p-ISSN 2615-787X e-ISSN 2615-790X
Accredited by Directorate General of Strengthening for Research and Development No: 30/E/KPT/2018

Tropical Animal Science Journal, December 2022, 45(4):436-442
DOI: https://doi.org/10.5398/tasj.2022.45.4.436
Available online at http://journal.ipb.ac.id/index.php/tasj

Estimation of Harvest Time of Forage Sorghum (Sorghum Bicolor) CV. Samurai-2 Using Decision Tree Algorithm

K. H. Suradiradja[a,*], I. S. Sitanggang[a], L. Abdullah[b], & I. Hermadi[a]
[a]Department of Computer Science, Faculty of Mathematics and Natural Science, IPB University
[b]Department of Nutrition and Feed Technology, Faculty of Animal Science, IPB University
Jalan Agatis, Kampus IPB Darmaga Bogor 16680, Indonesia
*Corresponding author: kahfi.heryandi@apps.ipb.ac.id
(Received 17-09-2021; Revised 22-12-2021; Accepted 18-01-2022)

ABSTRACT

Efforts to improve feed quality by adding additional nutritional supplements can increase production costs due to the increased concentrate prices. Therefore, one option is to combine the main feed with forages containing a high protein source at a low cost, such as Gramineae (e.g., sorghum). This study aims to estimate the harvest time of sorghum when the biomass content, nutrients, and digestibility for livestock are in good condition using a machine learning algorithm, namely a decision tree. The stages of this study include the collection of observation data in the field, preprocessing, modeling, evaluation, and validation. Images and field observations are the primary datasets used. These datasets become the model input for the decision tree algorithm. The results of this study are the classification model for estimating harvest time with an accuracy of 98.86% and the rule that is generated by the decision tree model, the right time to be harvested are in the condition (Day After Planting > 77.5 days AND Day After Planting ≤ 84 days AND Diameter > 26 mm) or (Day After Planting > 84 days AND Height ≤ 138.5 cm AND Leaves > 8.5 pieces) or (Day After Planting > 84 days AND Height > 138.5 cm). In conclusion, the rule generated from the decision tree algorithm can help estimate the fast harvest time of sorghum bicolor cv. Samurai 2.

***Keywords:** decision tree; estimated harvest; forage sorghum; machine learning; sorghum bicolor*

INTRODUCTION

In the livestock industry, the main concern is the availability of land, feed, and livestock. About 70% of the main feed for the cattle industry is grass. However, grass has low protein and high crude fiber contents, resulting in poor livestock digestibility. To improve the nutritional value of ruminant feed, grass as the main feed ingredient must be added to supplementary food or concentrated food, such as corn, dregs, and other similar foods, which increases the production costs. According to Abdullah (2010), the solutions are to provide quality forage in rations so that livestock performance can be maintained as well as to reduce input from concentrate materials to reduce the cost of the ration. The alternative source of high-protein forage at a more economical cost is a combination of the main feed and Gramineae, one of which is sorghum. According to Ardiansyah *et al.* (2016), sorghum can be cultivated and developed in marginal and dry areas in Indonesia. The selection of sorghum as a significant feed on marginal lands is the best solution for ruminants' forage supply. Etuk *et al.* (2012) stated that sorghum is often grown for forage or silage. Sorghum also has a greater biomass than corn (Rocateli *et al.*, 2012). Forage sorghum can be harvested as early as the flower or milk stage (Lyons *et al.*, 2019). To obtain high-quality biomass and fiber, harvesting must be done at the right time. However, we will receive a decision based on the laboratory analysis results in a few days, and the right time should not be missed.

In industrial era 4.0, information technology support is essential in various fields, including machine learning in the livestock sector. A decision tree, one of the machine learning algorithms, is a tree-like structure. Each node represents the value of the attribute being tested, each branch is the result of the test attribute, and the leaf represents a class or class distribution. The decision tree can be easily converted to the classification rules of Han *et al.* (2012). According to Sitanggang & Ismail (2011), two widely used decision tree algorithms are Quinlan's ID3 and C4.5, which is an extension of ID3 and CART (Classification and Regression Tree), respectively. Decision trees have been employed in several agricultural and animal husbandry studies (Maxwell *et al.*, 2018; Klompenburg *et al.*, 2020; McHugh *et al.*, 2020). According to Larose (2014), the two leading algorithms for building decision trees are the CART and C4.5. In this study, the algorithm used is the CART algorithm. The confusion matrix measurement method was employed as an evaluation method, with a measuring instrument in the form of a 2 (×) 2 matrix used to obtain dataset classification accuracy for active and inactive classes. According to Witten *et al.* (2017), the confusion matrix evaluates the classification model by determining which testing object is predicted to be true and not true. The confusion matrix is used to measure the accuracy, precision, and F-measure values. According to Gorunescu (2011), the accuracy value has a diagnosis level with a value between 0.90 and 1.00 indicating excellent classification; 0.80 and 0.90 for good classification; 0.70 and 0.80 for fair classification; 0.60 and 0.70 for poor classification; and 0.50 and 0.60 for a failure.

Several other studies related to sorghum use machine learning in addition to decision trees, such as those studies prediction of sorghum biomass (Masjedi *et al.,* 2019), Sorghum Head Detection and Counting (Ghosal *et al.,* 2019), and Yield Prediction of Sorghum and Machine Learning (Varela *et al.,* 2021). This study aimed to make a rule decision to estimate the harvest time using a decision tree algorithm. It is hoped that the algorithm model can help estimate the accurate harvest time for sorghum bicolor cv. Samurai 2.

MATERIALS AND METHODS

The research area for this plant is in Jonggol Animal Science Teaching and Research of IPB Unit, Singasari Village, Malati Village, Jonggol District, Bogor Regency, above sorghum research land of one hectare at coordinates latitude 6°28’ 9.4”– 6°28’ 12.3” and longitude 107°00’49.1” - 107°00’50.6”.

Sample Collection Dataset

The dataset obtained from the field observations is presented in a tabular format, as shown in Table 1, and images of the plant objects are presented in Figure 1. Every week, both are recorded and taken. A total of 1059 tabular data points were obtained from 55 plant samples throughout one cycle, which lasted from the fourth week after planting to the 13th week of the first ratoon.

Methods

The research methodology has five main phases: a preliminary study, data collection, preprocessing, modeling, and model evaluation—each phase in this research flow cycle chart explains each process.

Preliminary study. In the research understanding phase, an estimation model for harvest time is developed through problem identification, formulation of research objectives, identification of research scope, and literature review.

Data collection. The data used for the research were obtained from the field observations, starting from planting seeds to two harvests. Furthermore, the text data for recording observations in the field with the dataset attributes are presented in Table 1.

Preprocess. In the preprocessing stage, plant dataset records in which some data attributes have no or null value due to damaged sample data are deleted. Ignoring the tuple is not very effective unless the tuple contains

Table 1. Dataset sorghum of bicolor cv. Samurai-2

No	Attribute	Data type	Description
1	BLOCK	Text	Plant block
2	DATE	Date	Plant growth date
3	OBSERVED_BY	Text	Name of Observed
4	NO_SAMPLE	Text	No Sample, A and B are in one location where 2 trees grow
5	LATITUDE	Text	Latitude
6	LONGITUDE	Text	Longitude
7	HST	Numeric	Day After Planting
8	HEIGHT_CM	Decimal	Plant height in centimeters
9	DIAMETER_MM	Decimal	Plant stem diameter in millimeters
10	STEM	Numeric	Number of sticks
11	LEAVE	Numeric	Number of leaves in one tree
12	HARVESTED	Numeric	Target class (1 = Harvest time or 0 = not yet)

Figure 1. Sample image of sorghum bicolor cv. Samurai-2.

several attributes with missing values of Han *et al.* (2012). Table 1 presents the attributes of the dataset obtained from weekly observations, and the attributes ready for use in the modeling process are days after planting (HST), height (cm), diameter (mm), leaf, and harvest. Experts determine the attributes of the Harvest target class, namely, by assessing existing attribute data with plant images.

Modeling. The preprocessed dataset becomes the model input for the decision tree algorithm at this stage. The datasets were analyzed in Python 3.7 using the scikit-learn library module, which contains several main packages of classes for split dataset training and tests, classes for decision tree models, class confusion matrix, and cross-validation. The maximum depth parameter that forms the decision tree will not be overly complex in the decision tree model. The dataset was divided into two data groups. A classification algorithm analyzes training data, and test data are used to estimate the accuracy of the classification rules. If the accuracy is considered acceptable, the rules can be applied to classify new tuples (Han *et al.,* 2012). The modeling process cycle is from modeling to evaluation to analyze the accuracy level of the predictions generated from the model.

Evaluation. The evaluation process stage of the output of the modeling process uses confusion matrix validation, which aims to achieve model accuracy with computational efficiency. The success of this phase is indicated by an excellent accuracy value (0.90–1.00) of the resulting model (Gorunescu, 2011).

RESULTS

The observation dataset has 1,059 data, including statistical data, as presented in Table 2; it consists of 246 and 813 harvested and unharvested data, respectively. The dataset attributes used are HST_days, height_cm, diameter_mm, leaf, and one attribute of the harvest target class.

There were 794 and 265 rows of data for training and testing, respectively. A model was created with the DecisionTreeClassifier package from scikit-learn for the training process using training data. It set the maximum depth in the decision tree model to 4, making the formed decision tree less complex and the cross-validation parameter tenfold. Gini impurity was used by default in the function to measure the quality of a split in the DecisionTreeClassifier parameter. The Gini index of each attribute was first calculated during the modeling process of forming a decision tree. Then, the value of each Gini-Split was calculated. The attribute with the smallest Gini-Split value was the HST (0.355), so it was assigned to the root node, as presented in Figure 2. The same procedure was done for the next level to look for the selected attributes by calculating the Gini index and Gini-Split with the lowest value. Attributes with a

Table 2. Statistic dataset of sorghum bicolor cv. Samurai-2

	HST_days	Height_cm	Diameter_mm	Leaf
Count	1059.00	1059.00	1059.00	1059.00
Mean	67.46	147.82	16.43	9.78
Std	20.90	61.49	5.69	2.27
Min	32.00	0.00	0.00	0.00
25%	46.00	103.00	12.90	8.00
50%	73.00	154.00	17.00	10.00
75%	87.00	196.50	20.40	11.00
Max	99.00	366.00	58.00	17.00

Note: HST= days after planting.

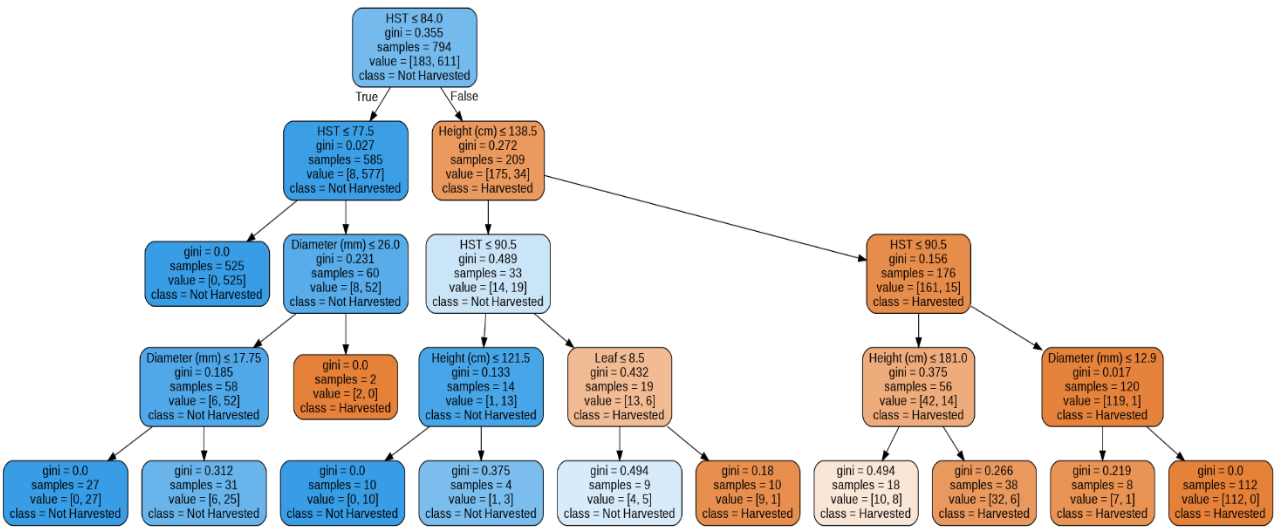

Figure 2. Result of the decision tree sorghum bicolor cv. Samurai-2; HST= days after planting.

Gini index of 0 do not need to be calculated and directly become leaf nodes. The results of the modeling process using the sorghum bicolor cv. Samurai-2 dataset produced a decision tree for estimating the harvest time or not the harvest time, producing 11 leaves and 23 nodes, as presented in Figure 2.

The decision tree result indicated that the resulting root node was in the HST attribute. That makes the leaf node the class in the Harvest target class attribute, containing "Harvested" and "Not Harvested."

In determining the classification for good accuracy, cross-validation was conducted ten times (Witten *et al.*, 2017). The results of the confusion matrix evaluation are presented in Table 3, with cross-validation k-fold is ten producing an accuracy value of 0.99 decision tree algorithm for estimation of harvest time of sorghum bicolor cv. Samurai-2.

DISCUSSION

Related research uses the same object, namely sorghum. However, the objectives, data, and algorithms used are different, such as research by Masjedi *et al.* (2019) and Ghosal *et al.* (2019). Meanwhile, in this study, we found the rules for harvesting sorghum based on the output of the decision tree model using plant observation data.

In the process, the target in the dataset is the classification of harvest time determined by experts based on tabular data in the field and plant images, namely, "Harvest" and "Not yet Harvested." Datasets that already have classes become input for training and testing data on machine learning models with one of the decision tree algorithms, CART. Generally, the machine learning methodology involves a learning process to learn from "experience" (training data) in carrying out tasks and is measured by performance metrics (Liakos *et al.*, 2018). The result of accuracy decision tree model of this research is 98.86, compared with research by Masjedi *et al.* (2019) using Recurrent Neural Network Gated Recurrent Unit (RNN-GRU) with a result of R^2 is 0.84 and research by Ghosal *et al.* (2019) using Convolution Neural Network algorithm with a result of R^2 is 0.88.

The output of the decision tree model produces decision tree rules that can be described and implemented as a logical model as follows:

Rule 1: IF Day After Planting ≤ 84 days THEN Class is **"Not Harvested"**

Rule 2: IF Day After Planting > 77.5 days AND Day After Planting ≤ 84 days AND Diameter ≤ 26 mm THEN Class is **"Not Harvested"**

Table 3. The results of the confusion matrix evaluation of the decision tree model

	Predicted	
	Harvested	Not Harvested
Real Harvested	57	6
Real Not Harvested	3	199

Notes: Cross-Validation, Fold: 10;Accuracy: 0.9886039886039887.

Rule 3: IF Day After Planting > 77.5 days AND Day After Planting ≤ 84 days AND Diameter > 26 mm THEN Class is "**Harvested"**

Rule 4: IF Day After Planting > 84 days AND Day After Planting ≤ 90.5 days AND Height ≤ 138.5 cm THEN Class is **"Not Harvested"**

Rule 5: IF Day After Planting > 84 cm AND Height ≤ 138.5 cm AND Leaves ≤ 8.5 THEN Class is **"Not Harvested"**

Rule 6: IF Day After Planting > 84 days AND Height ≤ 138.5 cm AND Leaves > 8.5 THEN Class is "**Harvested"**

Rule 7: IF Day After Planting > 84 days AND Height > 138.5 cm THEN Class is "**Harvested"**

This study used the confusion matrix to analyze how well our classifier can recognize tuples of different classes. The confusion matrix used for this research has two classes presented in Table 3. They are positive tuples when harvested and negative tuples when not harvested. The positive tuples correctly labeled by the classifier are referred to as the 189 true positives. Similarly, true negatives are the 57 negative tuples correctly labeled by the classifier. False positives are negative tuples that have been incorrectly labeled as those six tuples of class harvested that the classifier predicted would not be harvested. While false negatives are the incorrectly labeled positive tuples, the three tuples of class not harvested for which the classifier predicted as harvested. For example, three data in false negatives, are the first, 87 days after planting with height 202 cm have diameter 15.80 mm and have ten leaves. The second, 87 days after planting, has a diameter of 12.30 mm and eight leaves, and the third, 87 days after planting, has a diameter of 19.70 mm and 12 leaves. The reality in the field is that the three plants do not have harvest characteristics, such as the color of the seeds is not milky yet or the seed is not growing yet. The harvest is based on the logical rule model generated from the decision tree model that the plant age > 84 and height > 138.5 cm were harvested.

An example is the rule application from the decision tree for the results of the "Not Harvested" classification target class based on the tabular data in Table 4 and the observation image in Figure 3, sample number 33B on March 3rd, 2021, namely, on HST (Day After Planting) 94; the suitability of the decision tree in Figure 4 or the rule generated by the model expressed as follows:

Rule 1: IF Day After Planting ≤ 84 days THEN Class is **"Not Harvested"** (False)

Rule 2: IF Day After Planting > 77.5 days AND Day After Planting ≤ 84 days AND Diameter ≤ 26 mm THEN Class is **"Not Harvested"** (False)

Rule 3: IF Day After Planting > 77.5 days AND Day After Planting ≤ 84 days AND Diameter > 26 mm THEN Class is "**Harvested"** (False)

Rule 4: IF Day After Planting > 84 days AND Day After Planting ≤ 90.5 days AND Height ≤ 138.5 cm THEN Class is **"Not Harvested"** (False)

Rule 5: IF Day After Planting > 84 cm AND Height ≤

Table 4. Dataset date 03/03/2021 sample numbers 32–35

BLOCK	DATE	OBSERVED _BY	NO_ SAMPLE	LATITUDE	LONGITUDE	HST	HEIGHT _CM	DIAMETER _MM	STEM	LEAF	HARVEST
Samurai-2	3/3/2021	Kahfi	32B	−646852	10701099	94	222	20,2	1	10	0
Samurai-2	3/3/2021	Kahfi	33A	−646841	10701105	94	189	17,6	1	10	0
Samurai-2	**3/3/2021**	**Kahfi**	**33B**	**−646841**	10701105	94	120	12,8	1	8	1
Samurai-2	**3/3/2021**	**Kahfi**	**34A**	**−646829**	10701107	94	184	16,6	1	10	0
Samurai-2	3/3/2021	Kahfi	34B	−646829	10701107	94	212	17,2	1	9	0
Samurai-2	3/3/2021	Kahfi	35A	−646826	10701102	94	201	16	1	10	0

Note: HST= days after planting.

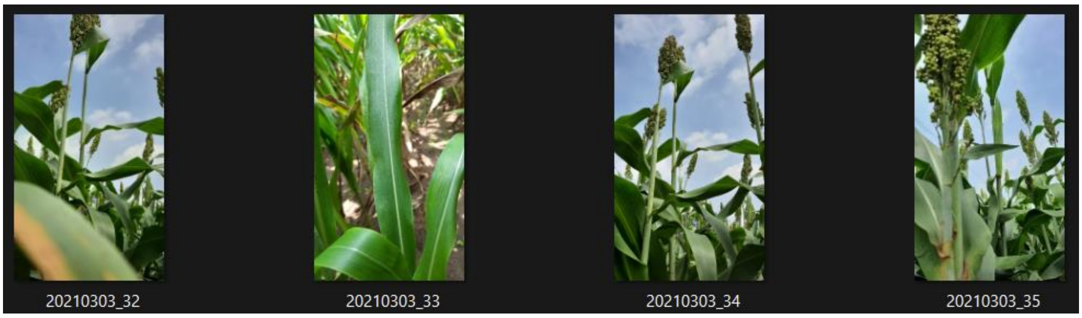

Figure 3. Citra Date March 3rd, 2021 sample numbers 32–35.

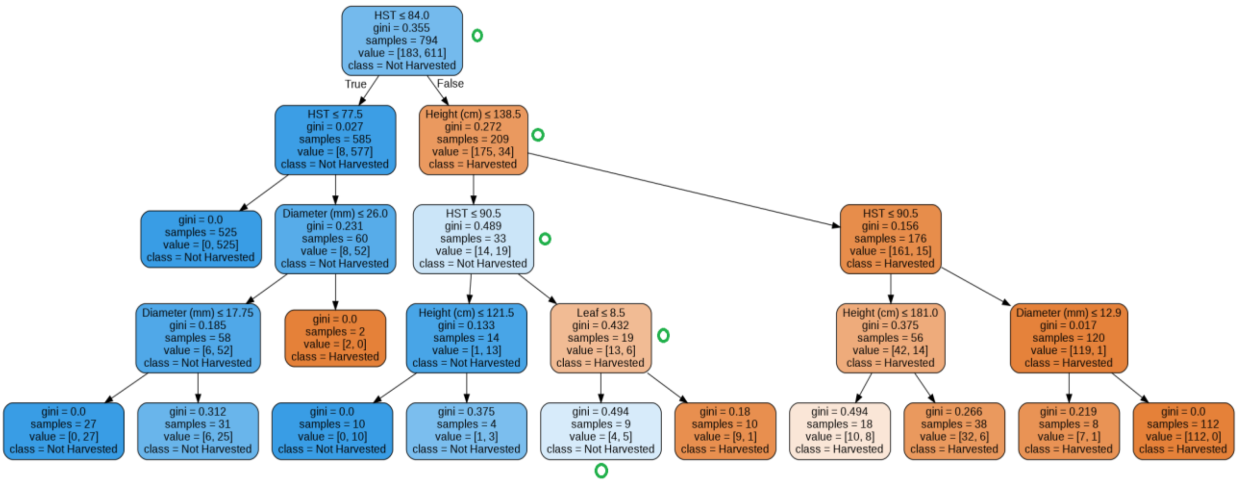

Figure 4. Decision tree for sample number 33B

138.5 cm AND Leaves ≤ 8.5 THEN Class is **"Not Harvested"** (True)

Rule 6: IF Day After Planting > 84 days AND Height ≤ 138.5 cm AND Leaves > 8.5 THEN Class is "**Harvested"**

Rule 7: IF Day After Planting > 84 days AND Height > 138.5 cm THEN Class is "**Harvested"**

Thus, the sample number 33B is not harvested based on the rule model as it matches rule 5. As can be seen from Figure 3, the color of the leaves is not dark green yet, and we cannot see seeds, although HST (Day After Planting) is 94 days.

The next example is the application of the rule from the decision tree for the target class "Harvest" based on the tabular data in Table 4 and the observation image in Figure 3, sample number 34A on March 3rd,2021, namely, at HST (Day After Planting) 94; the suitability of the decision tree in Figure 5 or the rule generated by the model as expressed as follows:

Rule 1: IF Day After Planting ≤ 84 days THEN Class is **"Not Harvested"** (False)

Rule 2: IF Day After Planting > 77.5 days AND Day After Planting ≤ 84 days AND Diameter ≤ 26 mm THEN Class is **"Not Harvested"** (False)

Rule 3: IF Day After Planting > 77.5 days AND Day After Planting ≤ 84 days AND Diameter > 26 mm THEN Class is "**Harvested"** (False)

Rule 4: IF Day After Planting > 84 days AND Day After Planting ≤ 90.5 days AND Height ≤ 138.5 cm THEN Class is **"Not Harvested"** (False)

Rule 5: IF Day After Planting > 84 cm AND Height ≤ 138.5 cm AND Leaves ≤ 8.5 THEN Class is **"Not Harvested"** (False)

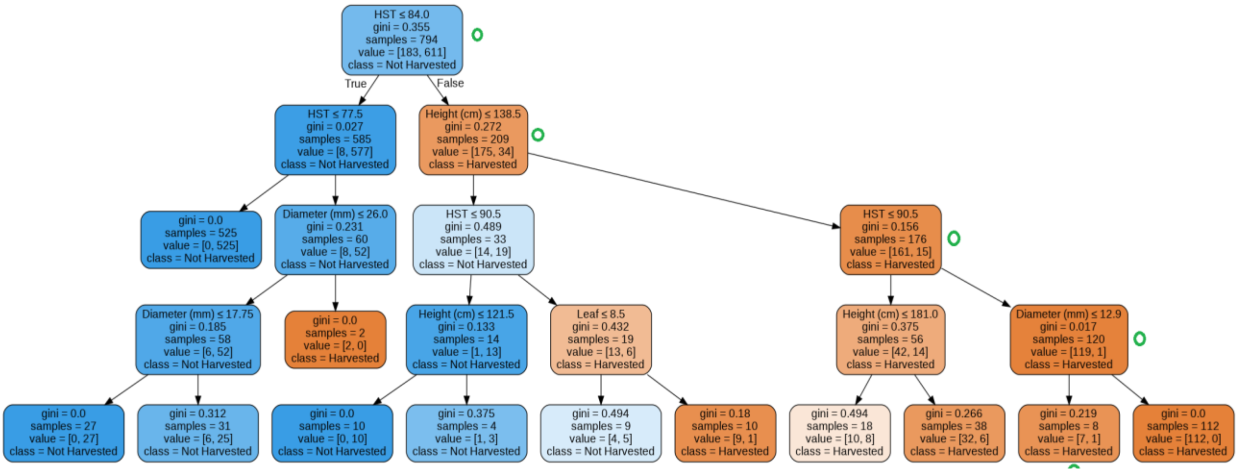

Figure 5. Decision tree for sample number 34A

Rule 6: IF Day After Planting > 84 days AND Height ≤ 138.5 cm AND Leaves > 8.5 THEN Class is "**Harvested**" (False)

Rule 7: IF Day After Planting > 84 days AND Height > 138.5 cm THEN Class is "**Harvested**"

(True) Based on the rule model, the sample number 34A is harvested as it matches rule 7. As can be seen from Figure 3, the color of the leaves is dark green, and we can see a milky white seed, indicating that it can already be harvested.

CONCLUSION

The decision tree algorithm was found to have an accuracy rate of 98.86%, related to an excellent classification category. The decision tree algorithm can estimate the harvest time of Sorghum bicolor cv Samurai-2. According to the rule generated by the decision tree model, the right time for harvest is the condition (Day After Planting > 77.5 days AND Day After Planting ≤ 84 days AND Diameter > 26 mm) or (Day After Planting > 84 days AND Height ≤ 138.5 cm AND Leaves > 8.5 pieces) or (Day After Planting > 84 days AND Height > 138.5 cm).

CONFLICT OF INTEREST

Luki Abdullah serves as an editor of the Tropical Animal Science Journal, but has no role in the decision to publish this article. We certify that there is no conflict of interest with any financial, personal, or other relationships with other people or organizations related to the material that has been discussed in the manuscript.

ACKNOWLEDGEMENT

We would like to thank the Ministry of Research and Technology/National Research and Innovation Agency for funding this doctoral dissertation research (No. 2003/IT3.L1/PN/2021).

REFERENCES

Abdullah, L. 2010. Herbage yield and quality of two vegetative parts of Indigofera at different times of first regrowth defoliation. Med. Pet. 33:44–49. https://doi.org/10.5398/medpet.2010.33.3.169

Ardiansyah, K. G. Wiryawan, & P. D. M. H. Karti. 2016. Silage quality of sorghum harvested at different times and its combination with mixed legumes or concentrate evaluated *in vitro*. Med. Pet. 39:53-60. https://doi.org/10.5398/medpet.2016.39.1.53

Etuk E. B., A. V. Ifeduba, U. E. Okata, I. Chiaka, Okoli, C. Ifeanyi, N. J. Okeudo, B. O. Esonu, A. B. I. Udedibie, & J. C. Moreki. 2012. Nutrient composition and feeding value of sorghum for livestock and poultry: A review. J. Anim. Sci. Adv. 2:510-524.

Ghosal S., B. Zheng, S. C. Chapman, A. B. Potgieter, D. R. Jordan, X. Wang, A. K. Singh, A. Singh, M. Hirafuji, & S. Ninomiya. 2019. A weakly supervised deep learning framework for sorghum head detection and counting. Plant Phenomics. 2019:1–14. https://doi.org/10.34133/2019/1525874

Gorunescu, F. 2011. Data Mining Concepts, Models, and Techniques. 12th ed. Springer. https://doi.org/10.1007/978-3-642-19721-5

Han, J., M. Kamber, & J. Pei. 2012. Data Mining: Concepts and Techniques. 3rd ed. In Morgan Kaufmann Publishers.

Klompenburg, T., A. Kassahun, & C. Catal. 2020. Crop yield prediction using machine learning: A systematic literature review. Comput. Electron. Agric. 177:105709. https://doi.org/10.1016/j.compag.2020.105709

Larose, D. T. 2014. Discovering Knowledge in Data: An Introduction to Data Mining. 2nd ed. John Willey and Son, Inc. https://doi.org/10.1002/9781118874059

Liakos, K. G., P. Busato, D. Moshou, D. Bochtis, & S. Pearson. 2018. Machine learning in agriculture: A review. Sensor. 18:2674. https://doi.org/10.3390/s18082674

Lyons, S. E., Q. M. Ketterings, G. S. Godwin, D. J. Cherney, J. H. Cherney, M. E. Amburgh, J. J. Meisinger, & T. F. Kilcer. 2019. Optimal harvest timing for brown midrib forage sorghum yield, nutritive value, and ration performance. J. Dairy Sci. 102:7134-7149. https://doi.org/10.3168/jds.2019-16516

Masjedi A., N. R. Carpenter, M. M. Crawford, M. R. Tuinstra. 2019. Prediction of Sorghum Biomass using UAV Time Series Data and Recurrent Neural Networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition

Workshops. pp. 2695–2702. https://doi.org/10.1109/CVPRW.2019.00327

Maxwell, A. E., T. A. Warner, & F. Fang. 2018. Implementation of machine-learning classification in remote sensing: An applied review. Int. J. Remote Sens. 39:2784–2817. https://doi.org/10.1080/01431161.2018.1433343

McHugh, O., J. Liu, F. Browne, P. Jordan, & D. McConnell. 2020. Data-Driven Classifiers for Predicting Grass Growth in Northern Ireland: A Case Study. In Communications in Computer and Information Science. Vol. 1237 CCIS. Springer International Publishing. https://doi.org/10.1007/978-3-030-50146-4_23

Rocateli, A. C., R. L. Raper, K. S. Balkcom, F. J. Arriaga, & D. I. Bransby. 2012. Biomass sorghum production and components under different irrigation/tillage systems for the southeastern U.S. Ind. Crops Prod. 36:589–598. https://doi.org/10.1016/j.indcrop.2011.11.007

Sitanggang, I. S. & M. H. Ismail. 2011. Classification model for hotspot occurrences using a decision tree method. Geomat Nat. Hazards Risk. 2:111–121. https://doi.org/10.1080/19475705.2011.565807

Varela, S., T. Pederson, C. J. Bernacchi, & A. D. B. Leakey. 2021. Understanding growth dynamics and yield prediction of sorghum using high temporal resolution UAV imagery time series and machine learning. Remote Sens. 13:1763. https://doi.org/10.3390/rs13091763

Witten, I. H., E. Frank, M. A. Hall, & C. J. Pal. 2017. Data Mining: Practical Machine Learning Tools and Techniques. 4th ed. In The Morgan Kaufmann Series in Data Management Systems.