A Decision Tree Combination from REP-Random in Creating a Rule to Determine Palm Oil Predictions Using FIS Tsukamoto

Tundo[a]
Universitas 17 Agustus 1945 Jakarta, Jl. Sunter Permai Raya, Sunter Agung, Tanjung Priok, Jakarta, Indonesia
Faculty of Engineering and Informatics, Universitas 17 Agustus 1945, Jakarta, Indonesia

Keywords: Decision Tree, Fuzzy Inference System Tsukamoto, Forecasting, Prediction.

Abstract: This research explains various decision trees and updates to decision trees by carrying out combination scenarios of decision trees that are used in taking a rule that is formed from palm oil production at PT Tapiana Nadenggan which is influenced by factors such as the amount of palm oil, existing demand, and available supplies. The decision trees used in this research, particularly REP and Random, were continued with the combination scenario concept of these decision trees by giving the combined name of the decision trees, specifically REP-Random. The purpose of the decision tree combination is an updated idea related to the combination concept to recognize a performance process from the selected decision tree combination, which in accuracy can exceed the results of the decision tree used or even the results are much worse. To answer this, it is necessary to have a method that is used to find out the performance comparison of decision trees and combinations. The method used is the Tsukamoto Fuzzy Inference System, to determine palm oil production output based on forecasting. After carrying out the forecasting process, the REP-Random combination decision tree is the most recommended alternative, producing the greatest accuracy with a value of 91.18%, followed by REP at 85.68%, and finally Random at 74.14%. The results that have been carried out mean that the decision tree combination scenario can be used to carry out forecasting in creating automatic rules based on the trials that have been carried out.

1 INTRODUCTION

Decision tree is one of the algorithms in data mining which is widely applied as a classification solution (Erol, Tyoden, & Erol, 2018),(Kumar & Chaturvedi, 2020),(Zuo & Guo, 2019). Apart from that, a decision tree is a classification method that uses a tree structure, where each node represents an attribute, the branches represent the attribute value, while the leaves are used to represent classes (Jalota & Agrawal, 2019),(Guler & Ozdemir, 2019),(An, Sun, & Wang, 2017). The top node of this decision tree is called the root (Ogundare & Wiggins, 2018),(Rahat, Kahir, Kaisar, & Masum, 2019). There are many types of decision trees (Hülsmann, Philip, Hammer, Kopp, & Botsch, 2018),(Mesarić & Šebalj, 2016), including those that will be discussed in this research, REP and Random in making rules based on data obtained from PT Tapiana Nadenggan is engaged in palm oil production. Apart from that, this research provides an update that was not available in previous research, which in this research uses a combination concept of decision trees used in combination with REP-Random. The process carried out to produce rules formed from a combination decision tree uses the concept (REP+Random)/2, which means that each result of the decision tree, whether in terms of accuracy, TP Rate, FP Rate, Precision, Recall for each classification class that is developed, will be divided two to obtain the average of the decision trees, the results of which will then be used by a combination of decision trees. All initial processes in creating automatic rules for each decision tree used use WEKA, the stages of the process in creating these rules in detail will be discussed in the next chapter.

After the rules are formed from the selected decision trees, either individual decision trees or combinations are then calculated using the

[a] https://orcid.org/0009-0002-6652-0335

Tundo, .
A Decision Tree Combination from REP-Random in Creating a Rule to Determine Palm Oil Predictions Using FIS Tsukamoto.
DOI: 10.5220/0012583800003821

In *Proceedings of the 4th International Seminar and Call for Paper (ISCP UTA '45 JAKARTA 2023)*, pages 416-421
ISBN: 978-989-758-691-0; ISSN: 2828-853X

Tsukamoto fuzzy inference system method to find out the best rules for determining palm oil production. Efforts to strengthen research and provide original answers to updates carried out by the author, along with similar research that the author took as material in making this research. First, according to (Bhatnagar & Kumar, 2018) conducted research related to the classification of short message service types using a machine learning algorithm that is capable of filtering spam with the aim of applying an effective classifier to classify SMS based on words, instead of running database queries that heavy to compile, load, and match stored regular expressions for each incoming SMS. The data set is available from Netcore Solutions Pvt. Ltd., India, was used to test the proposed approach using a rule-based classifier with a success accuracy of 98.7%. However, the shortcomings of this research are that it is still very minimal in explaining the process of rules being formed and the machine learning method used does not mention the algorithm. Second, research conducted by (Uyun & Choridah, 2018) research that emphasizes the feature selection process using data mining on the results of mammogram image feature extraction. Where the algorithm used to carry out mining, specifically using the classification algorithm: k-nearest neighbors, decision tree, and naive Bayes with a 10-fold cross validation scheme using stratified sampling. There are five parameters which are the best features and contribute to determining the classification of benign and malignant lesions, slices, integrated density, area fraction, model gray value, and center of mass. The best classification results based on the five parameters were produced by the decision tree algorithm with accuracy, sensitivity, specificity, FPR and TPR of 93.18%, 87.5%, 3.89%, respectively 6.33% and 92.11%. The shortcomings of this research are that it is still very minimal in explaining the rule process formed from the classification algorithm used, and a clear explanation has not yet been provided to obtain the best features in determining the classification of benign and malignant lesions. Third, research conducted by (Supianto, Julisar Dwitama, & Hafis, 2018) research which emphasizes the use of decision trees to classify student graduation at the Faculty of Computer Science, Brawijaya University. The classification algorithms used are, Random Tree, REPTree, and C4.5 and compare the accuracy of each selected algorithm. The results obtained show that the average accuracy using the C4.5 algorithm is greater with an accuracy level of 77.01% compared to Random Tree (74.70%) and REPTree (76.75%). The shortcomings of this research are that it is still very minimal in explaining the rule process that is formed from the classification algorithm used, and a clear explanation has not yet been provided until the student's graduation classification is obtained.

Based on the things explained above, this research modeling was carried out with the aim of shortening and speeding up the making of regulations without having to consult with experts. Apart from that, this research will show the process of making a classification of each class that is formed, as well as comparing the decision trees used. Especially, REP, Random and a combination of decision trees, REP-Random as a form of renewal carried out by the author as an effort to differentiate from previous research and as a new form of modeling that can be considered when creating rules for modeling.

2 METHODS

2.1 Dataset

This research will use two decision trees and a combination of decision trees consisting of REP and Random decision trees, while the combination decision tree is a combination of the results of REP-Random. The data collection to create rules for predicting palm oil was obtained from the Tapiana Nadenggan Company from 2014 to 2023 in a monthly period, where in 2023 it only lasted until April. The parameters used in making rules for predictions consist of the number of palm oil, consumer demand, existing supplies, and production quantities. Table 1 explains the dataset used. All data will be saved in *csv format to create rules using WEKA (Azizah, Pujianto, & Nugraha, 2018),(Hasan, Palaniappan, & Rafi, 2018).

Table 1: Dataset.

Palm Oil (kg)	Demand (liter)	Stock (liter)	Production (liter)
20,875,600	4,730,300	3,960,000	10,020,000
26,300,700	14,987,000	4,220,200	19,300,500
26,250,400	14,980,000	4,500,000	19,150,000
38,700,000	3,784,500	1,900,400	10,100,800
24,400,100	7,568,600	4,000,700	13,568,000
26,000,000	12,600,000	1,730,000	17,000,300
34,857,100	10,811,400	3,959,140	18,954,280
....	....	....	....
....	....	....	....
....	....	....	....
40,616,100	14,996,800	4,969,670	19,597,200

2.2 Preprocessing

Data that has been collected for preprocessing will be transformed so that it can be processed by WEKA by changing numerical data into linguistic values (Mishra & Ratha, 2016),(Zhong, Chang, & Zhang, 2018), which consist of Little, Medium, and Many. The division rule for determining the numerical value is to divide the membership function into three parts (Khan & Ahmad, 2019),(Fernandes et al., 2018), the first membership function from the minimum value to the middle value between Little-Medium, the second membership function between the value Little-Medium up to Medium-Many (Xmiddle), and the third membership function between Medium-Many (Xmiddle) up to the maximum value. Table 2 explains the data transformation that has been changed based on the previous explanation.

Table 2: Transformation.

Palm Oil (kg)	Demand (liter)	Stock (liter)	Production (liter)
Little	Little	Medium	Medium
Many	Little	Many	Little
Many	Little	Little	Little
Many	Medium	Medium	Little
Many	Many	Little	Many
Many	Many	Many	Medium
Many	Medium	Little	Medium
Little	Little	Many	Little
Many	Many	Many	Many
Many	Little	Little	Little
Many	Many	Many	Many
Little	Little	Medium	Medium
---------	---------	---------	---------
---------	---------	---------	---------
---------	---------	---------	---------
Many	Many	Little	Many

2.3 Decision Tree

The decision trees used to create rules for palm oil predictions include REP and Random. The update carried out in this research is by combining decision trees by carrying out REP-Random tree combination scenarios with the aim of finding out the level of accuracy produced between individual decision trees and combinations, apart from that as a new form of modeling that can be considered when making rules. for modeling, which then the rule will be linked to Tsukamoto's Fuzzy Inference System (FIS) to find predictions of palm oil production.

2.4 FIS Tsukamoto

FIS Tsukamoto in this case plays a role in producing the magnitude value of a prediction that is built (Tundo & Sela, 2018),(Selvachandran et al., 2019) using rules from the decision tree REP, Random and a combination of REP-Random, so that this research will be more focus on making rules using various kinds of decision trees and then comparing the results of the decision trees with the prediction output produced with FIS Tsukamoto by comparing directly with actual production (Sheena, Ramalingam, & Anuradha, 2017),(Geman, Chiuchisan, & Aldea, 2017). Table 3 explains Tsukamoto's FIS modeling in predicting palm oil with rules built using a decision tree.

Table 3: Tsukamoto's FIS modeling.

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

3 RESULTS AND DISCUSSION

This section presents the results of rule formation using a decision tree to be identified using WEKA (Sutopo et al., 2023),(Erol et al., 2018), where the data processed is data that has been transformed (Pristyanto, Pratama, & Nugraha, 2018). The following are the decision tree accuracy results obtained in detail, seen in Table 4 and Table 5 explaining the accuracy of the TP Rate, FP Rate, Precision, Recall and others from the decision tree used.

Table 4: Details of REP Rule Formation Accuracy.

Parameter	Class		
	Little	Medium	Many
TP Rate	0.783	0.909	1
FP Rate	0	0.077	0.059
Precision	1	0.714	0.935
Recall	0.783	0.909	1
F-Measure	0.878	0.8	0.967
ROC Area	0.954	0.937	0.971
Accuracy		Time	
Correctly Classified	90.48	0.03	
Incorrectly Classified	9.52		
Kappa Statistic	0.8489		
Mean Absolute Error	0.0995		
Root Mean Squared Error	0.2231		
Relative Absolute Error	23.841		
Root Relative Squared Error	48.899		

In Table 4, detailed accuracy results from the REP decision tree show an accuracy of 90.48%, MAE value 0.0995, RMSE 0.2231 in forming rules using WEKA.

Table 5: Details of Accuracy of Random Rule Formation.

Parameter	Class		
	Little	Medium	Many
TP Rate	0.913	0.909	1
FP Rate	0	0.019	0.059
Precision	1	0.909	0.935
Recall	0.913	0.909	1
F-Measure	0.955	0.909	0.967
ROC Area	0.985	0.978	0.98
Accuracy		Time	
Correctly Classified	95.24	0.03	
Incorrectly Classified	4.76		
Kappa Statistic	0.9234		
Mean Absolute Error	0.0536		
Root Mean Squared Error	0.1636		
Relative Absolute Error	12.826		
Root Relative Squared Error	35.866		

The REP-Random combination decision tree process obtained accuracy by dividing the average, where REP-Random was obtained from the average accuracy results of REP and Random. After getting the accuracy of the rule formed above 90%, the results were tested 6 times to generate the maximum rule. The following rules are formed from the decision tree used, shown in Table 6.

Table 6: Rule algorithm decision tree.

Algorithm	Node	Rule
Random	1	IF Demand Little AND Palm Oil Banyak THEN Production Little
	2	IF P Demand Little AND Palm Oil Little AND Stock Banyak THEN Production Little
	3	IF Demand Little AND Palm Oil Little AND Stock Little THEN Production Medium
	4	IF Demand Little AND Palm Oil Little AND Stock Medium THEN Production Medium
	5	IF Demand Many AND Stock Many THEN Production Many
	6	IF Demand Many AND Stock Little THEN Production Many
	7	IF Demand Medium AND Stock Many THEN Production Medium
REP	1	IF Demand Little AND Palm Oil Many THEN Production Little
	2	IF Demand Little AND Palm Oil Little THEN Production Medium
	3	IF Demand Many THEN Production Many
	4	IF Demand Medium THEN Production Medium
	5	IF Demand Little THEN Production Little

In Table 4, detailed accuracy results from the REP decision tree show an accuracy of 90.48%, MAE value 0.0995, RMSE 0.2231 in forming rules using WEKA.

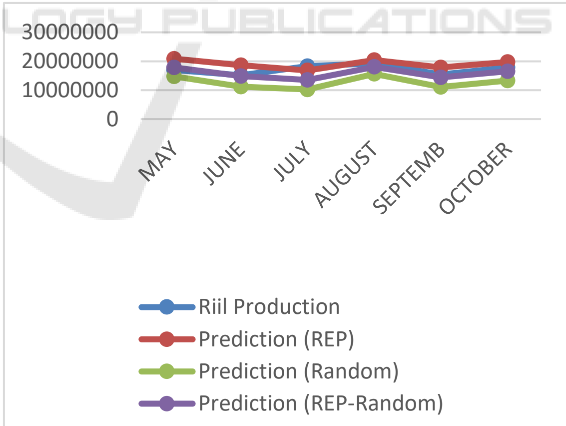

Figure 1: Algorithm Prediction Results are compared with actual production.

Based on Figure 1, the REP-Random and REP algorithms are trending closer to actual production. To find out clearly, error value calculations are used to find out which algorithm is closest to actual production by using the AFER value as shown in Table 7.

Table 7: Prediction Accuracy using AFER.

Algorithm	Rill Production	Prediction	| A – F |/A	Error (%)	Accuracy (%)
REP	16972100	20886261	0.230623258	14.33	85.67
	15170500	18594033	0.225670413		
	18350200	16828045	0.082950322		
	19275000	20446302	0.060767938		
	15432800	17829376	0.155291068		
	17876500	19753264	0.10498498		
Random	16972100	14762132	0.130211818	25.86	74.14
	15170500	11242067	0.258952111		
	18350200	10269662	0.440351495		
	19275000	15631145	0.189045655		
	15432800	11110671	0.280061233		
	17876500	13357868	0.25276939		
Combination					
REP-Random	16972100	17824196.5	0.05020572	8.82	91.18
	15170500	14918050	0.016640849		
	18350200	13548853.5	0.261650908		
	19275000	18038723.5	0.064138859		
	15432800	14470023.5	0.062385082		
	17876500	16555566	0.073892205		

4 CONCLUSIONS

Based on the results of experiments and tests, it can be concluded that decision trees and decision tree combination experiments can be used as an alternative method in assisting the process of predicting palm oil production in making fast, practical, and flexible rules without expert assistance, followed by looking for predicted values using FIS Tsukamoto. Prediction results using the REP-Random combination decision tree are the most recommended alternative, producing the greatest accuracy with a value of 91.18%, followed by REP with a value of 85.67%, and Random with a value of 74.14%.

Suggestions for future researchers with a similar theme include adding the parameters used and providing alternative decision tree combinations that are more diverse so that they can be used as references or findings that can be useful for future researchers.

REFERENCES

An, Y., Sun, S., & Wang, S. (2017). Naive Bayes classifiers for music emotion classification based on lyrics. In *Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017* (pp. 635–638). https://doi.org/10.1109/ICIS.2017.7960070

Azizah, E. N., Pujianto, U., & Nugraha, E. (2018). 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)* (pp. 18–22). IEEE.

Bhatnagar, S., & Kumar, A. (2018). A rule-based classification of short message service type. In *Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC 2018* (pp. 1139–1142). IEEE. https://doi.org/10.1109/ICISC.2018.8398982

Erol, H., Tyoden, B. M., & Erol, R. (2018). Classification Performances of Data Mining Clustering Algorithms for Remotely Sensed Multispectral Image Data. In *2018 IEEE (SMC) International Conference on Innovations in Intelligent Systems and Applications, INISTA 2018* (pp. 1–4). https://doi.org/10.1109/INISTA.2018.8466320

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, G. Van. (2018). Educational data mining : Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 1–9. https://doi.org/10.1016/j.jbusres.2018.02.012

Geman, O., Chiuchisan, I., & Aldea, R. T. (2017). Application of Adaptive Neuro-Fuzzy Inference System for Diabetes Classification and Prediction. In *The 6th IEEE International Conference on E-Health and Bioengineering - EHB 2017* (pp. 639–642).

Guler, E. R., & Ozdemir, S. (2019). Applications of Stream Data Mining on the Internet of Things: A Survey. In *International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism, IBIGDELFT 2018 - Proceedings* (pp. 51–55). https://doi.org/10.1109/IBIGDELFT.2018.8625289

Hasan, R., Palaniappan, S., & Rafi, A. (2018). Student Academic Performance Prediction by using Decision Tree Algorithm. In *2018 4th International Conference on Computer and Information Sciences (ICCOINS) Student* (pp. 1–5).

Hülsmann, F., Philip, J., Hammer, B., Kopp, S., & Botsch, M. (2018). Classification of motor errors to provide real-time feedback for sports coaching in virtual reality — A case study in squats and Tai Chi pushes. In *Computer & Graphics* (Vol. 10). https://doi.org/10.1016/j.cag.2018.08.003

Jalota, C., & Agrawal, R. (2019). Analysis of Educational Data Mining using Classification. In *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019* (pp. 243–247). IEEE. https://doi.org/10.1109/COMITCon.2019.8862214

Khan, I., & Ahmad, A. R. (2019). Tracking Student Performance in Introductory Programming by Means of Machine Learning. In *2019 4th MEC International*

Conference on Big Data and Smart City (ICBDSC) (pp. 1–6). IEEE.

Kumar, K., & Chaturvedi, K. (2020). An audio classification approach using feature extraction neural network classification approach. In *2nd International Conference on Data, Engineering and Applications, IDEA 2020* (pp. 1–6). https://doi.org/10.1109/IDEA49133.2020.9170702

Mesarić, J., & Šebalj, D. (2016). Decision trees for predicting the academic success of students. *Croatian Operational Research Review*, *7*(2), 367–388. https://doi.org/10.17535/crorr.2016.0025

Mishra, A. K., & Ratha, B. K. (2016). Study of Random Tree and Random Forest Data Mining Algorithms for Microarray Data Analysis. *International Journal on Advanced Electrical and Computer Engineering (IJAECE)*, *3*(4), 5–7.

Ogundare, O., & Wiggins, N. (2018). Identifying Sub-documents in a Composite Scanned Document Using Naive Bayes, Levenshtein Distance and Domain Driven Knowledge Base. In *5th International Conference on Soft Computing and Machine Intelligence, ISCMI 2018* (pp. 84–87). IEEE. https://doi.org/10.1109/ISCMI.2018.8703245

Pristyanto, Y., Pratama, I., & Nugraha, A. F. (2018). Data level approach for imbalanced class handling on educational data mining multiclass classification. In *2018 International Conference on Information and Communications Technology, ICOIACT 2018* (Vol. 2018-Janua, pp. 310–314). https://doi.org/10.1109/ICOIACT.2018.8350792

Rahat, A. M., Kahir, A., Kaisar, A., & Masum, M. (2019). Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. In *8th International Conference on System Modeling & Advancement in Research Trends* (pp. 266–270). IEEE Xplore.

Selvachandran, G., Quek, S. G., Thi, L., Lan, H., Son, L. H., Giang, N. L., & Ding, W. (2019). A New Design of Mamdani Complex Fuzzy Inference System for Multi-attribute Decision Making Problems. *IEEE Transactions on Fuzzy Systems*, *PP*(c), 1. https://doi.org/10.1109/TFUZZ.2019.2961350

Sheena, A. D., Ramalingam, M., & Anuradha, B. (2017). A Comprehensive Study on Fuzzy Inference System and its Application in the field of Engineering. *International Journal of Engineering Trends and Technology (IJETT)*, *54*(1), 36–40.

Supianto, A. A., Julisar Dwitama, A., & Hafis, M. (2018). 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*, 308–311. https://doi.org/10.1109/SIET.2018.8693158

Sutopo, J., Khanapi, M., Ghani, A., Burhanuddin, M. A., Septianti, A. N., & Tundo, T. (2023). Dance Gesture Recognition Using Laban Movement Analysis with J48 Classification. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *11*(2), 528–536. https://doi.org/10.52549/ijeei.v11i2.4314

Tundo, T., & Sela, E. I. (2018). APPLICATION OF THE FUZZY INFERENCE SYSTEM METHOD TO PREDICT The Number of Weaving Fabric Production. *(IJID) International Journal on Informatics for Development*, *7*(1), 1–9.

Uyun, S., & Choridah, L. (2018). Feature selection mammogram based on breast cancer mining. *International Journal of Electrical and Computer Engineering*, *8*(1), 60–69. https://doi.org/10.11591/ijece.v8i1.pp60-69

Zhong, S., Chang, C. I., & Zhang, Y. (2018). Iterative Support Vector Machine for Hyperspectral Image Classification. In *Proceedings - International Conference on Image Processing, ICIP* (pp. 3309–3312). IEEE. https://doi.org/10.1109/ICIP.2018.8451145

Zuo, L., & Guo, J. (2019). Customer Classification of Discrete Data Concerning Customer Assets Based on Data Mining. In *Proceedings - 2019 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2019* (pp. 352–355). IEEE. https://doi.org/10.1109/ICITBS.2019.00093