lingwal, Bhatia, & Singh, 2024). In research conducted by (Patil, Panpatil, & Kokate, 2022), the prediction results of the K-Nearest Neighbors algorithm outperformed Decision Tree with an accuracy rate of 89.4%. In research conducted by (Cedric et al., 2022), it was found that the decision tree and K-nearest neighbor algorithms showed better results compared to logistic regression with a coefficient of determination (R2) above 90%. Likewise, carried out by (Morales & Villalobos, 2023), the results show that random forest performance is better compared to artificial neural networks and linearly regularized.

In applying machine learning techniques, it is important to choose an algorithm that has good performance in making predictions. From several studies that have been described previously, there are three algorithms that are superior in predicting rice yields, namely: decision tree, k-nearest neighbor, and random forest. So this research aims to carry out a comparative analysis of the three machine learning algorithms for predicting rice production. To find out which algorithm is the best, a model evaluation will be carried out using the coefficient of determination (R2-score), mean absolute error (MAE), and mean squared error (MSE). The smaller the MAE and MSE values and the greater the R2-score, then the algorithm is the best algorithm.

METHOD

This research requires a method, which is a research procedure stage in the form of structured and systematic steps. The method used in this research is the result of a combination of several previous studies. In research conducted by (Cedric et al., 2022) and (Jiya et al., 2023), there is a need to carry out data preprocessing by applying feature engineering. It is also necessary to divide the dataset into two parts, namely training data and test data (Paudel et al., 2021). For this reason, the research steps are presented through the design shown in Figure 1.

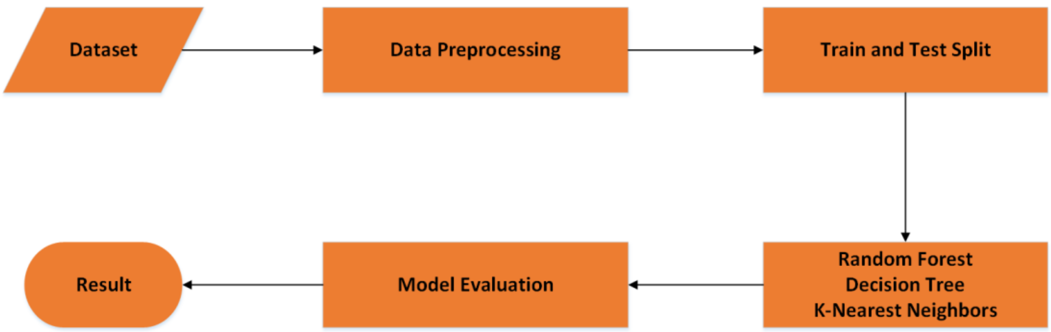


Fig 1. The Research Method Flow

Dataset

The data used in this research is a dataset of rice plants originating from eight provinces on the island of Sumatra, namely Aceh, North Sumatra, Riau, Jambi, South Sumatra, Bengkulu, and Lampung. The dataset was collected via the BPS website (Statistik, 2024). The data used for the rice dataset is from 2000 to 2020, which includes annual harvest area and annual production results. The harvest area variable is used because it can influence the amount of rice production (Nurzannah, Girsang, & El Ramija, 2020; Zahra & Cahyadi, 2020). Apart from the rice plant dataset, climate change data was also used, which included rainfall data (Statistik, 2022a),

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13666
e-ISSN : 2541-2019
p-ISSN : 2541-044X

humidity (Statistik, 2022b), and average temperature (Statistik, 2022c) from 2000 to 2020 obtained via the BPS website. The climate change variable is used because it can influence crop yields (Jiya et al., 2023; Xu et al., 2019).

Table 1. A part of Paddy Dataset

Province	Year	Production (ton)	Harvest Area (Ha)	Rainfall (mm)	Humidity (%)	Average Temperature (°C)
Aceh	2020	1861567.1	317869.41	1619.2	80.82	25.41
North Sumatera	2020	2076280.01	388591.22	1648.3	83.76	25.79
West Sumatera	2020	1450839.74	295664.47	4072.7	85.22	26.26
Riau	2020	269344.05	64733.13	2584.9	81.8	25.93
Jambi	2020	374376.27	84772.93	2303.8	78.24	25.85
South Sumatera	2020	2696877.46	551320.76	2300.2	81.1	26.14
Bengkulu	2020	296925.16	64137.28	4144	80.18	25.84
Lampung	2020	2604913.29	545149.05	2211.3	75.8	24.58

Table 1 shows part of the dataset from the total data, totaling 168 rows for each province over a period of 21 years. The dataset consists of seven features, namely, province, year, production, harvest area, rainfall, humidity, and average temperature. Province is the name of the provinces on the island of Sumatra. The year is the year of rice production. Production is the result of annual rice production. The harvest area is the area of crops that can be harvested. Rainfall is the average amount of rainfall in a year. Humidity is the average annual humidity level. The average temperature is the average temperature in a year.

Data Preprocessing

At this stage, the coding of the features in the dataset will be carried out. Categorical data on the province feature will be converted into numeric data. In this process, a one-hot encoding method is used, which will convert category features into numerical form so that the machine learning algorithm can make better predictions. One-hot encoding will convert the province column into a numeric array.

Train and Test Split

The dataset will be divided into two parts at this stage: the training dataset and the testing dataset. The training dataset uses 70%, and the testing dataset uses 30%. The test dataset is used to assess how well the machine learning algorithm is trained on the training dataset. Meanwhile, the training dataset is used to train the machine learning algorithm to make correct predictions.

Machine Learning

Three machine learning algorithm models (random forest, decision tree, and k-nearest neighbors) will be used in this process. To improve model performance based on the R2 score, hyperparameter tuning will be applied to each model. Grid Search Cross Validation or Randomized Search Cross Validation will be used, depending on the algorithm. A resampling method known as cross-validation is used to test machine learning models on limited data samples. The parameter k, which will be divided into specific data samples, is the only parameter used in this process.

Model Evaluation

Next, an evaluation will be carried out to select the best algorithm from the three that will be selected. To assess the model, the R2 score, mean absolute error, and mean squared error were used. The R2 score indicates how well the data points fit the curve or line. The mean absolute error measures the average of the residuals in a data set. The mean squared error measures the variance of the residuals. The smaller the mean absolute error and mean squared error values, but the greater the R2 score, the better the algorithm.

RESULT

This section will present the results of rice production predictions based on comparing the results of three machine learning algorithm models. All work, from processing the dataset to the results, was carried out using the Google Colab text editor in the Python programming language. The research results are described thoroughly using the research procedures described in the previous methodology section.

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13666
e-ISSN : 2541-2019
p-ISSN : 2541-044X

	Year	Production	Harvest Area	Rainfall	Humidity	Average Temperature
count	168.000000	1.680000e+02	168.000000	168.000000	168.000000	168.000000
mean	2010.000000	1.772174e+06	375287.420179	2405.413869	80.148631	27.053690
std	6.073403	1.225588e+06	238520.849168	1017.262403	5.199196	1.209537
min	2000.000000	4.293800e+04	63142.040000	222.500000	54.200000	22.190000
25%	2005.000000	5.535722e+05	145041.250000	1684.375000	78.022500	26.495000
50%	2010.000000	1.804400e+06	373551.500000	2308.800000	81.400000	26.970000
75%	2015.000000	2.603776e+06	546691.977500	3049.450000	83.855000	27.525000
max	2020.000000	4.881089e+06	872737.000000	5228.000000	90.600000	29.850000

Fig 1. Dataset Descriptive Statistics

From Figure 1, it can be seen that, for eight provinces on the island of Sumatra for 21 years, the average rice production yield was 1,772,173.70 metric tons, with the lowest yield being 42,938 metric tons and the highest yield being 4,881,089 metric tons. The average rice harvest area is 375,287.42 hectares. Average rainfall is 2,405.41 millimeters. The average humidity level is 80.14%, and the average temperature is 27.05 Celsius.

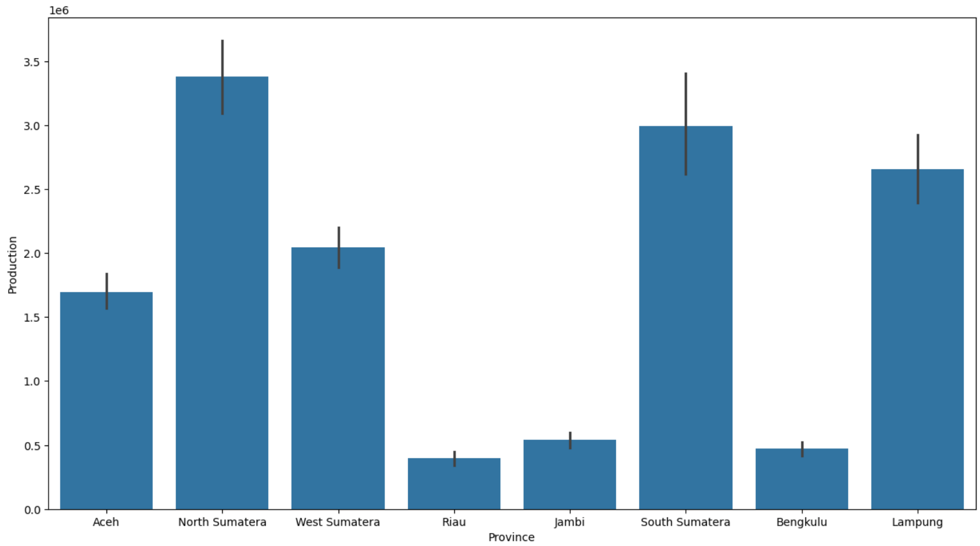


Fig 2. Rice Production in Sumatra

Figure 2 shows the average production dataset for each province on the island of Sumatra. It can be seen that the province of North Sumatra has the highest rice production results, with an average annual production of 3,380,909.65 tons. Meanwhile, the province producing the smallest rice production is Riau, with an average annual rice production of 398,03.02 tons.

	Harvest Area	Rainfall	Humidity	Average Temperature	Production	Province_Aceh	Province_Bengkulu	Province_Jambi	Province_Lampung	Province_North Sumatera	Province_Riau	Province_South Sumatera	Province_West Sumatera
163	390799.00	2317.6	79.40	26.45	3831923.00	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
164	396559.00	1825.1	77.04	26.36	4090654.00	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
165	511940.93	1385.8	76.05	25.50	2488641.91	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
166	464103.42	1706.4	78.03	27.23	2164089.33	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
167	545149.05	2211.3	75.80	24.58	2604913.29	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0

Fig 3. Data Preprocessing Results

Figure 3 shows the results of data preprocessing on the dataset. The province column, which originally contained categorical province names, has now changed to numeric data with a value of 0 or 1.

* Corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13666
e-ISSN : 2541-2019
p-ISSN : 2541-044X

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3, random_state=0)
print('x_train :',x_train.shape)
print('x_test :',x_test.shape)
print('y_train :',y_train.shape)
print('y_test :',y_test.shape)

x_train : (117, 12)
x_test : (51, 12)
y_train : (117, 1)
y_test : (51, 1)
```

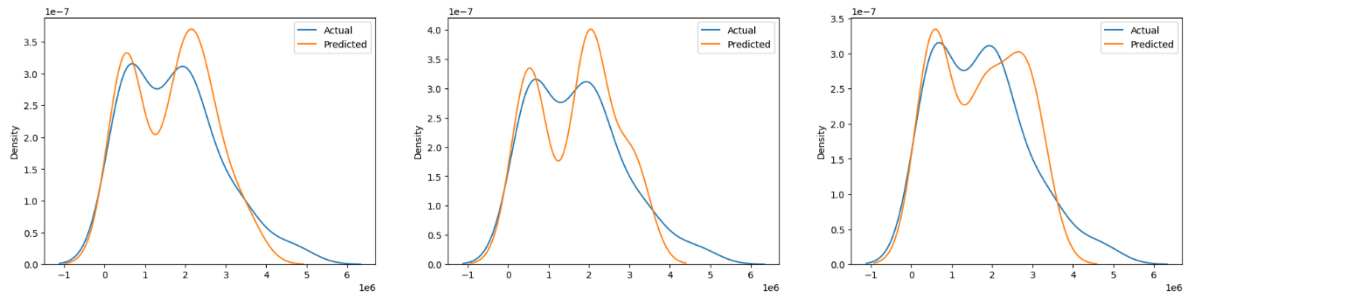
Fig 4. Train and Test Split

Figure 4 shows the results of the process of dividing training data by test data. By applying a 30% distribution of test data (test_size = 0.3) from the total dataset of 168 records, 117 records of training data were produced, and 51 records of test data were produced. The results of this data sharing will be applied to each machine learning algorithm.

Table 2. The Average Cross Validation Score

Model	before Hyperparameter tuning (%)	after Hyperparameter tuning (%)
Random Forest	84.71	85.83
Decision Tree	80.11	82.76
K-Nearest Neighbors	67.0	77.29

From table 2, it can be seen that there was an increase in the average cross-validation score after hyperparameter tuning was carried out. In the random forest algorithm, there was an increase in score of 1.12%. In the decision tree algorithm, there was an increase in score of 2.65%. Meanwhile, in the k-nearest neighbors algorithm, the score increased by 10.29%.



Random Forest *Decision Tree* *K-Nearest Neighbors*

Fig 5. Prediction Results

Figure 5 shows a graphical plot of the prediction results of each algorithm with the original data after hyperparameter tuning. In the graph, it can be seen that the level of conformity between the prediction results and the original data is shown by the random forest algorithm, followed by the k-nearest neighbors algorithm, and the decision tree.

Table 3. Model Evaluation

Model	R2-score (%)	MAE	MSE
Random Forest	82.38	261726.20	2.19495E+11
K-Nearest Neighbors	79.62	323257.99	2.49304E+11
Decision Tree	76.25	318433.42	2.90577E+11

Based on the model evaluation results shown in Table 3, the random forest algorithm model has the highest R2-score, namely, 82.38%. Meanwhile, the lowest R2-score is the decision tree algorithm. The random forest algorithm has the random forest algorithm has the smallest mean absolute error (MAE) value. Meanwhile, the algorithm with the algorithm with the largest MAE value is the k-nearest neighbors algorithm. The smallest MSE value is owned by the random forest algorithm. Meanwhile, the algorithm with the algorithm with the largest MSE value is the decision tree algorithm.

* Corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

This research has successfully implemented three machine learning algorithms for predicting rice production. From a dataset of 168 records and 7 features, not a single one has missing values, so there is no need to handle missing values. The overall distribution of data on the features used tends to be normal, and there are no outliers from the output features, namely production. Agricultural harvest area and average temperature have a positive correlation, meaning the greater the value of the independent variable, the greater the production, although this correlation is not significant. On the other hand, rainfall and humidity have a negative correlation, which means that the smaller the value of these two variables, the greater the production. At the data preprocessing stage, the one-hot encoding method has been successful in converting categorical data into numerical data. So the provincial features, which originally numbered one, were split into eight features according to the number of existing provinces. 70% of the test data in the dataset is able to make the prediction model better. From the results of applying hyperparameters to each algorithm, an increase in the percentage of cross-validation was obtained. This proves that hyperparameter tuning has a big influence on the average percentage of cross-validation and improvement in prediction results.

This research has limitations that can later be refined by further research. The dataset used only represents eight provinces on the island of Sumatra. In the future, it is necessary to use a dataset that represents all provinces in Indonesia. This is useful for gaining a more comprehensive understanding of rice production in Indonesia. The selection of features that influence rice production results needs to be developed further by adding other features that can influence rice production results, such as soil type (van Klompenburg et al., 2020), soil moisture (Anjana et al., 2021), wind data, irrigation data, and pollution data (Cedric et al., 2022). This was done to see a deeper correlation regarding the factors that influence rice production results. In the context of implementing cross-validation, further development is needed by applying a different number of k-folds. In this study, only k = 10 was used. This is useful for seeing the effect of the number k on the prediction accuracy results. This research uses R2, MAE, and MSE as evaluation matrices for machine learning algorithm test results; other matrices can be added to measure the accuracy of predictions by calculating the prediction rate and false prediction rate (Anjana et al., 2021).

CONCLUSION

From the research results described previously, a comparative analysis has been carried out on the random forest, decision tree, and k-nearest neighbors algorithms for predicting rice production. The random forest algorithm obtained a greater R2-score than other algorithms, namely 82.38%. With an MAE value of 261726.20 and an MSE value of 2.19495E+11, where the MAE and MSE values are smaller than the decision tree algorithm and the k-nearest neighbors algorithm. From the results of this research, it can be concluded that, in predicting rice production based on the dataset used in this research, the random forest algorithm is the best algorithm.

REFERENCES

Airlangga, G., Rachmat, A., & Lapihu, D. (2019). Comparison of exponential smoothing and neural network method to forecast rice production in Indonesia. *TELKOMNIKA*, *17*(3), 1367–1375. https://doi.org/10.12928/TELKOMNIKA.v17i3.11768

Anjana, K, A. K., Sana, A., Bhat, B. A., Kumar, S., & Bhat, N. (2021). An efficient algorithm for predicting crop using historical data and pattern matching technique. *Global Transitions Proceedings*, *2*(2), 294–298. https://doi.org/10.1016/j.gltp.2021.08.060

Badan Pusat Statistik. (2022). Pengeluaran untuk Konsumsi Penduduk Indonesia. In *bps.go.id*.

Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M., & Kimpolo, C. L. M. (2022). Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agricultural Technology*, *2*(March). https://doi.org/10.1016/j.atech.2022.100049

Choudhary, K., Shi, W., Dong, Y., & Paringer, R. (2022). Random Forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine. *Advances in Space Research*, *70*(8), 2443–2457. https://doi.org/https://doi.org/10.1016/j.asr.2022.06.073

Herwanto, H. W., Widiyaningtyas, T., & Indriana, P. (2019). Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *8*(4). Retrieved from https://journal.ugm.ac.id/v3/JNTETI/article/view/2563

Januarti, I., Yulian, J., & Erni, P. (2022). Forecasting Production and Consumption of Rice and Influence of Determinants To Increase Food Security in the South Sumatra Region, Indonesia. *Russian Journal of Agricultural and Socio-Economic Sciences*, *121*(1), 144–156. https://doi.org/10.18551/rjoas.2022-01.17

Jiya, E. A., Illiyasu, U., & Akinyemi, M. (2023). Rice Yield Forecasting: A Comparative Analysis of Multiple Machine Learning Algorithms. *Journal of Information Systems and Informatics*, *5*(2), 785–799. https://doi.org/10.51519/journalisi.v5i2.506

Khusna, I. M., & Mariana, N. (2021). Sistem Pendukung Keputusan Pemilihan Bibit Padi Berkualitas Dengan

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13666
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Metode AHP Dan Topsis (Study Kasus Desa Sambongbangi). *Jurnal Sisfokom (Sistem Informasi Dan Komputer), 10*(2), 162–169. https://doi.org/10.32736/sisfokom.v10i2.1145

Lingwal, S., Bhatia, K. K., & Singh, M. (2024). A novel machine learning approach for rice yield estimation. *Journal of Experimental & Theoretical Artificial Intelligence*, *36*(3), 337–356. https://doi.org/10.1080/0952813X.2022.2062458

Mardianto, M. F. F., Tjahjono, E., Rifada, M., Herawanto, A., Putra, A. L., & Utama, K. A. (2018). The Prediction of Rice Production in Indonesia Provinces for Developing Sustainable Agriculture. *The First International Conference of Food and Agriculture*, (0), 325–333. Retrieved from https://doi.org/10.1016/j.rse.2021.112679%0Ahttps://doi.org/10.1080/08839514.2023.2175113%0Ahttp://dx.doi.org/10.1371/journal.pone.0283452%0Ahttps://journal.agrimetassociation.org/index.php/jam/article/view/75%0Ahttps://journal.agrimetassociation.org/inde

Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., … Omodei, E. (2022). Machine learning can guide food security efforts when primary data are not available. *Nature Food*, *3*(9), 716–728. https://doi.org/10.1038/s43016-022-00587-8

Morales, A., & Villalobos, F. J. (2023). Using machine learning for crop yield prediction in the past or the future. *Frontiers in Plant Science*, *14*(March), 1–13. https://doi.org/10.3389/fpls.2023.1128388

Muhamad, N. (2023). *Konsumsi Beras Indonesia Terbanyak Keempat di Dunia pada 2022/2023*. Retrieved from https://databoks.katadata.co.id/datapublish/2023/07/13/konsumsi-beras-indonesia-terbanyak-keempat-di-dunia-pada-20222023

Nasution, F., Lubis, Y., & Syaifuddin. (2020). Peranan Kinerja Penyuluh Pertanian Terhadap Peningkatan Produktivitas Padi Sawah di Kabupaten Labuhanbatu Utara. *AGRISAINS: Jurnal Ilmiah Magister Agribisnis*, *2*(2), 116–128.

Nurzannah, S. E., Girsang, M. A., & El Ramija, K. (2020). FAKTOR-FAKTOR YANG MEMPENGARUHI PRODUKSI PADI SAWAH (Oryza sativa L.) DI KABUPATEN SERDANG BEDAGAI. *Jurnal Pengkajian Dan Pengembangan Teknologi Pertanian*, *23*(1), 11–24.

Patil, P., Panpatil, V., & Kokate, S. (2022). Crop Prediction System using Machine Learning Algorithms. *International Research Journal of Engineering and Technology (IRJET)*, *07*(02), 33–41. https://doi.org/10.1007/978-981-19-2069-1_3

Paudel, D., Boogaard, H., Wit, A. De, Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, *187*(October 2020), 103016. https://doi.org/10.1016/j.agsy.2020.103016

Putra, H., & Ulfa Walmi, N. (2020). Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *6*(2), 100–107. https://doi.org/10.25077/teknosi.v6i2.2020.100-107

Singh Boori, M., Choudhary, K., Paringer, R., & Kupriyanov, A. (2023). Machine learning for yield prediction in Fergana valley, Central Asia. *Journal of the Saudi Society of Agricultural Sciences*, *22*(2), 107–120. https://doi.org/https://doi.org/10.1016/j.jssas.2022.07.006

Statistik, B. P. (2022a). Jumlah Curah Hujan dan Jumlah Hari Hujan di Stasiun Pengamatan BMKG. Retrieved March 18, 2024, from bps.go.id website: https://www.bps.go.id/id/statistics-table/1/MTk1OSMx/jumlah-curah-hujan-dan-jumlah-hari-hujan-di-stasiun-pengamatan-bmkg-2016-2020.html

Statistik, B. P. (2022b). Kecepatan Angin dan Kelembaban di Stasiun Pengamatan BMKG. Retrieved March 18, 2024, from bps.go.id website: https://www.bps.go.id/id/statistics-table/1/MTk2MCMx/kecepatan-angin-dan-kelembaban-di-stasiun-pengamatan-bmkg-2011-2015.html

Statistik, B. P. (2022c). Suhu Minimum, Rata-Rata, dan Maksimum di Stasiun Pengamatan BMKG (oC). Retrieved March 18, 2024, from bps.go.id website: https://www.bps.go.id/id/statistics-table/1/MTk2MSMx/suhu-minimum--rata-rata--dan-maksimum-di-stasiun-pengamatan-bmkg--oc---2011-2015.html

Statistik, B. P. (2024). Luas Panen, Produksi, dan Produktivitas Padi Menurut Provinsi. Retrieved March 18, 2024, from bps.go.id website: https://www.bps.go.id/id/statistics-table/2/MTQ5OCMy/luas-panen--produksi--dan-produktivitas-padi-menurut-provinsi.html

Sujarwo, Putra, A. N., Setyawan, R. A., Teixeira, H. M., & Khumairoh, U. (2022). Forecasting Rice Status for a Food Crisis Early Warning System Based on Satellite Imagery and Cellular Automata in Malang, Indonesia. *Sustainability*, *14*(15). https://doi.org/10.3390/su14158972

Tosida, E. T., Wihartiko, F. D., Hermadi, I., Nurhadryani, Y., & Feriadi. (2020). Model Manajemen Big Data Komoditas Beras untuk Kebijakan Pangan Nasional. *Jurnal RESTI (Rekayasa Sistem Informasi)*, *4*(1), 142–154.

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*(August), 105709. https://doi.org/10.1016/j.compag.2020.105709

Wijaya, D. Y., Furqon, M. T., & Marji. (2022). Peramalan Jumlah Produksi Padi Menggunakan Metode

* Corresponding author

DOI : https://doi.org/10.33395/sinkron.v8i2.13666
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Backpropagation. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *6*(3), 1129–1137.

Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., … Wu, X. (2019). Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China. *Ecological Indicators*, *101*, 943–953. https://doi.org/https://doi.org/10.1016/j.ecolind.2019.01.059

Zahra, N., & Cahyadi, E. R. (2020). Forecasting and Establishing National Rice Production Targets in Indonesia. *Proceedings of the 1st International Conference on Sustainable Management and Innovation, ICoSMI*. https://doi.org/10.4108/eai.14-9-2020.2304658

Zhi, J., Cao, X., Zhang, Z., Qin, T., Qu, L., Qi, L., … Fu, X. (2022). Identifying the determinants of crop yields in China since 1952 and its policy implications. *Agricultural and Forest Meteorology*, *327*, 109216. https://doi.org/https://doi.org/10.1016/j.agrformet.2022.109216

* Corresponding author