Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 6, December 2024, pp. 4182~4192
ISSN: 2302-9285, DOI: 10.11591/eei.v13i6.7773

Prediction of palm oil production using hybrid decision tree based on fuzzy inference system Tsukamoto

Tundo[1], Shoffan Saifullah[2,5], Mesra Betty Yel[1], Opi Irawansah[3], Zulfikar Yusya Mubarak[3], Andi Saidah[4]

[1]Department of Informatics, STIKOM Cipta Karya Informatika, Jakarta, Indonesia
[2]Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Yogyakarta, Indonesia
[3]Department of Digital Business, Universitas Al-Irsyad, Cilacap, Indonesia
[4]Faculty of Engineering and Informatics, Universitas 17 Agustus 1945, Jakarta, Indonesia
[5]Faculty Computer Sciences, AGH University of Krakow, Krakow, Poland

Article Info

Article history:

Received Oct 25, 2023
Revised Apr 4, 2024
Accepted May 16, 2024

Keywords:

Agricultural forecasting
Fuzzy inference system
Hybrid decision tree
Palm oil production
Predictive modeling
Tsukamoto

ABSTRACT

This research addresses the challenge of optimizing rule creation for palm oil production at PT Tapiana Nadenggan. It deals with the complexity of diverse agricultural variables, environmental factors, and the dynamic nature of palm oil production. The existing problem lies in the limitations of conventional decision tree models—J48, reduced error pruning (REP), and Random—in capturing the nuanced relationships within the intricate palm oil production system. The study introduces hybrid decision tree models—specifically J48-REP, REP-Random, and Random-J48—to address this challenge via combination scenarios. This approach aims to refine and update the rule creation process, enabling the recognition of nuanced performance processes within the selected decision tree combinations. To comprehensively tackle this challenge and problem, the study employs Tsukamoto's fuzzy inference system (FIS) for a sophisticated performance comparison. Despite the complexity, intriguing results emerge after the forecasting process, with the standalone J48 decision tree achieving 85.70% accuracy and the combined J48-REP excelling at 93.87%. This highlights the potential of decision tree combinations in overcoming the complexities inherent in forecasting palm oil production, contributing valuable insights for informed decision-making in the industry.

Corresponding Author:

Tundo
Department of Informatics, STIKOM Cipta Karya Informatika
St. Radin Inten II No 8.5, Duren Sawit, East Jakarta, 13440, Indonesia
Email: asna8mujahid@gmail.com

1. INTRODUCTION

Palm oil production is a cornerstone of global agriculture and is pivotal in various industries and economies worldwide [1]. Despite its significance, optimizing palm oil production presents a multifaceted challenge characterized by the intricate interplay of agricultural variables and environmental factors [2]. The variability in palm oil production, influenced by diverse factors such as climate conditions, soil quality, and agricultural practices, underscores the need for sophisticated forecasting methods to facilitate informed decision-making [3]. While numerous studies have explored the application of decision trees in agricultural forecasting, the dynamic nature of palm oil production necessitates a more nuanced approach [4]. Existing research often needs to look more into the potential of combining decision trees to enhance predictive capabilities, thereby leaving a critical gap in the field. This research identifies this gap and seeks to address it by introducing a novel framework that harnesses the strengths of individual decision trees while exploring synergies through combination scenarios.

Decision trees, renowned for their efficacy in data mining for classification problems, offer a structured approach characterized by a hierarchical tree-like configuration [5]. In this study, we delve into three key decision tree types–J48 [6], reduced error pruning (REP) [7], and random [8]–leveraging their capabilities in rule creation based on empirical data from PT Tapiana Nadenggan, a significant player in palm oil production. Introducing a layer of innovation to existing methodologies, we propose a combination concept that explores scenarios like J48-REP, REP-Random, and Random-J48. This unique approach involves averaging the results of decision trees using combinations such as (J48+REP)/2, (REP+Random)/2, and (Random+J48)/2. By doing so, we aim to create more robust decision tree combinations that offer a comprehensive outlook on palm oil production forecasting [9]. Following the formation of rules from selected decision trees, individual or combined, the Tsukamoto fuzzy inference system (FIS) emerges as a crucial tool for discerning optimal rules for determining palm oil production. This method brings a nuanced perspective to the analysis, considering accuracy, TP rate, FP rate, precision, and recall for each classification class.

In the broader context of research methodologies, our work aligns with contemporary trends and draws inspiration from various innovative approaches. Notably, studies by Bhatnagar and Kumar [10] on SMS classification, Uyun and Choridah [11] on feature selection using decision trees [12], and Supianto *et al*. [13] on classifying student graduation serve as benchmarks for algorithmic accuracy assessment and offer valuable insights into decision tree applications. In the realm of SMS classification, Bhatnagar and Kumar [10] adopted machine learning algorithms to filter spam messages, achieving an impressive 98.7% success rate by classifying messages based on textual content. This innovative approach streamlined the classification process by eliminating the need for resource-intensive database queries and intricate regular expression matching. The dataset from netcore solutions Pvt. Ltd. in India was pivotal in validating their proposed method. Uyun and Choridah [11] work on feature selection for mammogram image extraction showcased the decision tree algorithm as a top performer, achieving a notable 93.18% classification accuracy. Their study leveraged data mining techniques and focused on the crucial facet of selecting pertinent features in the context of mammogram images. Moving to Supianto *et al*. [13] research on student graduation classification, comparing decision tree algorithms revealed that C4.5 outperformed random tree and REPTree, boasting an accuracy of 77.01%. These studies underscore the versatility and efficacy of decision tree methodologies across diverse domains, providing a nuanced understanding of their applications and reinforcing their relevance in contemporary research landscapes.

This research aims to bridge the gap by introducing a comprehensive framework integrating hybrid decision trees–J48, REP, and random–with the FIS Tsukamoto [14], [15]. Our proposed solution seeks to refine the accuracy of individual decision trees and unlock the potential of decision tree combinations in forecasting palm oil production. By leveraging the capabilities of FIS Tsukamoto [16], our approach aims to provide a nuanced understanding of the complex relationships within the agricultural ecosystem. Building upon the current state-of-the-art in agricultural forecasting, our study aims to establish a more robust and adaptive predictive model aligned with the dynamic nature of palm oil production. Through empirical trials and systematic analyses, we seek to contribute valuable insights that advance precision in forecasting, offering practical implications for decision-making in the palm oil industry.

This paper is structured as follows: section 2 delineates the methods employed in our hybrid decision tree and Tsukamoto algorithms. Section 3 presents the results and discussion of our experiments, shedding light on the outcomes of the decision tree combinations and Tsukamoto FIS. Finally, section 4 summarizes the conclusions drawn from our experiment and outlines potential avenues for future research.

2. METHOD

This section offers a detailed overview of our research methodology, as depicted in Figure 1. It begins with precise data collection from PT Tapiana Nadenggan, followed by thorough preprocessing. The subsequent stage involves a combination of decision tree and FIS Tsukamoto for a comprehensive evaluation of forecasting. This integrated approach ensures a robust analytical process (see Figure 1).

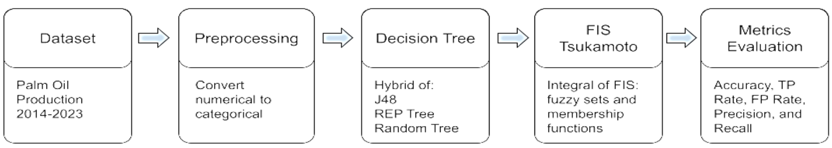

Figure 1. Process flow for forecasting with hybrid decision tree and FIS Tsukamoto rules

2.1. Dataset and preprocessing

The dataset employed in this research is derived from PT Tapiana Nadenggan, a key player in the palm oil production sector. This collaborative effort ensures access to a robust dataset covering the period from January 2014 to April 2023, with monthly records providing a comprehensive temporal overview of palm oil production dynamics. The dataset encompasses essential parameters, including palm oil quantities, consumer demand, existing stock levels, and production volumes, meticulously selected to facilitate accurate prediction modeling.

Structured in CSV format, the dataset aligns with the WEKA software [17], facilitating seamless integration for subsequent rule creation and analysis [18]. This comprehensive dataset, outlined in Table 1, serves as the foundation for forecasting palm oil production. Table 1 offers a detailed snapshot of monthly metrics that shape the palm oil production landscape.

Table 1. Dataset

Month	Year	Palm oil	Demand	Stock	Production
Jan	2014	20,875,600	4,730,300	3,960,000	10,020,000
Feb	2014	26,300,700	14,987,000	4,220,200	19,300,500
Mar	2014	26,250,400	14,980,000	4,500,000	19,150,000
Apr	2014	38,700,000	3,784,500	1,900,400	10,100,800
May	2014	24,400,100	7,568,600	4,000,700	13,568,000
Jun	2014	26,000,000	12,600,000	1,730,000	17,000,300
Jul	2014	34,857,100	10,811,400	3,959,140	18,954,280
....	....	....	....	....	....
....	....	....	....	....	....
Apr	2023	40,616,100	14,996,800	4,969,670	19,597,200

Source of data: PT Tapiana Nadenggan

The preprocessing phase, a critical component of refining the dataset, employs meticulous techniques. Handling missing values with mean imputation and predictive modeling, rigorous outlier detection using Z-score analysis and the interquartile range (IQR) method and adept temporal aggregation techniques contribute to dataset optimization [19]. Advanced feature scaling, normalization methods, and categorical variable handling ensure precision in model comparison. Integration of time-series decomposition methods, including seasonal decomposition of time series (STL), enriches the understanding of underlying patterns [20]. Strategically dividing the dataset into training and testing sets ensures the robustness and generalizability of forecasting models. A linguistic transformation, converting numerical data into refined linguistic values [21] (low, medium, and high), adheres to well-defined division rules, emphasizing seamless integration with the WEKA platform. This meticulous approach establishes a robust foundation for precise forecasting, leveraging hybrid decision tree models.

2.2. Decision tree construction with ensemble learning techniques

This section explores the decision tree construction process, intricately integrating ensemble learning techniques to augment precision and adaptability in forecasting palm oil production.

2.2.1. Individual decision tree algorithms

Our method strategically employs three robust decision tree algorithms, each with its unique attributes:

- J48 (C4.5 algorithm) [18]: recognized for its efficiency and scalability, particularly adept at handling large and complex datasets.
- REP [7]: emphasizes minimizing overall error rates by judiciously pruning branches, enhancing the tree's predictive accuracy.
- Random tree [22]: injects an element of randomness by selecting subsets of features at each split point, fostering diversity to mitigate overfitting.

2.2.2. Ensemble learning for improved accuracy

Moving beyond individual trees, our research delves into ensemble learning. This involves combining multiple decision trees [23], [24] synergistically to forge a more resilient and accurate prediction model. The following combination scenarios are meticulously explored:

- J48-REP: a symbiosis of J48's computational efficiency and REP's adept pruning capabilities, with the aim of achieving enhanced accuracy and model simplicity.

- Random-J48: this combination strategically harnesses the diversity introduced by random trees and the structural integrity offered by J48, mitigating overfitting concerns while preserving interpretability.
- REP-Random: an amalgamation of REP's precise pruning abilities and the randomness introduced by random trees, aiming for a robust model with improved generalization capabilities.

2.2.3. Evaluation and selection of the best ensemble

Our evaluation protocol employs a comprehensive range of metrics to assess decision trees and ensemble models rigorously [25] for integration with the FIS. We begin by evaluating model accuracy in identifying positive instances and minimizing false positives through the true positive rate (TPR) and false positive rate (FPR), which are calculated using (1) and (2), respectively.

$$TPR = \frac{TP}{TP+FN} \tag{1}$$

$$FPR = \frac{FP}{FP+TN} \tag{2}$$

Precision, recall, and F-measure further elucidate predictive accuracy and balance, as defined by (3) to (5):

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{FP}{FP+TN} \tag{4}$$

$$F - Measure = \frac{2\ x\ Precision\ x\ recall}{Precision\ +\ Recall} \tag{5}$$

We then assess discriminative ability through the receiver operating characteristic (ROC) area, calculated using the ROC curve. Accuracy and Kappa statistics evaluate overall correctness and agreement beyond chance, represented by (6) and (7):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$Kappa = \frac{P_o - P_e}{1-P_e} \tag{7}$$

Lastly, prediction errors are analyzed using mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE), defined by (8) to (11), respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

$$RAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} |y_i - \hat{y}|} \tag{10}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}} \tag{11}$$

2.2.4. Theoretical considerations

The theoretical underpinnings include a profound exploration of bagging and boosting, which are integral to ensemble learning. Bagging, exemplified in random forests, involves training on random data subsets with replacement, fostering diverse predictions that are then aggregated for the outcome. Boosting, a sequential training approach focuses on iteratively improving subsequent trees based on the errors of preceding ones [25]. While our research champions ensemble learning through the fusion of decision tree algorithms, we conscientiously acknowledge alternative ensemble methods like stacking and voting, incorporating diverse model types, potentially further enhancing predictive accuracy.

2.2.5. Integration with fuzzy inference system

The culmination of our decision tree model's journey involves seamless integration with the Tsukamoto FIS. Decision trees, providing discrete classifications, are intricately woven with FIS, harnessing fuzzy logic to navigate inherent uncertainties in agricultural data. This fusion ensures a nuanced and robust prediction framework for palm oil production. By intricately amalgamating ensemble learning and fuzzy logic with decision trees, our research pioneers a hybrid model, strategically leveraging the strengths of both approaches to elevate accuracy, flexibility, and interpretability in predicting palm oil production.

2.3. Fuzzy inference system Tsukamoto

In pursuing precise palm oil production predictions, this research seamlessly integrates the FIS with the Tsukamoto method, providing a nuanced and sophisticated analytical layer. The synergy between fuzzy logic and decision tree outcomes is explored, emphasizing the Tsukamoto method's role in refining predictions from various decision trees [26], including J48, REP, random, J48-REP, Random-J48, and REP-Random. The Tsukamoto method, integral to FIS, leans on fuzzy sets and membership functions. These functions, employing shapes like triangles and trapezoids, convert raw data into fuzzy propositions, enabling a more flexible and nuanced understanding of input variables. FIS operates on IF-THEN fuzzy rules, articulating intricate relationships between input and output variables. The antecedents present fuzzy propositions about the inputs, while the consequents contribute to the system's output, enhancing the interpretability of the decision-making process.

The Tsukamoto method unfolds through a meticulous inference process. It involves evaluating antecedents, clipping consequent functions, calculating a weighted average, and defuzzifying for crisp output. The step-by-step process ensures transparency and clarity in transforming fuzzy propositions into precise predictions. Beyond its inherent capabilities, FIS Tsukamoto assumes a pivotal role in refining predictions derived from decision trees. This integration focuses on comparing the outcomes of diverse decision trees with FIS Tsukamoto predictions against actual production data, aiming for a comprehensive understanding of the predictive models' strengths and limitations. Table 2 delineates the criteria, fuzzy sets, and domains employed in Tsukamoto's FIS modeling for palm oil production prediction. This meticulous integration signifies a synergy between decision tree methodologies and fuzzy logic, contributing to an enriched and nuanced forecasting approach.

Table 2. Tsukamoto's FIS modeling for palm oil production

Parameter	Criteria	Fuzzy set	Domain
Input	Palm oil	Low	16,572,300–30,036,150
		Medium	30,036,150–43,500,000
		High	>30,036,150
	Demand	Low	3,153,333–11,036,666.5
		Medium	11,036,666.5–18,920,000
		High	>18,920,000
	Stock	Low	833,333–2,916,666.5
		Medium	2,916,666.5–5,000,000
		High	>5,000,000
Output	Production	Low	6,000,000–12,900,000
		Medium	12,900,000–19,800,000
		High	>19,800,000

3. RESULTS AND DISCUSSION

This section presents a detailed examination of the results of various decision tree algorithms that predict palm oil production. The analysis encompasses individual decision trees (J48, REP, and random) and hybrid combinations (J48-REP, Random-J48, and REP-Random). The discussion explores each model's accuracy, metrics, and implications, shedding light on their effectiveness in forecasting palm oil production dynamics. Additionally, this section discusses the limitations inherent in the models and provides insights into potential areas for improvement.

3.1. Results of decision tree based on each method

Analyzing individual decision tree methods provides valuable insights into their predictive performance for palm oil production. The J48 decision tree stands out prominently, achieving a remarkable accuracy of 95.24%. Delving deeper into the metrics presented in Table 3, such as true positive (TP) rate, false positive (FP) rate, precision, recall, F-measure, and ROC area, reveals a robust performance across the low (91.3%), medium (90.9%), and high (100%) production classes. The high precision and low error rates underscore the J48 decision tree’s ability to effectively classify instances, making it a formidable choice for accurate predictions.

Table 3. Accuracy results for each decision tree method

Method	Class	TP rate	FP rate	Precision	Recall	F-measure	ROC area	Accuracy (%)	Kappa	MAE	RMSE	RAE (%)	RRSE (%)
J48	Low	0.913	0	1	0.913	0.955	0.977	95.24	0.923	0.057	0.17	13.70	37.07
	Medium	0.909	0.019	0.909	0.909	0.909	0.969						
	High	1	0.059	0.935	1	0.967	0.971						
REPTree	Low	0.783	0	1	0.783	0.878	0.954	90.48	0.848	0.099	0.22	23.84	48.89
	Medium	0.909	0.077	0.714	0.909	0.8	0.937						
	High	1	0.059	0.935	1	0.967	0.971						
Random tree	Low	0.913	0	1	0.913	0.955	0.985	95.24	0.923	0.053	0.16	12.83	35.87
	Medium	0.909	0.019	0.909	0.909	0.909	0.978						
	High	1	0.059	0.935	1	0.967	0.980						

Similarly, leveraging REP, the REP decision tree attains an accuracy of 90.48%. Examination of TP rates (78.3%, 90.9%, and 100% for low, medium, and high classes, respectively) indicates a degree of specificity in minimizing false positives, particularly in the low class. While the model exhibits 100%, 71.4%, and 93.5% precision rates for the respective classes, there is room for enhancement, especially in refining predictions within certain classes. Additionally, the random tree algorithm, injecting randomness into decision tree construction, yields notable results, achieving an accuracy of 95.24%. Detailed scrutiny of metrics like TP rate, FP rate, precision, recall, F-measure, and ROC area demonstrates a well-balanced and accurate prediction across the diverse production classes. In a broader context, the individual decision tree methods—J48, REP, and Random Tree—demonstrate distinct strengths. The J48 decision tree excels in precision and robustness, showcasing its prowess in intricate classification tasks. The REP decision tree, emphasizing REP, exhibits specificity in minimizing false positives, though areas for improvement are identified. The random tree algorithm introduces an element of randomness while maintaining high accuracy, contributing a unique approach to the prediction process. The nuanced analysis of these each methods establishes their performance benchmarks and provides a foundation for discussions on potential synergies in hybrid models.

3.2. Results of our proposed methods with the hybrids

Our innovative hybrid approach, integrating ensemble learning techniques and fuzzy logic through the Tsukamoto FIS method, aims to surpass the predictive prowess of individual decision tree methods. The amalgamation of decision trees, namely J48-REP, Random-J48, and REP-Random, undergoes a meticulous evaluation to gauge their efficacy in predicting palm oil production (Table 4). The J48-REP hybrid decision tree demonstrates a noteworthy accuracy of 92.86%, showcasing its superiority over individual decision tree methods. A closer examination of Table 4 reveals intricate details about the model’s performance across various production classes. The TP rate, ranging from 1.69% to 100%, signifies the model’s ability to identify TPs in different production scenarios effectively. Precision rates, especially in the medium and high classes, highlight a balanced and accurate prediction of positive instances. The ROC area values further affirm the model’s discriminative ability, solidifying its reliability in distinguishing between different production categories. The comprehensive metrics in Table 4 underscore the nuanced and accurate predictive capabilities of the J48-REP hybrid decision tree.

Table 4. Accuracy results for each decision tree method

Method	Class	TP rate	FP rate	Precision	Recall	F-measure	ROC area	Accuracy (%)	Kappa	MAE	RMSE	RAE (%)	RRSE (%)
J48-REP	Low	0.017	0	1	0.848	0.917	0.966	92.86	0.886	0.057	0.19	18.77	42.98
	Medium	0.909	0.048	0.8115	0.909	0.855	0.953						
	High	1	0.059	0.935	1	0.967	0.971						
Random-J48	Low	0.913	0	1	0.913	0.955	0.981	95.24	0.923	0.055	0.16	13.26	36.46
	Medium	0.909	0.019	0.909	0.909	0.909	0.974						
	High	1	0.059	0.935	1	0.967	0.974						
REP-Random	Low	0.848	0.909	1	0.848	0.909	1	92.86	0.886	0.077	0.19	18.33	42.38
	Medium	0	0.048	0.059	0	0.048	0.059						
	High	1	0.812	0.935	1	0.812	0.935						

The random-J48 hybrid decision tree attains an impressive accuracy of 95.24%, closely aligning with the individual random tree method. A detailed metrics analysis in Table 4 elucidates the model's performance across low, medium, and high production classes. Precision rates, especially in the low and medium classes, demonstrate the model's robust identification of positive instances, contributing significantly to its accuracy. ROC Area values validate the model's discriminative capability, emphasizing its effectiveness in predicting diverse production outcomes. The comprehensive metrics in Table 4 validate the efficacy of the Random-J48 hybrid decision tree in enhancing predictive accuracy.

The REP-Random hybrid decision tree achieves a commendable accuracy of 92.86%, showcasing its potential in leveraging the strengths of REP and randomness. An in-depth analysis of detailed metrics reveals consistent and robust performance across different production classes, with notable precision rates. The ROC Area values further support the model's ability to discriminate between production categories, solidifying its reliability. The comprehensive metrics in Table 4 underscore the harmonious integration of decision tree methodologies in the REP-Random hybrid, contributing to enhanced accuracy in palm oil production prediction.

3.3. Discussion and limitation

Our exploration into palm oil production prediction through decision tree methodologies and hybrid models has unearthed valuable insights. However, it is essential to delve into the nuances of our findings and acknowledge inherent limitations. The decision tree rule formation process, as detailed in Table 5, offers a transparent and interpretable representation of the predictive logic. Each algorithm, including J48, REP, and random tree, formulates rules based on input parameters like demand, palm oil, and stock. These rules, outlined in Table 5, are the foundation for predictive models, guiding the decision-making process in palm oil production estimation.

Table 5. Decision tree algorithm rules for palm oil production prediction

Algorithm	Node	Rule
J48	1	IF Demand Low AND Palm Oil High THEN Production Low
	2	IF Demand Low AND Palm Oil Low THEN Production Medium
	3	IF Demand High THEN Production High
	4	IF Demand Medium AND Stock High THEN Production Medium
	5	IF Demand Medium AND Stock Low THEN Production Medium
	6	IF Demand Medium AND Stock Medium THEN Production Low
	7	IF Demand Low THEN Production Low
REP	1	IF Demand Low AND Palm Oil High THEN Production Low
	2	IF Demand Low AND Palm Oil Low THEN Production Medium
	3	IF Demand High THEN Production High
	4	IF Demand Medium THEN Production Medium
	5	IF Demand Low THEN Production Low
Random	1	IF Demand Low AND Palm Oil High THEN Production Low
	2	IF P Demand Low AND Palm Oil Low AND Stock High THEN Production Low
	3	IF Demand Low AND Palm Oil Low AND Stock Low THEN Production Medium
	4	IF Demand Low AND Palm Oil Low AND Stock Medium THEN Production Medium
	5	IF Demand High AND Stock High THEN Production High
	6	IF Demand High AND Stock Low THEN Production High
	7	IF Demand High AND Stock Medium THEN Production Low
	8	IF Demand Medium AND Stock High THEN Production Medium
	9	IF Demand Medium AND Stock Low THEN Production Medium
	10	IF Demand Medium AND Stock Medium THEN Production Low
	11	IF Demand Low THEN Production Low

The decision tree rules provide a clear understanding of how various input parameters contribute to predicting palm oil production. For instance, rule 3 in J48 suggests that if the demand is categorized as 'High,' the predicted production is also 'High.' These rules form the basis for subsequent analysis and model comparison. Table 6 provides a comprehensive overview of the accuracy achieved by each decision tree algorithm, considering various metrics such as error percentage and accuracy percentage. J48 emerges as a robust performer with an accuracy of 85.70%, while the J48-REP combination surpasses all, boasting an accuracy of 93.87%. The careful selection of decision tree algorithms or their combinations is critical in ensuring the precision of predictions. Notably, the accuracy calculation employs AFER, contributing to a nuanced evaluation of predictive performance.

The optimal decision tree algorithm or combination selection involves a detailed examination of various accuracy metrics. J48-REP stands out for its high accuracy and low error percentage, indicating a reliable predictive model. The consideration of multiple metrics ensures a comprehensive evaluation of model performance. Figures 2 to 4 visually juxtapose the prediction results of different decision tree algorithms against actual production data. These graphical representations offer an intuitive understanding of how well the models align with real-world outcomes. The comparison underscores the effectiveness of hybrid models, particularly J48-REP, in achieving predictions closely aligned with actual production levels. Visual representations provide a qualitative assessment of model predictions. Figure 4, which compares all decision tree algorithms, highlights the consistent performance of J48-REP across different production categories. This aligns with our quantitative accuracy findings and reinforces the reliability of the hybrid approach. While our study provides a robust foundation for palm oil production prediction, certain limitations warrant consideration. The performance of our models relies heavily on the quality and representativeness of the training dataset—an evolving industry landscape, unforeseen external factors, and dynamic production dynamics challenge model generalizability. Continuous monitoring and recalibration of models will be essential to adapt to changing patterns and ensure sustained accuracy. Acknowledging limitations is crucial for a holistic interpretation of the results. The dependence on historical data assumes consistent patterns, which unforeseen events may disrupt. The call for continuous monitoring aligns with the dynamic nature of agricultural systems, emphasizing the need for adaptive modeling strategies.

Table 6. Accuracy evaluation for each decision tree algorithm in palm oil production prediction

Algorithm	Rill production (A)	Prediction (F)	|A–F|	|A–F|/A	Error (%)	Accuracy (%)
J48	16972100	17136609	164509	0.009692908	14.30	85.70
	15170500	11825034	3345466	0.220524439		
	18350200	12315559	6034641	0.328859685		
	19275000	17659804	1615196	0.083797458		
	15432800	13806863	1625937	0.10535593		
	17876500	15910728	1965772	0.109964031		
REP	16972100	20886261	3914161	0.230623258	14.33	85.67
	15170500	18594033	3423533	0.225670413		
	18350200	16828045	1522155	0.082950322		
	19275000	20446302	1171302	0.060767938		
	15432800	17829376	2396576	0.155291068		
	17876500	19753264	1876764	0.10498498		
Random	16972100	14762132	2209968	0.130211818	25.86	74.14
	15170500	11242067	3928433	0.258952111		
	18350200	10269662	8080538	0.440351495		
	19275000	15631145	3643855	0.189045655		
	15432800	11110671	4322129	0.280061233		
	17876500	13357868	4518632	0.25276939		
J48-REP	16972100	17136609	2039335	0.120158083	6.13	93.87
	15170500	11825034	39033.5	0.002572987		
	18350200	12315559	3778398	0.205905004		
	19275000	17659804	221947	0.01151476		
	15432800	13806863	385319.5	0.024967569		
	17876500	15910728	44504	0.002489525		
Random-J48	16972100	14762132	1022729.5	0.060259455	19.92	80.08
	15170500	11242067	3636949.5	0.239738275		
	18350200	10269662	7057589.5	0.38460559		
	19275000	15631145	2629525.5	0.136421556		
	15432800	11110671	2974033	0.192708582		
	17876500	13357868	3242202	0.18136671		
REP-Random	16972100	20886261	852096.5	0.05020572	8.82	91.18
	15170500	18594033	252450	0.016640849		
	18350200	16828045	4801346.5	0.261650908		
	19275000	20446302	1236276.5	0.064138859		
	15432800	17829376	962776.5	0.062385082		
	17876500	19753264	1320934	0.073892205		

Our discussion sheds light on the intricacies of decision tree rule formation, accuracy comparisons, and the visual representation of predictive models. The hybrid approach, exemplified by the J48-REP combination, demonstrates superior predictive accuracy. Future research endeavors could explore additional variables, refine existing models, and consider advanced machine learning techniques to enhance the predictive capabilities of palm oil production estimation models. Our findings offer a robust foundation for palm oil production prediction, and the emphasis on continuous improvement aligns with the evolving nature of agricultural systems. The success of J48-REP warrants further exploration into hybrid models, potentially incorporating more advanced machine-learning techniques for increased predictive power.

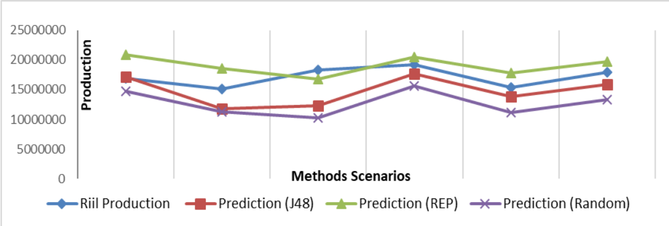

Figure 2. The prediction results of the decision tree algorithm are compared with the actual production

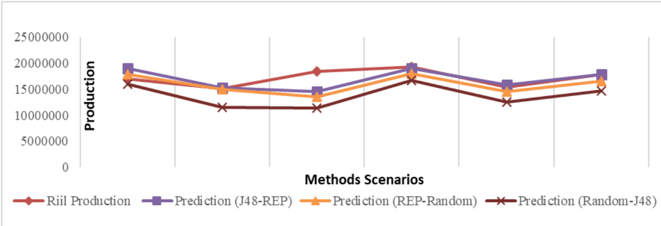

Figure 3. The prediction results of the combination decision tree algorithm are compared with actual production

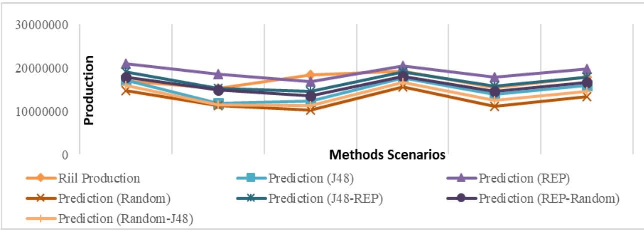

Figure 4. The prediction results of all decision tree algorithms are compared with actual production

4. CONCLUSION

This study presents a comprehensive exploration of palm oil production prediction, utilizing decision tree methodologies and hybrid models. The integration of the FIS with the Tsukamoto method, along with decision trees such as J48, REP, and random tree, resulted in a nuanced and accurate predictive framework. The J48-REP hybrid model emerged as a standout performer with an accuracy of 93.87%, demonstrating robust predictive capabilities across different production categories. However, the study acknowledges limitations related to dependency on historical data and the potential impact of unforeseen events. Continuous monitoring and adaptive modeling strategies are essential. Overall, these findings contribute valuable insights to palm oil production prediction, emphasizing the need for ongoing improvement and adaptation in dynamic agricultural systems.

REFERENCES

[1] S. Saifullah, D. B. Prasetyo, Indahyani, R. Drezewski, and F. A. Dwiyanto, "Palm Oil Maturity Classification Using K-Nearest Neighbors Based on RGB and L*a*b Color Extraction," *Procedia Computer Science*, vol. 225, pp. 3011–3020, 2023, doi: 10.1016/j.procs.2023.10.294.
[2] S. Y. Foong *et al.*, "Microwave processing of oil palm wastes for bioenergy production and circular economy: Recent advancements, challenges, and future prospects," *Bioresource Technology*, vol. 369, p. 128478, Feb. 2023, doi: 10.1016/j.biortech.2022.128478.
[3] P. Durán-Díaz, "Sustainable Land Governance for Water–Energy–Food Systems: A Framework for Rural and Peri-Urban Revitalisation," *Land*, vol. 12, no. 10, p. 1828, Sep. 2023, doi: 10.3390/land12101828.
[4] S. O. Araújo, R. S. Peres, J. C. Ramalho, F. Lidon, and J. Barata, "Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives," *Agronomy*, vol. 13, no. 12, p. 2976, Dec. 2023, doi: 10.3390/agronomy13122976.
[5] V. Sharma, M. L. Tsai, C. W. Chen, P. P. Sun, P. Nargotra, and C. Di Dong, "Advances in machine learning technology for sustainable biofuel production systems in lignocellulosic biorefineries," *Science of the Total Environment*, vol. 886, p. 163972, Aug. 2023, doi: 10.1016/j.scitotenv.2023.163972.
[6] A. D. Amirruddin, F. M. Muharam, M. H. Ismail, M. F. Ismail, N. P. Tan, and D. S. Karam, "Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (Elaeis guineensis) based on frond numbers: Analysis of decision tree and random forest," *Computers and Electronics in Agriculture*, vol. 169, p. 105221, 2020, doi: 10.1016/j.compag.2020.105221.

[7] A. Maachi, T. Nagata, and J. M. F. Silva, "Date palm virus A: first plant virus found in date palm trees," *Virus Genes*, vol. 56, no. 6, pp. 792–795, Dec. 2020, doi: 10.1007/s11262-020-01801-0.
[8] A. Ismail, H. M. A. Rashid, R. Gholami, and A. Raza, "Characterization based machine learning modeling for the prediction of the rheological properties of water-based drilling mud: an experimental study on grass as an environmental friendly additive," *Journal of Petroleum Exploration and Production Technology*, vol. 12, no. 6, pp. 1677–1695, Jun. 2022, doi: 10.1007/s13202-021-01425-6.
[9] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction," *IEEE Access*, vol. 9, pp. 63406–63439, 2021, doi: 10.1109/ACCESS.2021.3075159.
[10] S. Bhatnagar and A. Kumar, "A rule-based classification of short message service type," in *Proceedings of the 2nd International Conference on Inventive Systems and Control ICISC 2018*, 2018, pp. 1139–1142, doi: 10.1109/ICISC.2018.8398982.
[11] S. Uyun and L. Choridah, "Feature selection mammogram based on breast cancer mining," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 1, pp. 60–69, 2018, doi: 10.11591/ijece.v8i1.pp60-69.
[12] S. Saifullah, R. Dreżewski, F. A. Dwiyanto, A. S. Aribowo, and Y. Fauziah, "Sentiment Analysis Using Machine Learning Approach Based on Feature Extraction for Anxiety Detection," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14074 LNCS, Berlin, Heidelberg: Springer-Verlag, pp. 365–372, 2023, doi: 10.1007/978-3-031-36021-3_38.
[13] A. A. Supianto, A. Julisar Dwitama, and M. Hafis, "Decision Tree Usage for Student Graduation Classification: A Comparative Case Study in Faculty of Computer Science Brawijaya University," *3rd International Conference on Sustainable Information Engineering and Technology, SIET 2018 - Proceedings*, 2018, pp. 308–311, doi: 10.1109/SIET.2018.8693158.
[14] F. P. Schena, V. W. Anelli, D. I. Abbrescia, and T. Di Noia, "Prediction of chronic kidney disease and its progression by artificial intelligence algorithms," *Journal of Nephrology*, vol. 35, no. 8, pp. 1953–1971, May 2022, doi: 10.1007/s40620-022-01302-3.
[15] J. Sutopo, M. K. Abd Ghani, M. A. Burhanuddin, A. N. Septiani, and Tundo, "Dance Gesture Recognition Using Laban Movement Analysis with J48 Classification," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 11, no. 2, pp. 528–536, 2023, doi: 10.52549/ijeei.v11i2.4314.
[16] D. P. Alamsyah, Y. Ramdhani, and R. D. Nurbeni, "Implementation of the Fuzzy Inference System Tsukamoto Method in the Decision Support System," *ISESD 2022 - 2022 International Symposium on Electronics and Smart Devices, Proceeding*, pp. 1–6, Nov. 2022, doi: 10.1109/ISESD56103.2022.9980745.
[17] A. T. Akbar, R. Husaini, B. M. Akbar, and S. Saifullah, "A proposed method for handling an imbalance data in classification of blood type based on Myers-Briggs type indicator," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 4, pp. 276–283, Oct. 2020, doi: 10.14710/jtsiskom.2020.13625.
[18] E. N. Azizah, U. Pujianto, E. Nugraha, and Darusalam, "Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a Virtual Learning Environment," in *2018 4th International Conference on Education and Technology ICET 2018*, 2018, no. 1, pp. 18–22. doi: 10.1109/ICEAT.2018.8693928.
[19] V. S. Kumar *et al.*, "Optimizing solar power plant efficiency through advanced analytical framework and comparative analysis," *Journal of Intelligent and Fuzzy Systems*, vol. 46, no. 2, pp. 4995–5011, Feb. 2024, doi: 10.3233/JIFS-235578.
[20] C. Zhang, M. Li, S. Sun, L. Tang, and S. Wang, "Decomposition Methods for Tourism Demand Forecasting: A Comparative Study," *Journal of Travel Research*, vol. 61, no. 7, pp. 1682–1699, Sep. 2022, doi: 10.1177/00472875211036194.
[21] S. Zhong, C. I. Chang, and Y. Zhang, "Iterative Support Vector Machine for Hyperspectral Image Classification," in *Proceedings International Conference on Image Processing ICIP*, 2018, pp. 3309–3312, doi: 10.1109/ICIP.2018.8451145.
[22] S. O. Giwa, R. T. Taziwa, and M. Sharifpur, "Dependence of composition-based approaches on hybrid biodiesel fuel properties prediction using artificial neural network and random tree algorithms," *Renewable Energy*, vol. 218, p. 119324, Dec. 2023, doi: 10.1016/j.renene.2023.119324.
[23] A. Arabameri *et al.*, "Decision tree based ensemble machine learning approaches for landslide susceptibility mapping," *Geocarto International*, vol. 37, no. 16, pp. 4594–4627, Aug. 2022, doi: 10.1080/10106049.2021.1892210.
[24] Y. Ang *et al.*, "A novel ensemble machine learning and time series approach for oil palm yield prediction using Landsat time series imagery based on NDVI," *Geocarto International*, vol. 37, no. 25, pp. 9865–9896, Dec. 2022, doi: 10.1080/10106049.2022.2025920.
[25] S. Saifullah, R. Dreżewski, F. A. Dwiyanto, A. S. Aribowo, Y. Fauziah, and N. H. Cahyana, "Automated Text Annotation Using a Semi-Supervised Approach with Meta Vectorizer and Machine Learning Algorithms for Hate Speech Detection," *Applied Sciences*, vol. 14, no. 3, p. 1078, Jan. 2024, doi: 10.3390/app14031078.
[26] G. Selvachandran *et al.*, "A New Design of Mamdani Complex Fuzzy Inference System for Multiattribute Decision Making Problems," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 4, pp. 716–730, Apr. 2021, doi: 10.1109/TFUZZ.2019.2961350.

BIOGRAPHIES OF AUTHORS

Tundo completed his undergraduate education in 2017 at Yogyakarta Technology University, Informatics Study Program, then Masters (S2) Informatics Study Program in 2020 at Sunan Kalijaga State Islamic University, Yogyakarta. Currently, he is actively teaching at the STIKOM CKI since 2024 as a permanent lecturer in the Informatics study program. The focus of the research carried out is in the fields of machine learning, artificial intelligence, decision support systems, data mining, and fuzzy logic. He is also active as a reviewer in various accredited national journals (SINTA 4–SINTA 6), as well as non-accredited national journals. Apart from that, he is active in the Higher Education Cooperation Forum (FKPT) and several scientific organizations. He can be contacted at email: asna8mujahid@gmail.com.

Shoffan Saifullah received a Bachelor's degree in Informatics Engineering from Universitas Teknologi Yogyakarta, Indonesia, in 2015 and a Master's Degree in Computer Science from Universitas Ahmad Dahlan, Yogyakarta, Indonesia, in 2018. He is a lecturer at Universitas Pembangunan Nasional "Veteran" Yogyakarta, Indonesia. His research interests include image processing, computer vision, and artificial intelligence. He is currently a Ph.D. student at AGH University of Krakow, Poland, with a concentration in the field of artificial intelligence (bio-inspired algorithms), image processing, and medical image analysis. He can be contacted at email: shoffans@upnyk.ac.id and saifulla@agh.edu.pl.

Mesra Betty Yel was born on 16 May 1972 in Bukit Tinggi. Studied at Universitas Terbuka, Department of Economy, then took a Master of Computer Science at the Universitas Budi Luhur, and completed his Ph.D. (Computer Science) in the Department of Computer Science at the University of Universitas Sumatra Utara. The author is interested in the field of Digital Business, Computer Science, currently active as a lecturer at the University of STIKOM CKI in the Information System Study Program. She can be contacted at email: mesrabettyyel.mby@gmail.com.

Opi Irawansah was born in Langam November 20th 1988. He Completed the Bachelor's Program in Islamic Religious Education at the Al-Aqidah Islamic Institute, Jakarta. He also took a Master's degree in Islamic education management at Muhammadiyah University, Surakarta. He received his Doctoral Program at Sunan Kalijaga State Islamic University, Yogyakarta, in the field of Islamic education. Currently, he is active as a lecturer at Al-Irsyad University Cilacap in the Faculty of Social Sciences, Digital Business Study Program. During his time as a lecturer, he actively carried out research and service in the fields of management and digital business using an Islamic education approach. Among his writings are about Islamic laws relating to the digital world and business. He can be contacted at email: opi_irawansah@universitasalirsyad.ac.id.

Zulfikar Yusya Mubarak born in Cilacap 21 May 1989. Completed the Bachelor's Program in Informatics Engineering at STMIK Tasikmalaya. Next, he took a Master's degree in Informatics Engineering Management at AMIKOM University, Yogyakarta. Currently active as a lecturer at Al-Irsyad University Cilacap in the Faculty of Social Sciences, Digital Business Study Program. During his time as a lecturer, he actively carried out research and service in the fields of digital business technology, digital marketing, UI UX design and data analysis. Among his writings are about technologies used in business digitalization, such as e-commerce platforms, business management software (ERP), data analysis, and cloud computing technology. He can be contacted at email: zulfikar@universitasalirsyad.ac.id.

Andi Saidah was born on December 19 1967 in South Sulawesi. Studied at Hasanuddin University Makassar Faculty of Engineering, Department of Engineering, then took a Master of Engineering, Department of Mechanical Engineering at the University of Indonesia. She is interested in the field of composite materials and renewable energy, currently active as a lecturer at the University of 17 August 1945 Jakarta in the Mechanical Engineering Study Program. She can be contacted at email: andi.saidah@uta45jakarta.ac.id.