IJCAT - International Journal of Computing and Technology, Volume 7, Issue 3, March 2020
ISSN (Online) : 2348-6090
www.IJCAT.org
Impact Factor: 0.835

Evaluation of Decision Tree Algorithms in Precision Agriculture

[1] Kalesanwo Olamide; [2] Awodele Oludele; [3] Eze Monday; [4] Kuyoro 'Shade; [5] Ajaegbu Chigozirim

[1, 2, 3, 4, 5] Babcock University, Ilishan Remo
Ogun State, Nigeria

Abstract - Precision Agriculture (PA) is a farm management approach which ensures that plants and animals' needs are adequately met. One of the advances in PA is Autonomous Irrigation System (AIS). Various AISs have been proposed to ensure effective management of water resources, soil water content optimization and increasing crop yield. However, these systems require some level of decision-making algorithm in order to make the appropriate irrigation decisions that are critical. Decision tree algorithms were evaluated and results are presented in this study. The results of evaluation showed that CART recorded an increase of 0.06% and 0.26% compared to C5.0 and ID3 respectively in terms of accuracy. CART also records an increase of 0.38% and 1.76% against the C5.0 and ID3 respectively with regard to precision. Evaluating the recall, CART records an increase of 0.12% and 0.33% in comparison to C5.0 and ID3 respectively. The F-measure of CART also records an increase of 0.24% and 1.05% against the C5.0 and ID3 respectively.

Keywords - *Autonomous Irrigation Systems, Data validation, Decision making, Irrigation, Precision Agriculture.*

1. Introduction

Artificial Intelligence (AI) is aimed at the development and design of machines that have the ability to behave intelligently or make systems and machines perform activities that require some level of intelligence [1]. It has penetrated various sectors of which Agriculture is one of them. The agricultural sector has grown to utilize intelligent systems in carrying out agricultural practices ranging from planting, harvesting, storage, disease identification, control practices, livestock management approach. Agriculture was once a field that had the least technological intervention; it was regarded as the least digitized industry in the world [1]. Over time, the agricultural sector had developed with the introduction of disrupting technologies which have given rise to a new technological trend known as "agritech". Agritech is a term coined from the words Agriculture and Technology and it refers to technologies that can be used in agriculture to enhance efficiency and increase profitability. Technology application in the field of agriculture has the potential of ensuring that precise requirements are given to crops and plants. This will no doubt increase productivity [2], reduce food wastage [3], improve the supply and distribution chain [4], ensure effective crop and livestock management, to mention a few. Unlike Agritech that allows all forms of technology, Precision Agriculture (PA) is a farm (be it livestock or plantation farm) management approach that utilizes Information Technology (IT) to ensure that optimum need of plants and animals are met in an efficient and effective way while guaranteeing maximum productivity and optimum health of farm produce Smart devices and technologies are incorporated so as to evaluate the entities that contribute to the growth and development of farm produce (be it plants or animals). Based on the outcome of this evaluation, standardized farm practices and activities are carried out, strategic decisions are made to increase yield and production and ensure maximal utilization of available resources.

With the PA approach, intelligent systems could be used to sense and understand the farm environment, and decisions will be taken based on information acquired. The aim of PA is to determine the precise needs and practices for agricultural products so as to efficiently increase productivity and maximally utilize resources with the help of intelligent devices. The role of data analytics in PA cannot be undermined as multiple data are collected from sensors and there is a need to interpret data and make meaning out of acquired data. Several data analytics approaches can be adopted based on the type of data collected. Using these data, several tools are developed to aid decision making for efficient farming practices [5]. Decisions such as type of crop variety to grow, rate of sowing, when to irrigate, fertilizer and herbicide application, and other agronomy practices are key to improving crop yield and productivity. For effective management of available resources, the decision on appropriate proportion is also required. The value of data analytics lies in the information and new insights that could be obtained from acquired data [6]. In other words, the effectiveness of the decision(s) made is reliant on the type of data collected by the sensors which can be integrated to the decision support system or processed for further analytics [5]. Machine learning models are used to analyze this data and solve prediction and classification problems. In making valid decisions in PA, some of the existing decision-making approaches include Neural Network (NN) which mimics the human brain. NN use a multifactorial analytic approach [7]. Rule-based approach is another decision making technique which is the most common type of technique used in decision making. Decisions are made based on available information, Bayesian network is yet another approach that is used in reasoning when there is a level of uncertainty. It is dependent on probabilistic inference- [8], [9], Semantic analysis [10], uses the similarity measure. With data analytics, smart decisions and management are supported. However, [11] noted that the creation of an integration platform for big data analytics is yet to be implemented.

The everincreasing population, limited resources, and degrading environmental conditions are major challenges facing the increase in food production. In overcoming the challenge of food shortage, several advancements have been made in the agricultural sector especially in PA to ensure increased food production. One of such advancement is the Autonomous Irrigation Systems (AIS). A number of researchers have proposed and developed several AIS in the PA domain to ensure effective management of water resources, optimizing water content in the soil, and increasing crop yield as a result of appropriate and timely irrigation practices. These systems make decisions based on acquired data that affects quality of decision. Decisions made by AISs need to be dependable as they affect food production. This study therefore evaluates some decision making algorithms for AISs.

2. Related Works

A framework proposed by [12] used fuzzy logic to control the valve control systems of an irrigation system after collecting diverse environmental data (such as soil temperature, environmental humidity, environmental temperature, sunlight intensity, soil moisture, CO2). Structural similarity, as well as neural network based predictions, were used to determine the required soil moisture content to be supplied to the farmland. With structural similarity, farm regions having water deficiency can easily be identified. This will aid decisions as regards when to irrigate the farmland.

A hybrid decision support system for crop-specific irrigation in PA was proposed by [12]. The proposed system utilized the Partial Least Square Regression (PLSR) and fuzzy logic. The proposed system constantly acquires real-time soil and environmental data using WSNs that utilizes Zigbee. The collected data are analyzed and used in making hourly predictions of soil moisture contents. The performance of the decision support system was evaluated using the R-squared, Root Mean Square Error (RMSE), Mean Square Error (MSE) and Ratio of Performance to Deviation (RPD). The developed model can only be used for soil moisture predictions.

Using the neural network and fuzzy logic approach a decision-making model for irrigation notification and control in PA was proposed by [13]. The soil moisture was predicted and the algorithm was implemented by collecting field data on an hourly basis. The model helps to compensate for the amount of water lost to evapotranspiration. However, this model does not seem to be site-specific and some areas of the farmland could be overflooded.

Using the Association Rule Mining (ARM) technique to make decisions as regards irrigation practices, this approach was proposed by [14]. In their research, sensors were used to collect data from the field and these data were sent to the base station. At the base station, the ARM was used to generate rule from the massive amount of data gathered by the sensor. The rules generated were ranked based on the data acquired. Based on these generated rules and its ranking, decisions to irrigate or not were made.

Using the C5.0 Advanced Decision Tree (ADT) classifier algorithm, [15] developed a system that could be used to predict crop selections based on the soil fertility level. The data used for the analyses were captured from sensor devices that collect information about the soil moisture, soil PH value, type of soil. This system helps farmers in deciding what type of crops to plant in the bid of increasing productivity.

The data are collected from the database where collected data are stored. The collected data are stored in the cloud environment. From these data, the data are preprocessed and thereafter features are extracted. From these extracted features, the classifier algorithm (C5.0) is used to predict the soil fertility level and based on that, crops are suggested to the farmers.

3. Methodology

The decision tree algorithms selected for evaluation are; C5.0, Iterative Dichotomiser 3 (ID3) and Classification and Regression Tree (CART). These decision tree algorithms were evaluated using the accuracy, precision and recall. These algorithms were selected because of the type of data that was processed which is a multivariate time series data. Based on the review of related works, these algorithms have been recorded to be used in analyzing these types of data. R 3.6.2 version was used to analyze the algorithms. Datasets were acquired from https://eip.ceh.ac.uk/data. Based on the dataset acquired, features were selected by weighing each attributes/features present in the data set using the value of information they provided which is dependent on the information theory. Feature extraction helps to eradicate overfitting of dataset, speedup training time, enhance explainability of model, as well as improving accuracy [15].

In selecting the features, entropy, information gain, and gain ratio of each attributes were considered. Entropy is a measure from information theory. It helps determine the level of purity or homogeneity of an arbitrary collection of data. Entropy values vary smoothly between 1 and 0. Entropy value is 0 when the data is completely pure and homogeneous and it is 1 when the data is completely impure and heterogeneous [16]. Entropy is mathematically defined as shown in (1)

$$H(X) = \sum_{x \in range(X)} p_x(x).\, log_2 \frac{1}{p_x(x)} \quad (1)$$

Where $p_x(x)$ = Pr $[X = x]$

Information Gain is the change in information entropy of an attribute or feature. It is the amount of information before split occurred excluded from information left after split occurred. It is mathematically denoted in (2) below

$$IG(X, a) = H(X) - H(X|a) \quad (2)$$

Gain Ratio: Gain ratio takes a step further than information gain by sizer of branches when selecting attributes/features. it helps reduce bias towards multivalued attributes. This is mathematically denoted in (3) below

$$IG(X, a) = H(X) - \sum_{v \in range(a)} \left(\frac{|\{x \in X | values(x, a) = v\}|}{|x|} .\, H(\{x \in X | values(x, a) = v\}) \right)$$

The accuracy and precision of the algorithm were calculated using the formula in Equation 4 and Equation 5 respectively, while the recall and F-measure was calculated using Equation 6 and 7 respectively.

$$Accuracy = \frac{True\ positive\ + True\ Negative}{True\ positive + False\ positive + False\ Negative + True\ Negative}$$

(4) [17]

$$precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (5)$$

$$recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (6)$$

$$F - measure = 2 * \frac{precision * recall}{precision + recall} \quad (7)$$ [18]

4. Outcome of the Study

Choosing the right features from your data sets can make all the difference between medium performance which has recorded a long training time and great efficiency with short training time [15], [19]. Amidst thirteen attributes present in the dataset to be used for evaluation, feature selection approach was used to determine the attributes that influence the decision on the dataset. Information gain, gain ratio and entropy of the dataset were used with the ranker search method to weigh the relevance of each attribute. Table 1 shows the result of the information gain and gain ratio of each attribute. Figure 1 depicts the graphical presentation of the information gain for each attribute while Figure 2 depicts the graphical presentation of the Gain ratio for each attribute. Figure 3 shows the entropy value derived for each attribute while Figure 4 shows the average of the information gain and gain ratio for each attribute.

IJCAT - International Journal of Computing and Technology, Volume 7, Issue 3, March 2020
ISSN (Online) : 2348-6090
www.IJCAT.org
Impact Factor: 0.835

Table 1: Information Gain, Gain Ratio and Entropy of attributes

List of attributes	Information Gain	Gain Ratio	Average	Entropy
Soil State	0.4462	0.71201	0.579105	0.6266996
Air Temperature	0.4527	0.43773	0.445215	13.65697
Humidity Level	0.0863	0.39921	0.242755	0.2160794
Soil Temperature	0.2393	0.06334	0.15132	10.11912
Relative Humidity(%)	0.2361	0.06188	0.14899	9.814261
Wetness (%)	0.1823	0.03305	0.107675	7.117012
Volumetric Water Content	0.1635	0.03301	0.098255	10.48187
Water Present Status	0.0363	0.0496	0.04295	0.7310237
Record	0.0673	0.01508	0.04119	6.15908
DateTime	0.0689	0.00412	0.03651	6.551994

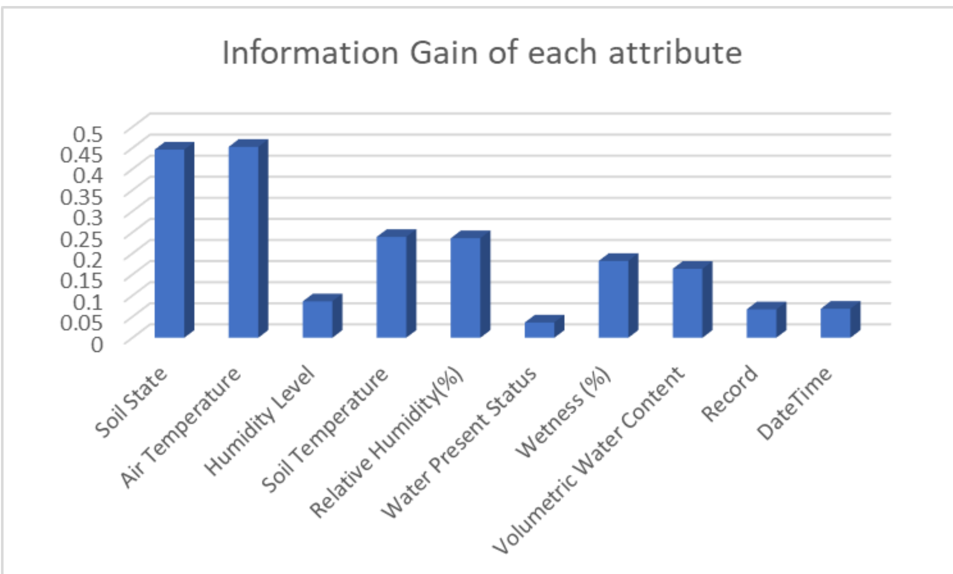

Figure 1: Information Gain for each attribute

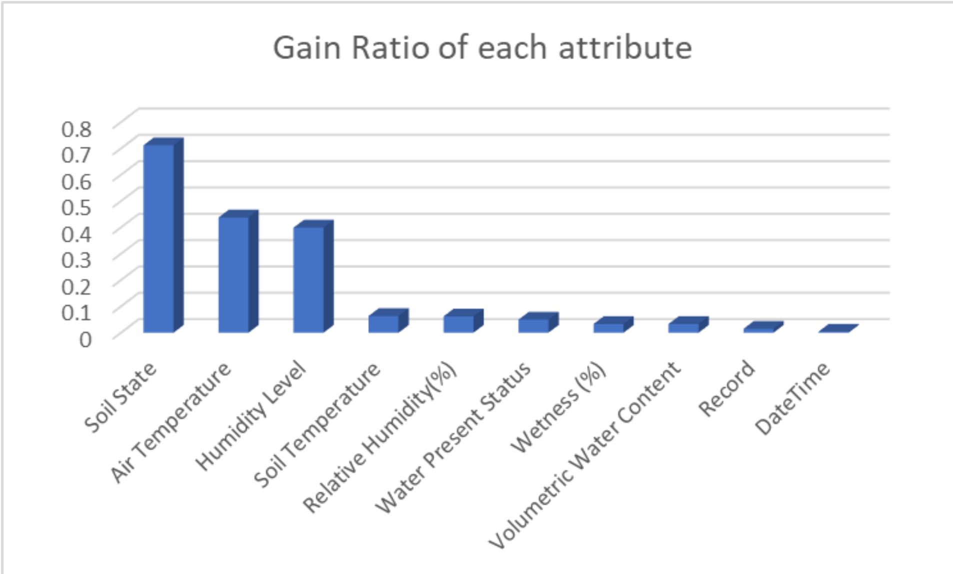

Figure 2: Gain Ratio for each attribute

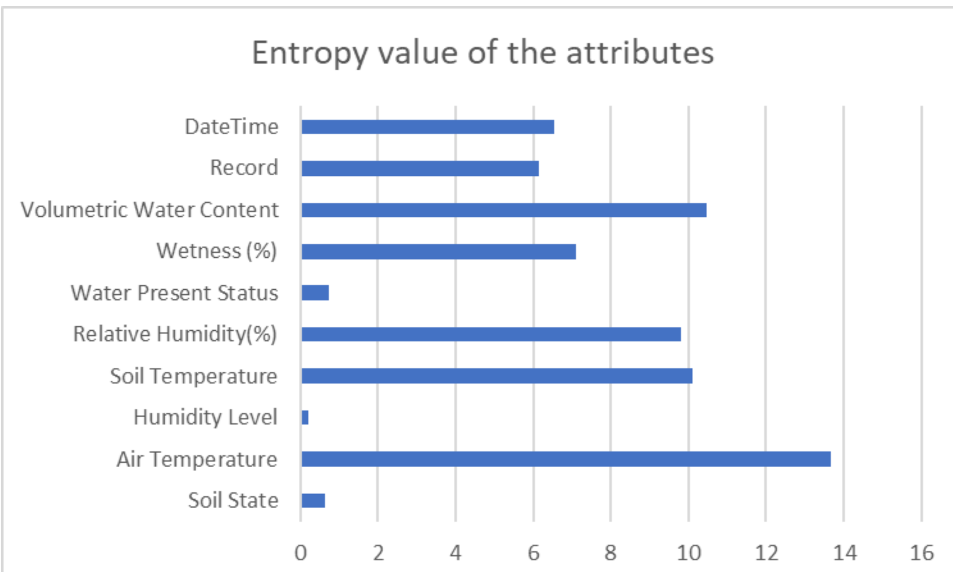

Figure 3: Entropy value for each attribute

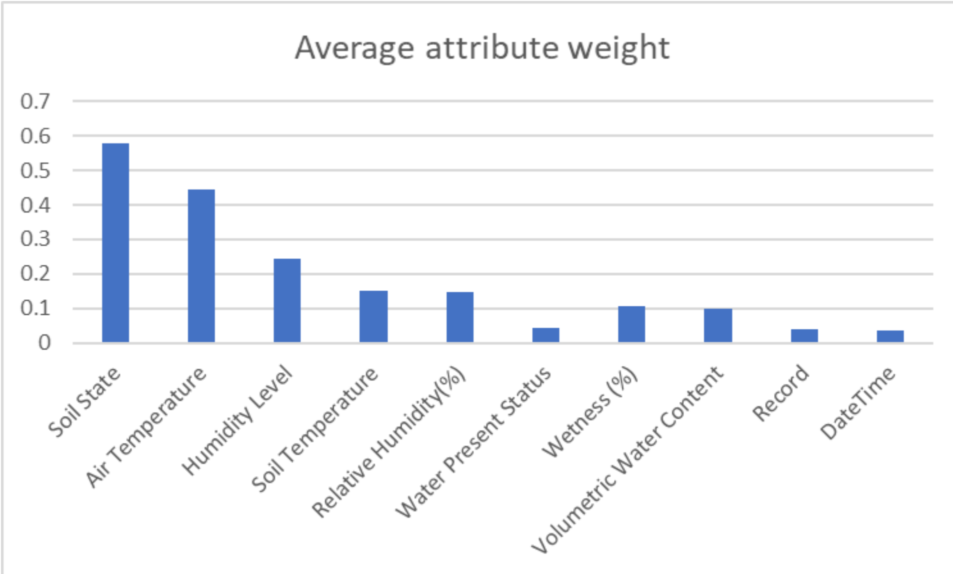

Figure 4: Average weight for each attribute

The entropy of an attribute helps determine what level of uncertainty of the attribute with respect to the attribute class. The lower the entropy, the higher the certainty of the attribute. With the results obtained, it is visible that soil state, humidity level and water present status have the lowest entropy and hence, their uncertainty is lower. Comparing with the result gotten from the information gain, the air temperature, soil state and soil temperature are the topmost attribute that provides the most information. The soil state, air temperature and humidity level have the highest gain ratio amidst other attributes. The gain ratio helps determine the ratio of the information gain to the inherent information. In evaluating the algorithms, the 5 attributes that have the most information were used in evaluating the algorithms. These attributes are; soil state, air temperature, humidity level, soil temperature and relative humidity.

Following the selected attributes, the datasets were trained and tested with the ID3, C5.0 and CART algorithms. The datasets were shuffled so as to reduce the possibility of class imbalance that could arise while splitting the dataset to test and train data. Dataset was further cleaned by dropping attributes that had no significant influence on decision making. The datasets were split into 80% train dataset which serves as the dataset that will be used to train the model and 20% test dataset to test the model in making a prediction. The trees gotten were further pruned to avoid overfitting. Figure 5 shows the tree model built with the ID3 algorithm while Figure 6 shows the decision tree drawn by the C5.0 algorithm and Figure 7 shows the tree gotten from the CART model. A true value indicates irrigation while false value indicates no irrigation.

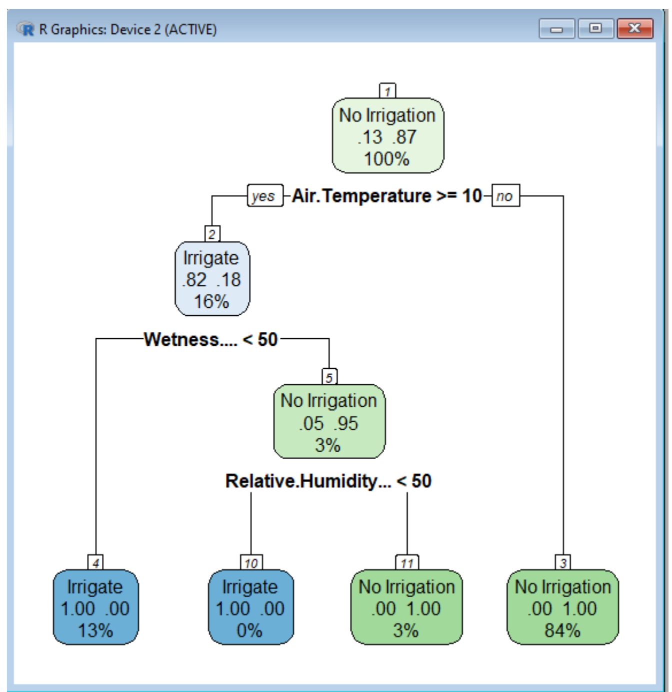

Figure 5: Average weight for each attribute

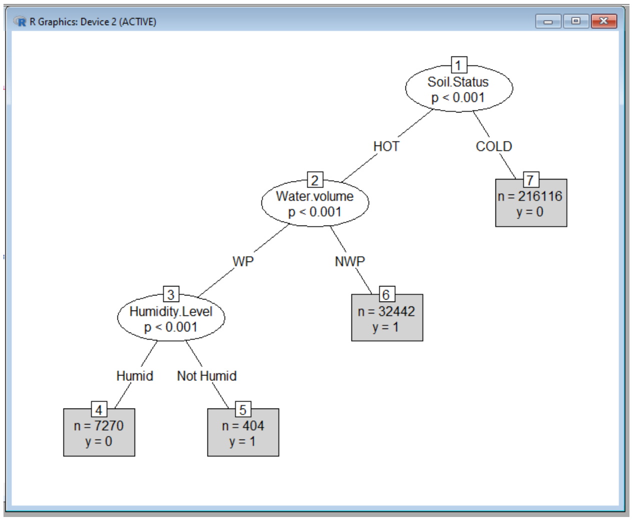

Figure 6: Decision Tree derived from C5.0 model.

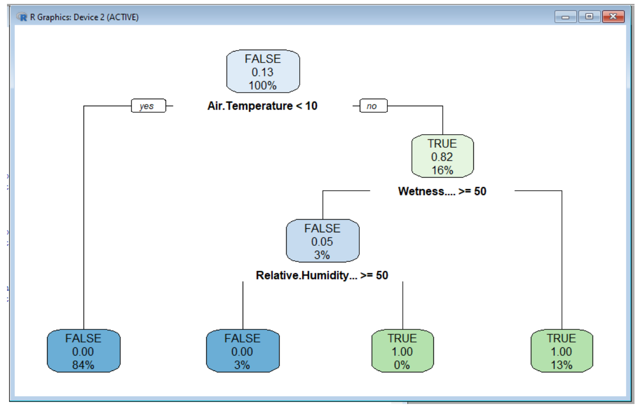

Figure 7: Decision Tree derived from the CART model

The data set used had 320,164 instances and using the split percentage (80% for training and 20% for evaluating the model), 256,131 was used as test data while 64,033 were used as the test data. The results obtained were evaluated and analyzed in order to choose the optimal algorithm. Table 2 shows the evaluation results.

Table 2: Results of the Evaluated Algorithms

Algorithm	**Accuracy**	**Precision**	**Recall**	**F-measure**
CART	0.9957	0.9783	0.9884	0.9833
C5.0	0.9951	0.9746	0.9872	0.9809
ID3	0.9931	0.9614	0.9851	0.9731

The results obtained showed that the CART algorithm records an increase of 0.06% compared to the C5.0 and 0.26% compared with the ID3 in terms of accuracy. The CART also records an increase of 0.38% against the C5.0 and 1.76% as against the ID3 with respect to the precision. Evaluating the recall, the CART outperformed the C5.0 and the ID3 with 0.12% and 0.33% respectively. The F-measure of the CART also recorded highest outperforming the C5.0 and ID3 with 0.24% and 1.05% respectively. In light of this, the CART outperformed the C5.0 and ID3 and hence, it is used as the decision tree algorithm in the decision making module. This is clearly shown in Figure 8

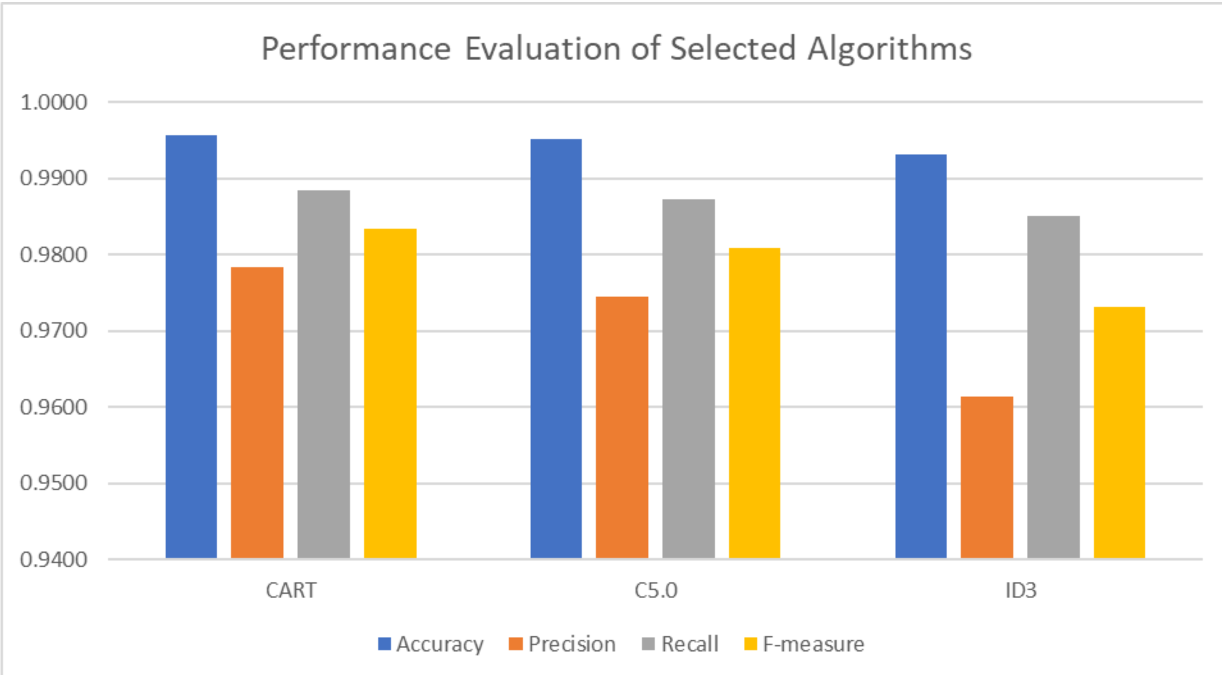

Figure 8: Performance evaluation of the selected algorithm.

The confusion matrix used for the evaluation of the results is presented in Table 3 while the confusion matrix for each algorithm is presented in Tables 4,5 and 6

Table 3: Confusion Matix Definition

		Predicted Value	
		Positive	Negative
Actual Value	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Table 4: Confusion Matrix for the CART algorithm

CART		Predicted Value	
		irrigate	No Irrigation
Actual Value	irrigate	8082	95
	No Irrigation	179	55677

Table 5: Confusion Matrix for the C5.0 algorithm

C5.0		Predicted Value	
		irrigate	No Irrigation

Actual Value	irrigate	8051	104
	No Irrigation	210	55668

Table 6: Confusion Matrix for the ID3 algorithm

ID3		Predicted Value	
		irrigate	No Irrigation
Actual Value	irrigate	7942	120
	No Irrigation	319	55652

5. Conclusions

From the outcome of the study, it was discovered that temperature state of the soil (hot, cold or warm), air temperature and humidity level are key determinant in making irrigation decisions. Also, the CART algorithm outperforms the remaining decision tree algorithm evaluated and this algorithm can be utilized in developing the decision module of AISs.

References

[1] R. Leclerc, "How to Invest In Agritech," Forbes. 2016.

[2] J. Lakshmi and G. Naresh, "A review on developing tech-agriculture using deep learning methods by applying UAVs," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 3, no. 1, pp. 1858–1863, 2018.

[3] S. Campbell and A. Tully, "Artificial intelligence helps supply chains minimize waste in food and medicine," SAP news center, 2018. [Online]. Available: https://news.sap.com/2018/05/artificial-intelligence-helps-supply-chains-minimize-waste-food-medicine/.

[4] M. Irimia, "Five ways agriculture could benefit from artificial intelligence," IBM Watson, 2016. [Online]. Available: https://www.ibm.com/blogs/watson/2016/12/five-ways-agriculture-benefit-artificial-intelligence/.

[5] P. Taechatanasat and L. Armstrong, "Decision Support System Data for Farmer Decision Making," in Proceedings of Asian Federation for Information Technology in Agriculture, 2014, pp. 472–486.

[6] C. Lokhorst, R. De-Mol, and C. Kamphuis, "Invited review: Big Data in precision dairy farming," Anim. Consort., vol. 14, pp. 1–10, 2019.

[7] N. Karayiannis and A. Venetsanopoulos, "Decision making using neural networks," Neurocomputing, vol. 6, no. 3, pp. 363–374.

[8] N. Zhifang, L. Phillips, and G. Hanna, "The use of bayesian networks in decision making," Key Top. Surg. Res. Methodol., vol. 1, no. 9, pp. 1–10, 2010.

[9] O. Awodele, O. Kalesanwo, F. Osisanwo, and S. Kuyoro, "Performance Evaluation of Some Mobile Adhoc Network Routing Protocols," Int. J. Sci. Eng. Res., vol. 8, no. 4, pp. 766–770, 2017.

[10] Y. Duan, J. Edwards, and Y. Dwivedi, "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda," Int. J. Inf. Manage., vol. 48, pp. 63–71, 2019.

[11] B. Keswani et al., "Adapting weather conditions based IoT enabled smart irrigation technique in precision agriculture mechanisms," Neural Comput. Appl., vol. 31, pp. 277–292, 2019.

[12] A. Mohapatra and S. Lenka, "Hybrid decision support system using PLSR-fuzzy model for GSM-based site-specific irrigation notification and control in precision agriculture," Int. J. Intell. Syst. Technol. Appl., vol. 15, no. 1, pp. 4–18, 2016.

[13] A. Mohapatra, S. Lenka, and B. Keswani, "Neural network and fuzzy logic based smart DSS model for irrigation notification and control in precision agriculture," in Proceedings of the National Academy of Sciences India, 2019, vol. 89, no. 1, pp. 67–76.

[14] M. Maharajan, T. Abirami, and S. Anitha, "Energy efficient wireless sensor network for precision agriculture," Adv. Eng. Res., vol. 142, no. 18, pp. 131–136, 2018.

[15] P. Ippolito, "Feature extraction techniques," Towards Data Science, 2019. [Online]. Available: towardsdatascience.com/feature-extraction-techniques-d619b56e31be.

[16] R. O'Donnell, "Information Theory," Carnegie Mellon University, 2013.

[17] A. Sarah, "Basic evaluation measures from the confusion matrix," class eval, 2016. [Online]. Available: https://classeval.wordpress.com/introduction/basic-evaluation-measures/.

[18] K. Shung, "Accuracy, Precision, Recall or F1?," Towards Data Science, 2018. [Online]. Available: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9.

[19] J. Brownlee, "Feature selection with the Caret R package," Machine Learning Mastery, 2014. [Online]. Available: https://machinelearningmastery.com/feature-selection-with-the-caret-r-package/.