Jurnal Sifo Mikroskil (JSM)
Volume 27, No 1, April 2026 – Hal 165 - 180
DOI: https://doi.org/10.55601/jsm.v27i1.2017

e-ISSN: 2622-8130
ISSN: 1412-0100

Comparison of Machine Learning Methods with Optimization for Paddy Production Prediction

Roni Yunis[1], Irpan Adiputra Pardosi[2]
[1]Department of Information System, Universitas Mikroskil, Medan, Indonesia
[2]Department of Informatics, Universitas Mikroskil, Medan, Indonesia
e-mail: [1]roni@mikroskil.ac.id, [2]irpan@mikroskil.ac.id

Dikirim: ***16-04-2026*** *| Diterima:* ***29-04-2026*** *| Diterbitkan:* ***30-04-2026***

Abstract

Food security remains a critical challenge in Indonesia, where rice serves as the primary staple and demand continues to rise with population growth. Fluctuations in paddy production pose significant risks to supply stability, highlighting the need for accurate and reliable forecasting models. This study presents a comparative evaluation of Random Forest and Support Vector Regression (SVR) for paddy production prediction using national-level production data aggregated from provincial statistics in Indonesia, incorporating Grid Search and Random Search to optimize model performance. Experimental results demonstrate that SVR optimized with Random Search achieves superior predictive accuracy, yielding an RMSE of 27,478.58 and a MAPE of 0.05%, indicating both low absolute and relative errors. This performance suggests that SVR is more effective in modeling the non-linear and continuous dynamics inherent in paddy production data. Furthermore, Random Search consistently outperforms Grid Search, reflecting its ability to efficiently explore complex hyperparameter spaces. These findings underscore the critical role of both model selection and optimization strategy in improving forecasting reliability. The proposed approach provides a robust framework for data-driven agricultural planning and offers practical value for policymakers in managing food supply, reducing uncertainty, and enhancing national food security.

***Keywords**: paddy production prediction, SVR, random forest, hyperparameter optimization, random search*

1. INTRODUCTION

Food is a basic human need that must be fulfilled by everyone [1]. Ensuring food availability is essential to maintain social security, political stability, and national resilience [2]. Therefore, governments continuously strive to guarantee adequate food supply so that people can meet their daily needs [3]. In Indonesia, food plays a particularly critical role in household expenditure, accounting for the largest proportion of total spending [4]. Recent data from Badan Pusat Statistik indicate that food expenditure consistently represents approximately 56–58% of total household expenditure, highlighting the dominant role of food consumption in the economic structure and emphasizing the importance of strengthening food security and agribusiness development as key priorities in agricultural and regional policies [5]. Consequently, improving food security and developing agribusiness remain essential components of national agricultural development strategies [6].

Paddy is the most important food commodity in Indonesia, as it is consumed by the majority of the population [7]. National paddy consumption continues to increase in line with population growth [1]. However, fluctuations in paddy production can pose significant risks to food availability and supply stability. Inaccurate estimation of production levels may lead to supply shortages, price instability, and ineffective policy interventions. Therefore, the development of accurate paddy production prediction systems is essential to support strategic decision-making and improve agricultural planning.

In this context, machine learning models have emerged as effective approaches for predicting agricultural production due to their ability to capture complex and non-linear relationships. Several commonly used models include Decision Tree, Random Forest, Support Vector Regression (SVR), K-Nearest Neighbors (KNN), and Naive Bayes. This study focuses on comparing two widely used models, namely Random Forest and SVR, in predicting paddy production. Random Forest is known for its capability to handle complex data structures and reduce overfitting [8], while SVR is effective in modeling both small and large datasets with strong generalization performance [9].

Several previous studies have demonstrated the effectiveness of Random Forest (RF) and SVR models in prediction tasks. Researcher [10] applied Random Forest Regression (RFR) to forecast parameters in E-jet printing and achieved low error rates, indicating high prediction accuracy. Similarly, researcher [11] showed that RF outperforms logistic regression and ultrasound methods in predicting macrosomia, achieving high sensitivity and specificity. In addition, researcher [12] applied a hierarchical Random Forest-based approach to predict cardiovascular disease with high accuracy.

In contrast, researcher [9] demonstrated that Support Vector Regression (SVR) provides accurate predictions in laser tube bending processes, achieving low error values and high determination coefficients. Furthermore, researcher [13] applied SVR to predict instant noodle demand and reported improved performance compared to baseline models. Researcher [14] also compared SVR with ANFIS in predicting the performance of aerobic granular sludge reactors and found that SVR achieved significantly better accuracy. Although previous studies have shown that both Random Forest and SVR are capable of producing accurate predictions, most existing research focuses on individual model performance without systematically evaluating the impact of different hyperparameter optimization strategies. In particular, the comparison between Grid Search and Random Search in improving model performance for paddy production prediction remains limited.

Therefore, this study aims to compare the performance of Random Forest and Support Vector Regression models in predicting paddy production by incorporating two optimization techniques, namely Grid Search and Random Search. The contribution of this study lies in providing a systematic evaluation of model performance under different optimization strategies, as well as offering practical insights for selecting appropriate predictive models in agricultural applications. The results are expected to support more accurate decision-making and contribute to improving paddy production and food security in Indonesia.

2. LITERATURE REVIEW

2.1 OSEMN

This study applies the OSEMN method, which is an approach in data science. OSEMN consists of a series of sequential steps, including data collection (Obtain), data cleansing (Scrub), data exploration (Explore), model building (Model), and result interpretation (iNterpret) [15]. Figure 1 below shows the stages of OSEMN.

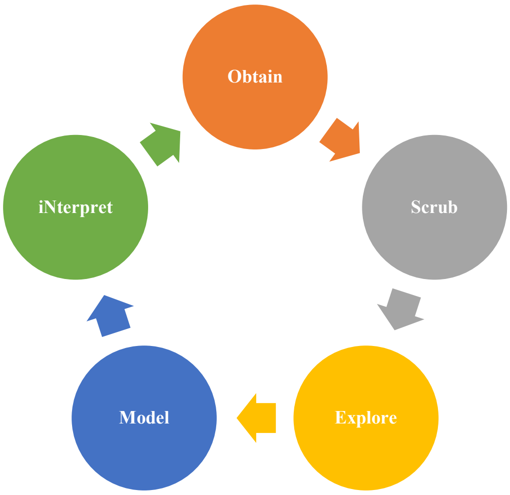

Figure 1. OSEMN Methods

a. Obtain
The first stage in the OSEMN process is "Obtaining Data", which refers to the initial step of collecting all the data relevant to the set goal or target. This can be done using various sources such as databases, APIs, or web scraping [16].

b. Scrub
Scrub in the OSEMN method refers to data cleaning and preprocessing steps to ensure that the data is appropriate and ready for analysis. This involves removing duplicates, handling missing values, and adjusting the format of the data to make it usable. This scrub process is very important as it helps to ensure that the quality of the data is maintained high and any errors or inconsistencies are corrected before analysis [16].

c. Explore
Data exploration is the process of investigating data before moving to Machine Learning (ML) modeling [16]. This stage involves a comprehensive examination of the data, focusing on all its characteristics. This process is known as Exploratory Data Analysis (EDA), where we analyze data cleanliness, readiness for modeling, and the types of features that need to be investigated using statistical methods [16].

d. Model
After completing the previous three steps in the OSEMN process, the next step is the data modeling stage. At this stage, the main goal is to create a model that can accurately predict the results by utilizing unique patterns and trends in the data. In this study, two types of models, namely Random Forest and SVR, were used to forecast paddy production.

e. Interpret
Data interpretation is the final step in the OSEMN method. Interpretation refers to the presentation of data to answer business questions and provide insights that can be implemented through data science [16]. The process of interpreting data in Random Forest and SVR models involves comparing the performance evaluation results of these two models to determine the more accurate one.

2.2 Random Forest

Random Forest (RF) or random decision forests are ensemble learning methods for classification, regression, and other tasks. Random Forest operates by building multiple decision trees at different training times, and outputting a class that represents the classification or regression mode of each tree. The Random Forest algorithm incorporates the established classification and regression trees (CARTs) technique. Each CART is constructed using a random subset of the features.

DOI: https://doi.org/10.55601/jsm.v27i1.2017

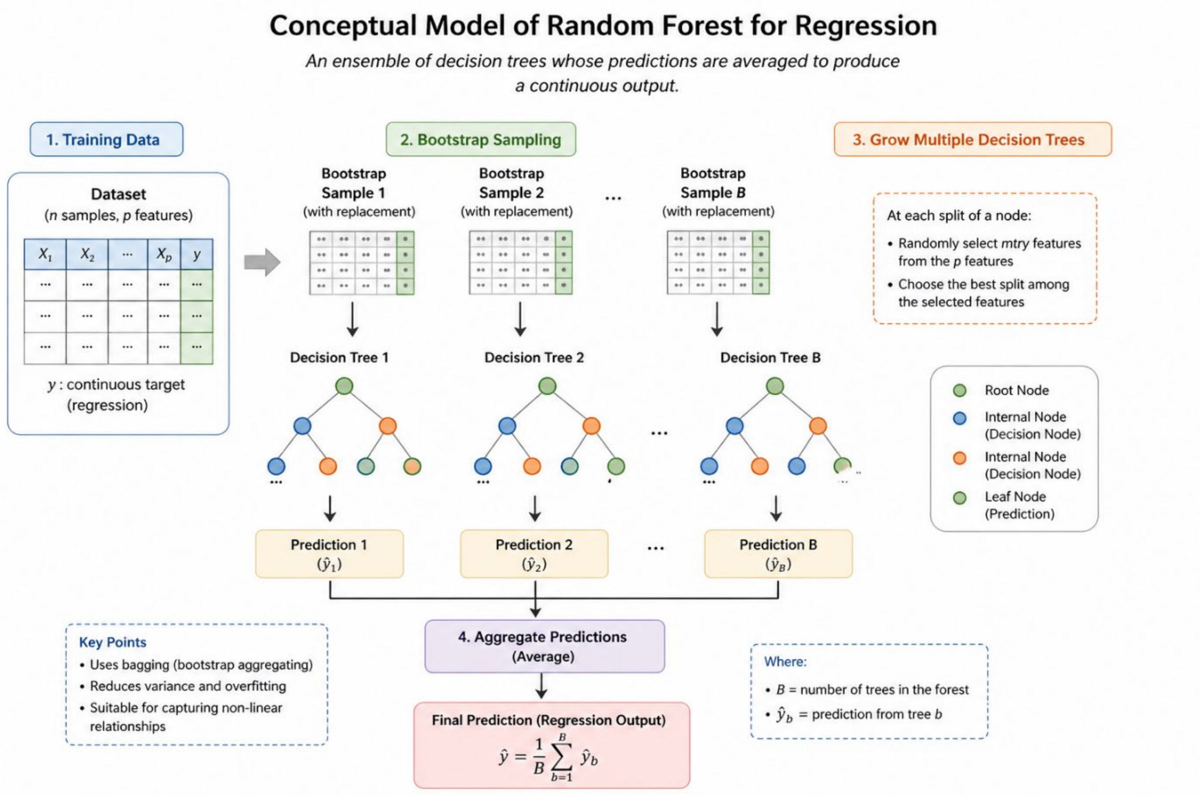

Figure 2. Random Forest Regression (RFR)

For a Random Forest classifier model, the main parameters are the number of decision trees and the number of features used at each node as the tree grows. During model training, the number of decision trees is determined in advance. More trees are generally better, but take longer to calculate. A lower number of features leads to a greater reduction in variance. The number of features can be defined using the empirical formula $N_f = \sqrt{M}$, where M denotes the total number of features [17]. Random Forest can be applied to both classification and regression problems, depending on the type of tree used whether it is a classification or regression tree. Assuming that the model consists of T regression trees to perform regression prediction (Figure 2), the final result of the Random Forest regression model is as follows [17]:

$$H(x) = \frac{1}{T} - \sum_{i=1}^{T} h_i(x) \tag{1}$$

In this case, T is the number of regression trees, and $h_i(x)$ is the output of the i-th regression tree on sample x. Thus, the prediction of Random Forest is the average value of all the tree predictions.

2.3 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a concept derived from machine learning theory, which was previously used to solve classification problems with Support Vector Machine (SVM). In a simpler context, SVM is used to categorize data, while SVR is used to make numerical predictions or regressions (Figure 3). The main goal of SVR is to find a function that fits the input data with the minimum amount of error possible [18]. In other words, SVR aims to provide highly accurate predictions with the smallest possible error rate when analyzing regression data. The regression function of the SVR model can be written as follows [19].

$$f(x) = w\,\varphi(x) + b \tag{2}$$

2622-8130 Jurnal Sifo Mikroskil (JSM) - Volume 27, No 1, April 2026
DOI: https://doi.org/10.55601/jsm.v27i1.2017

Description:
w : Weight vector
φ(x) : Function that maps x in a dimension
b : Bias

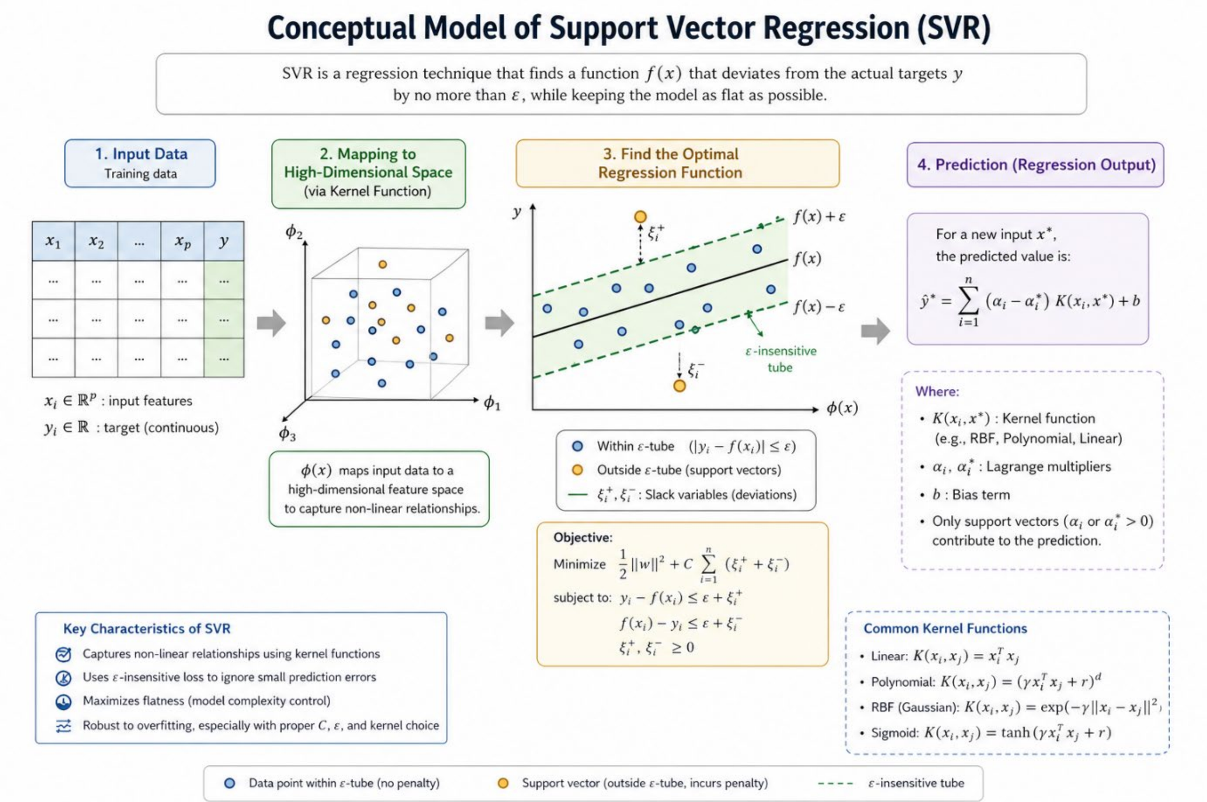

Figure 3. Support Vector Regression (SVR)

In handling non-linear problems using the SVR algorithm, we use a function called Kernel. Kernel in data mining is an inner product function that operates in the feature space [20]. Its function is to allow solving complex problems in higher dimensions without having to directly calculate detailed mappings. Several types of Kernel functions are commonly used, namely [19]:

a. Radial Basic Function (RBF) Kernel

The Radial Basis Function (RBF) kernel is widely chosen because of its ability to understand complex relationships in non-linear data. This kernel is suitable for irregular data and is well known for its prediction accuracy in applications such as exchange rate prediction [21], carbon price [22], and cantilever retaining wall stability [23]. RBF kernels are very effective in capturing non-linear patterns and have shown high prediction accuracy in various studies, even reaching 95.94% accuracy in exchange rate prediction [21].

$$K(x_i, x_j) = \exp(-y(x_i, x_j)^2) \tag{3}$$

Here, *y* is a free parameter that determines how much influence two points have on each other. RBF expands towards infinite dimensions as it undergoes rapid exponential growth [21].

b. Linear Kernel

Linear kernels are the most basic among all types of kernels. In its use, there is no technical projection of data to higher dimensions. This kernel only involves the product of *x* and *y* with a constant value *c*, without the addition of new dimensions. The advantage of the Linear Kernel is that it is very simple and has a constant parameter *c* as the only component [21].

$$K(x_i, x_j) = x_i, x_j + c \tag{4}$$

c. Polynomial Kernel

Unlike the Linear Kernel, the polynomial Kernel involves multiplying the results of the representation in a higher dimensional space. The polynomial kernel formula can be expressed as

$$K(x_i, x_j) = (\alpha\, x_i, x_j\ c)^d \tag{5}$$

where there are three parameters: α, c, and d. The most common degree (d) used is 2 as larger degrees can lead to overfitting [21].

d. Sigmoid Kernel

The sigmoid kernel is used to capture non-Linear relationships and is suitable for data with non-Linear patterns. This kernel has been used in SVR for applications such as stock price prediction [24].

$$K(x_i, x_j) = \tanh(y(x_i, x_j) + \theta) \tag{6}$$

Here, y is the slope and θ is the intercept constant [24].

2.4 Hyper-Parameter Tuning

Hyperparameter tuning is the process of selecting optimal values for hyperparameters of machine learning algorithms to improve their performance. Hyperparameters are parameters that are not learned from the data, but rather are set by the user before training the model. Hyperparameter tuning is very important because it can significantly affect the model's ability to generalize and make accurate predictions. The goal of hyperparameter tuning is to find the best combination of hyperparameter values that maximizes model performance on the validation set. There are three methods in hyperparameter tuning [25]:

a. Grid Search

Grid Search is a method of determining the combination of models and hyperparameters by testing one model combination at a time and validating each combination [26]. In the process, Grid Search tries all possible combinations of hyperparameter values. This method can take a long time, especially for models with many hyperparameters [25].

b. Random Search

Random Search is a method similar to Grid Search, but the hyperparameter values are chosen randomly. This method is faster than Grid Search, but the results obtained may not be optimal. This method can be effective if the search space is large, but it does not guarantee that the optimal hyperparameter value will be found [25].

c. Bayesian optimization

Bayesian optimization is another method that uses probabilistic models to guide the search for optimal hyperparameters. This method balances exploration and exploitation to efficiently find the best configuration [25].

2.5 Model Evaluation

Evaluation is conducted using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to measure how well these models predict production values in the test data [9],[27],[28], .

a. Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{Y}_i - Y_i| \tag{7}$$

b. Mean Square Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{8}$$

c. Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \tag{9}$$

d. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n}} \tag{10}$$

3. RESEARCH METHOD

At this stage (Figure 4), the main goal is to create a model that can accurately predict the results by utilizing unique patterns and trends in the data. In this study, two types of models, namely Random Forest and SVR, were used to forecast paddy production. This process is done with the help of R Studio software and installing additional packages such as Random Forest and e1071. The Random Forest and SVR modeling process can be described as follows.

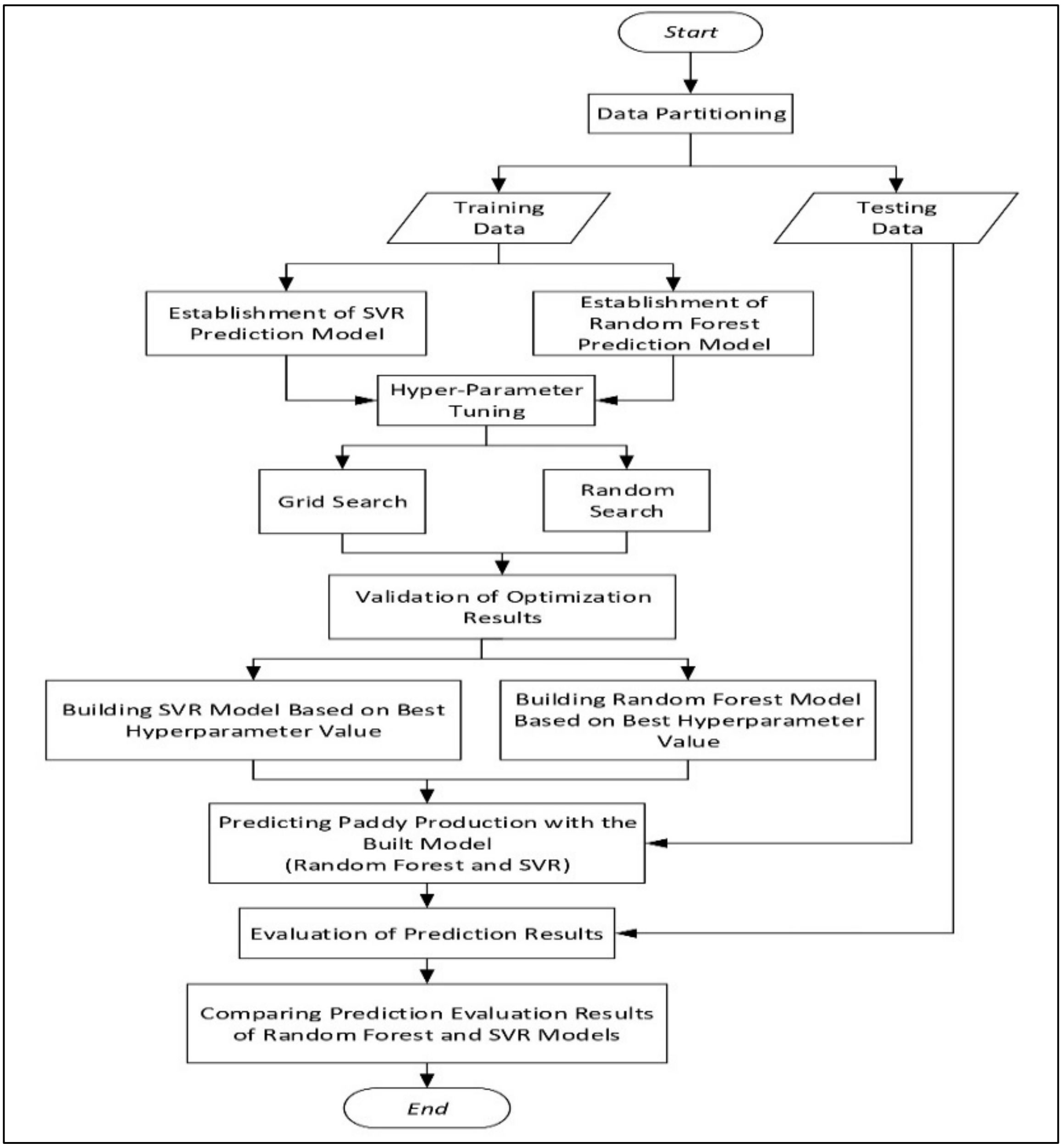

Figure 4. Research Method

3.1 Data Partitioning

In this study, paddy production data from 1993 to 2022 were collected from the official website of Badan Pusat Statistik (BPS) Indonesia (https://www.bps.go.id) in Excel (.xlsx) format. The dataset consists of provincial-level observations, which were subsequently aggregated to obtain national-level production values for each year, ensuring consistency with the modeling objective. The aggregated dataset was then partitioned into training and testing sets using the *createDataPartition()* function from the caret package in R. A 70:30 split ratio was applied, where 70% of the data were used for model training and 30% for testing. This proportion is commonly adopted in machine learning to balance model learning and evaluation, allowing sufficient data for training while maintaining an independent test set to assess model generalization performance [14].

3.2 Training Model Building

At this stage, two machine learning models, namely SVR and Random Forest, are formed using the training data. The SVR model is formed using the *svm* function from package e1071. and the Kernels used are RBF and Linear. While the random forest model is formed using the Random Forest package.

3.3 Hyperparameter Tuning

In this step, the hyperparameters of the model will be optimized using Grid Search and Random Search techniques. Hyperparameters are parameters that can be adjusted to improve model performance. The hyperparameter to be optimized for the Random Forest model is mtry. In the SVR model, the hyperparameters to be optimized for the RBF Kernel are *sigma* and *cost (C)*, while the Linear Kernel will optimize the *cost (C)* value hyperparameter.

3.4 Validation of Optimization Results

At this stage, the hyperparameter optimization results are validated to determine the best hyperparameter value. Then, the best hyperparameter value is used to build Random Forest models and SVR models in predicting paddy production.

3.5 Building Prediction Model

At this stage, the Random Forest model and SVR model are built using the best hyperparameter values obtained from the previous step.

3.6 Predicting Paddy Production

After the model is built, the next step is to predict the test data using the Random Forest and SVR models that have been made previously using the *predict()* function. Then, visualize the prediction results. This visualization aims to compare the actual production value with the predicted results of the Random Forest and SVR models.

3.7 Evaluation

At this stage, the prediction results of the model are evaluated to determine its performance. Evaluation is done using various metrics, such as MAE, MSE, RMSE, and MAPE.

4. RESULTS AND DISCUSSION

The dataset used in this study consists of paddy production data obtained from the official website of Badan Pusat Statistik (BPS) Indonesia (https://www.bps.go.id), covering the period from 1993 to 2022 in Excel (.xlsx) format. The dataset comprises 930 observations with six variables, namely Year, Province, Commodity Type, Production, Productivity, and Harvested Area. Table 1 presents the description of the research variables.

Table 1. Research Data Variables

Variable	Example Data
Year	2015
Province	West Java
Commodity.Types	Paddy
Production	11373144
Productivity	61.22
Harvested.Area	1857612

Prior to modeling, provincial-level data were aggregated to obtain national-level production values for each year. The dataset was then partitioned into training and testing sets using a 70:30 ratio. The training set was used for model development and hyperparameter tuning, while the testing set was used to evaluate model performance. The modeling workflow involves training the model, optimizing hyperparameters using 10-fold cross-validation, and evaluating the final model on the testing dataset. All evaluation metrics (MSE, RMSE, MAE, and MAPE) are computed based on the comparison between predicted and actual production values in the testing data. Although a 70:30 split is applied, alternative partition strategies (e.g., 80:20) may influence the results and could be explored in future work.

4.1 Evaluation of the Random Forest Model

The Random Forest model was optimized using both Grid Search and Random Search methods. For Grid Search, the optimal parameter was obtained at *mtry = 2*, with the model constructed using 1000 trees. This configuration suggests that a limited number of predictors at each split is sufficient to capture the underlying structure of the data.

Figure 5 illustrates the relationship between actual and predicted values for the Random Forest model optimized using Grid Search. The close alignment of data points along the diagonal line indicates that the model achieves a reasonable level of predictive accuracy in capturing the relationship between input variables and paddy production.

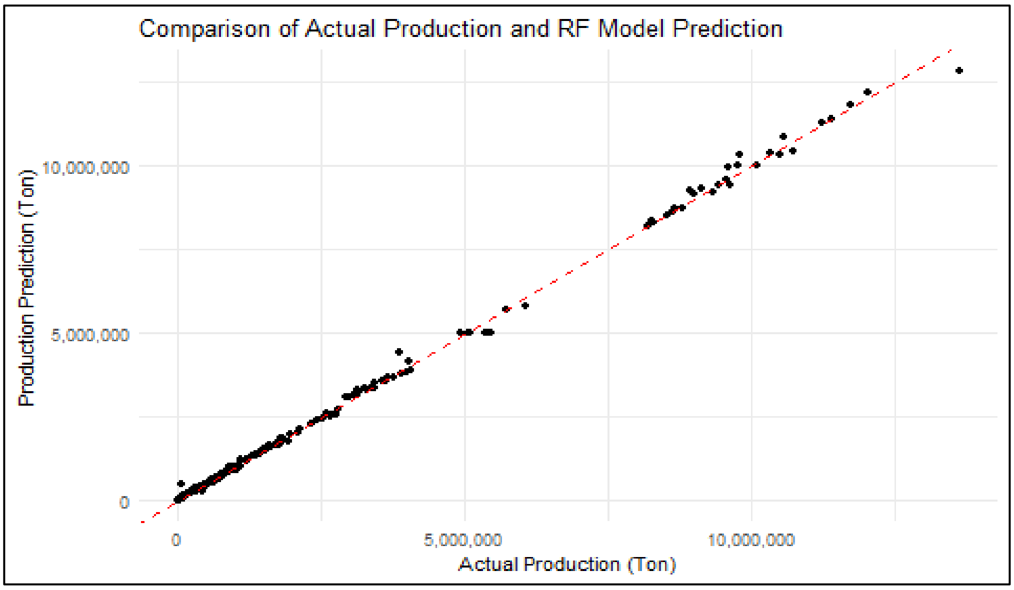

Figure 5. RF Predictions vs Actual (Grid Search)

Table 2 presents the performance evaluation results. The reported MSE values are calculated from the squared differences between predicted and actual production values in the testing dataset. Due to the large magnitude of paddy production data, the resulting MSE values appear numerically high. Therefore, RMSE and MAPE are considered more interpretable measures of model accuracy.

Table 2. Performance of Random Forest (Grid Search)

Method	MSE	RMSE	MAE	MAPE
Grid Search	11,747,484,893.46	108,385.81	50,092.75	0.08 %

Using Random Search, the model also identified *mtry = 2* as the optimal parameter. As shown in Figure 6, the predicted values are closely distributed around the diagonal line, indicating comparable performance to the Grid Search approach.

Table 3. Performance of Random Forest (Random Search)

Method	MSE	RMSE	MAE	MAPE
Random Search	11,626,623,757.42	107,826.82	49,810.99	0.08 %

Table 3 shows that Random Search produces slightly lower error values compared to Grid Search. This improvement indicates that Random Search is more effective in exploring a wider range of hyperparameter combinations, thereby increasing the likelihood of identifying more optimal parameter settings.

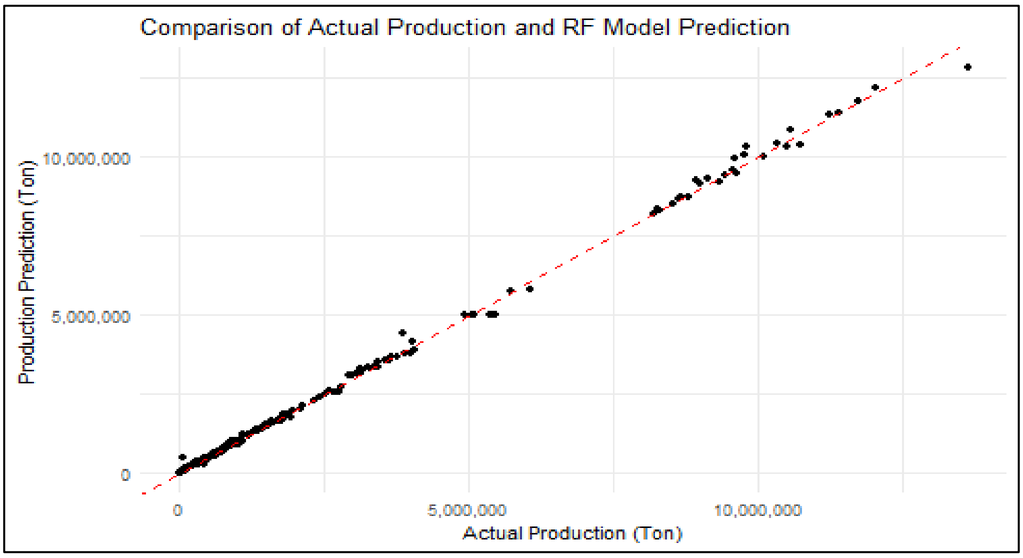

Figure 6. RF Predictions vs Actual (Random Search)

4.2 Evaluation of the SVR Model

The SVR model was optimized using both Grid Search and Random Search approaches. Grid Search identified the optimal parameters at *sigma = 0.2* and *C = 10*, which provide a balanced trade-off between model complexity and generalization capability.

Figure 7 shows the relationship between actual and predicted values for the SVR model optimized using Grid Search. The majority of data points lie close to the diagonal line, indicating strong predictive performance.

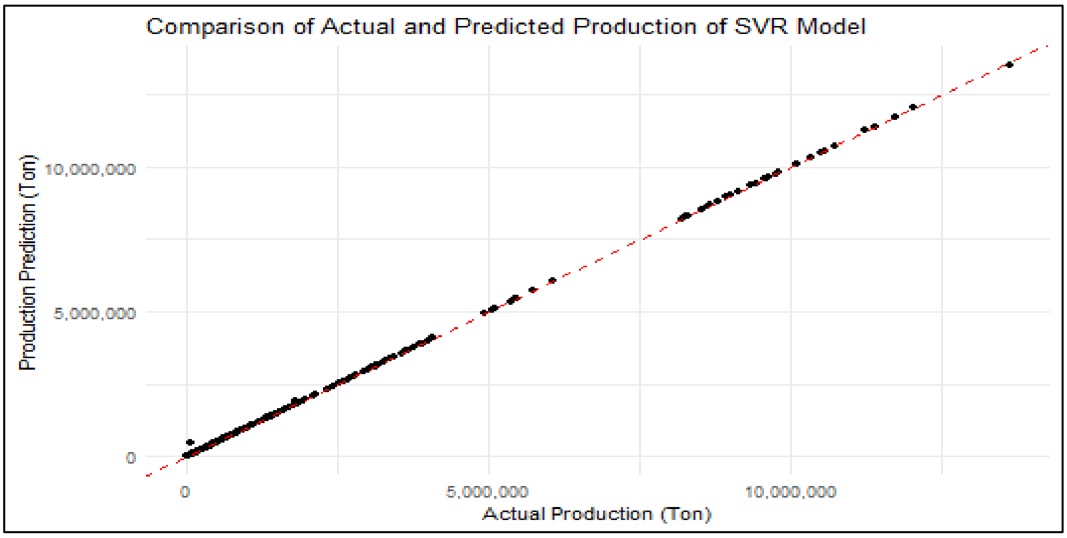

Figure 7. SVR Predictions vs Actual (Grid Search)

Table 4 summarizes the evaluation results, showing relatively low error values compared to the Random Forest model.

Table 4. Performance of SVR (Grid Search)

Method	MSE	RMSE	MAE	MAPE
Grid Search	832,253,877.15	28,848.81	3,723.42	0.09 %

Using Random Search, the optimal parameters were identified as *sigma = 0.145* and *C = 65.30*. The larger value of *C* indicates a more flexible model that allows smaller training errors, which can improve predictive accuracy for complex and non-linear relationships.

As illustrated in Figure 8, the SVR model optimized with Random Search shows a tighter clustering of data points around the diagonal line compared to Grid Search. Table 5 confirms that Random Search significantly reduces prediction error across all evaluation metrics.

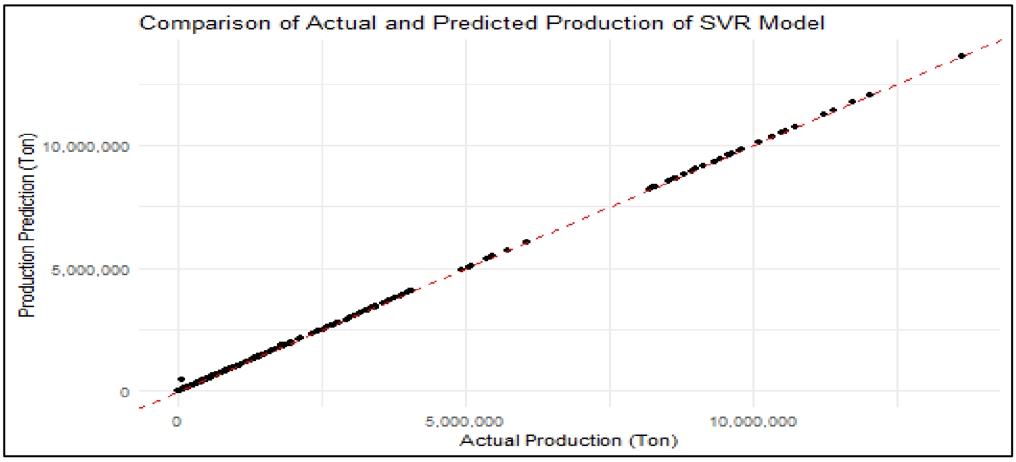

Figure 8. SVR Predictions vs Actual (Random Search)

It should be noted that the scatter-based visualization focuses on the overall relationship between actual and predicted values rather than the temporal sequence of the data. Therefore, the results are interpreted as an assessment of global model performance rather than time-dependent prediction behavior.

Table 5. Performance of SVR (Random Search)

Method	MSE	RMSE	MAE	MAPE
Random Search	755,072,528.71	27,478.58	2,924.59	0.05 %

4.3 Comparison of Model Evaluation

Figure 9 presents the overall comparison between Random Forest and SVR models using both optimization methods. The results show that SVR consistently achieves lower error values than Random Forest across most evaluation metrics. In particular, SVR optimized using Random Search yields the best performance, with the lowest RMSE and MAPE values. Although MSE values are substantially larger in magnitude, this is primarily due to the scale of the production data. Therefore, RMSE and MAPE provide more meaningful interpretations of prediction accuracy.

The superior performance of Random Search compared to Grid Search can be attributed to its ability to explore the hyperparameter space more efficiently. Unlike Grid Search, which evaluates a fixed set of parameter combinations, Random Search samples a broader range of configurations, increasing the likelihood of identifying near-optimal solutions. Overall, the SVR model optimized with Random Search demonstrates the best predictive performance, indicating its effectiveness in capturing complex and non-linear patterns in paddy production data.

A more detailed examination of the results reveals that the performance gap between SVR and Random Forest is not marginal but substantial across all error metrics. For instance, the RMSE of SVR is approximately four times lower than that of Random Forest, indicating a significantly smaller deviation from actual values. Similarly, the MAE values suggest that SVR produces much more stable predictions with lower absolute errors. This consistent pattern across multiple metrics strengthens the robustness of the conclusion that SVR is better suited for this prediction task. Moreover, the relatively small difference between Grid Search and Random Search within each model suggests that model choice has a greater impact on performance than the optimization strategy itself. This implies that selecting an appropriate learning algorithm is more critical than exhaustive hyperparameter tuning, particularly when the underlying data exhibits strong non-linear characteristics.

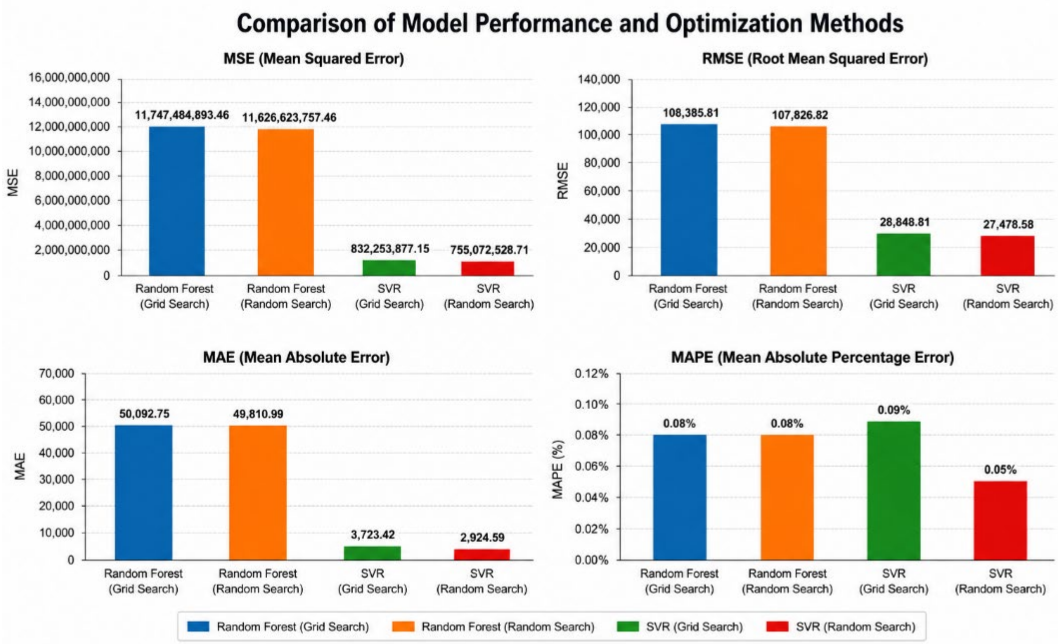

Figure 9. Comparison of RF and SVR Performance Across Optimization Methods

5. CONCLUSION

In the comparison of machine learning models for paddy production prediction, the results show that Support Vector Regression (SVR) outperforms Random Forest across all evaluation metrics. In particular, SVR optimized using Random Search achieves the lowest error values, with an RMSE of 27,478.58, and MAPE of 0.05%, significantly lower than those of Random Forest. These results indicate that SVR is more effective in capturing the non-linear and continuous patterns in paddy production data. The superior performance of Random Search compared to Grid Search is attributed to its ability to explore the hyperparameter space more efficiently, allowing it to identify better parameter configurations in complex models.

From a practical perspective, lower prediction errors imply more reliable forecasts, which are essential for supporting government decisions related to food supply planning, stock management, and price stability. High prediction errors may lead to inaccurate estimation of production levels, potentially resulting in inefficient policy interventions and supply imbalances. Therefore, SVR combined with Random Search provides the most effective approach for paddy production prediction in this study, highlighting the importance of both model selection and optimization strategy in improving forecasting accuracy.

6. FUTURE WORKS

Future research can be extended by addressing several limitations of this study. First, the use of time-aware modeling approaches such as Long Short-Term Memory (LSTM) and Transformer-based architectures may improve the ability to capture temporal dependencies in paddy production data. Second, incorporating additional explanatory variables, including climate conditions and socio-economic factors, could enhance model robustness and predictive performance. Third, more advanced optimization strategies, such as metaheuristic-based tuning, can be explored to improve hyperparameter search efficiency in complex models. These improvements are expected to contribute to the development of a more accurate and comprehensive paddy production forecasting system for supporting national food security planning in Indonesia.

ACKNOWLEDGEMENTS

We are grateful for the help provided in this research process by Institute for Research and Community Service (LPPM) Universitas Mikroskil.

REFERENCES

[1] D. A. Usdianto and R. Setiyowati, "Mathematical Model of Rice Commodity Supply Chain in Indonesia," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 1139–1148, 2023, doi: 10.30598/barekengvol17iss2pp1139-1148.

[2] S. Olena, M. National, B. Anna, and B. Alla, "Global Aspects of Food Security in the World: Economic Dimensions," vol. 38, no. April, pp. 178–185, 2023.

[3] Suparno, "Policy Implementation on the Rice of in Order to Increase Food Stock in Rembang District," *Mod. Appl. Sci.*, vol. 12, no. 3, p. 84, 2018, doi: 10.5539/mas.v12n3p84.

[4] S. T. Pandangwati and D. Widiyanto, "The Nexus of Planning and Food Security in Indonesia: a Review of the Literature," *Built Environ. Stud.*, vol. 4, no. 1, pp. 11–20, 2023, doi: 10.22146/best.v4i1.5931.

[5] Badan Pusat Statistik, "Percentage of Monthly Average Per Capita Expenditure by Commodity Group (Indonesia, 1999–2023)," 2023. [Online]. Available: https://www.bps.go.id/en/statistics-table/1/OTM3IzE%3D/percentage-of-monthly-average-per-capita-expenditure-by-commodity-group--indonesia--1999--2002-2023.html. [Accessed: Apr. 27, 2026]..

[6] S. R. Topildiev and A. M. Kamilovna, "Priority Ways Of Agricultural Sector Development," *Int. J. Manag. Econ. Fundam.*, vol. 02, no. 02, pp. 14–18, 2019, doi: https://doi.org/10.37547/ijmef/Volume03Issue02-01.

[7] A. T. Rianti, A. Bafadal, and A. Abdi, "The Forecasting Analysis of Rice Production and Sufficiency Consumption of Rice (Oriza sativa) in Konawe District," *J. Ilm. Membangun Desa dan Pertan.*, vol. 8, no. 3, pp. 96–101, 2023, doi: 10.37149/jimdp.v8i3.131.

[8] X. Guo and P. Hao, "Using a random forest model to predict the location of potential damage on asphalt pavement," *Appl. Sci.*, vol. 11, no. 21, 2021, doi: 10.3390/app112110396.

[9] M. Safari, A. H. Rabiee, and J. Joudaki, "Developing a Support Vector Regression (SVR) Model for Prediction of Main and Lateral Bending Angles in Laser Tube Bending Process," *Materials (Basel).*, vol. 16, no. 8, 2023, doi: 10.3390/ma16083251.

[10] Y. Chen, Z. Lao, R. Wang, J. Li, J. Gai, and H. You, "Prediction of Both E-Jet Printing Ejection Cycle Time and Droplet Diameter Based on Random Forest Regression," *Micromachines*, 2023, doi: 10.3390/mi14030623.

[11] F. Wang, Y. Wang, X. Ji, and Z. Wang, "Effective Macrosomia Prediction Using Random Forest Algorithm," *Int. J. Environ. Res. Public Health*, vol. 19, no. 6, 2022, doi: 10.3390/ijerph19063245.

[12] M. Aslam, "Hierarchical Random Forest Formation with Nonlinear Regression Model for Cardiovascular Diseases Prediction," *Int. Conf. Comput. Commun. Informatics*, pp. 28–32, 2021, doi: 10.1109/ICCCI50826.2021.9402571.

[13] E. Fradinata, Z. M. Kesuma, S. Rusdiana, and N. Zaman, "Forecast Analysis of Instant Noodle Demand using Support Vector Regression (SVR)," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 506, no. 1, 2019, doi: 10.1088/1757-899X/506/1/012021.

[14] M. S. Zaghloul, R. A. Hamza, O. T. Iorhemen, and J. H. Tay, "Comparison of adaptive neuro-fuzzy inference systems (ANFIS) and support vector regression (SVR) for data-driven modelling of aerobic granular sludge reactors," *J. Environ. Chem. Eng.*, vol. 8, no. 3, p. 103742, 2020, doi: 10.1016/j.jece.2020.103742.

[15] I. A. Putra, O. D. Nurhayati, and D. Eridani, "Human Action Recognition (HAR) Classification Using MediaPipe and Long Short-Term Memory (LSTM)," *Teknik*, vol. 43, no. 2, pp. 190–201, 2022, doi: 10.14710/teknik.v43i2.46439.

[16] K. Kumari, M. Bhardwaj, and S. Sharma, "OSEMN Approach for Real Time Data Analysis," *Int. J. Eng. Manag. Res.*, vol. 10, no. 02, pp. 107–110, 2020, doi: 10.31033/ijemr.10.2.11.

[17] H. Liu, Q. Li, D. Yu, and Y. Gu, "Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms," *Appl. Sci.*, vol. 9, no. 19, 2019, doi: 10.3390/app9194069.

[18] M. P. Kusuma and A. Kudus, "Penerapan Metode Support Vector Regression (SVR) pada Data Survival KPR PT. Bank ABC, Tbk.," *Bandung Conf. Ser. Stat.*, vol. 2, no. 2, pp. 167–172, 2022, doi: 10.29313/bcss.v2i2.3614.

[19] Z. Rais, "Analisis Support Vector Regression (Svr) Dengan Kernel Radial Basis Function (Rbf) Untuk Memprediksi Laju Inflasi Di Indonesia," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 1, pp. 30–38, 2022, doi: 10.35580/variansiunm13.

[20] M. Ghazali and R. Purnamasari, "Pencarian Kernel Terbaik Support Vector Regression Pada Kasus Data Kemiskinan Di Indonesia Dengan User Interface (Gui) Matlab," *Statistika*, vol. 9, no. 1, pp. 1–8, 2021.

[21] M. Alida and M. Mustikasari, "Rupiah Exchange Prediction of US Dollar Using Linear, Polynomial, and Radial Basis Function Kernel in Support Vector Regression," *J. Online Inform.*, vol. 5, no. 1, pp. 53–60, 2020, doi: 10.15575/join.v5i1.537.

[22] X. Yang, J. L. Mi, J. Jiang, J. W. Li, Q. S. Zhang, and M. M. Geng, "Carbon Sink Price Prediction Based on Radial Basis Kernel Function Support Vector Machine Regression Model," *Int. J. Low-Carbon Technol.*, vol. 17, no. August, pp. 1075–1084, 2022, doi: 10.1093/ijlct/ctac088.

[23] R. Alias, S. J. Matlan, and A. Ibrahim, "Prediction of Cantilever Retaining Wall Stability using Optimal Kernel Function of Support Vector Machine," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 433–438, 2021, doi: 10.14569/IJACSA.2021.0120648.

[24] F. Fahmi, "Model Support Vector Regression (SVR) Berdimensi Tinggi dengan Pendekatan Fungsi Kernel Berbeda untuk Peramalan Harga Saham TLKM: Sebuah Pemodelan Deret Waktu Selama Masa Pandemi Covid-19," *J. Infomedia*, vol. 5, no. 2, p. 44, 2021, doi: 10.30811/jim.v5i2.2033.

[25] E. Bartz, T. Bartz-Beielstein, M. Zaefferer, and O. Mersmann, *Hyperparameter Tuning for Machine and Deep Learning with R: A Practical Guide*. 2023. doi: 10.1007/978-981-19-5170-1.

[26] Fatmawati and N. A. K. Rifai, "Klasifikasi Penyakit Diabetes Retinopati Menggunakan Support Vector Machine dengan Algoritma Grid Search Cross-validation," *J. Ris. Stat.*, pp. 79–86, 2023, doi: 10.29313/jrs.v3i1.1945.

[27] Z. Guo, B. Yu, M. Hao, W. Wang, Y. Jiang, and F. Zong, "A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient," *Aerosp. Sci. Technol.*, vol. 116, 2021, doi: 10.1016/j.ast.2021.106822.

[28] J. Cai, K. Xu, Y. Zhu, F. Hu, and L. Li, "Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest," *Appl. Energy*, vol. 262, 2020, doi: 10.1016/j.apenergy.2020.114566.