© JAN 2026 | IRE Journals | Volume 9 Issue 7 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1713470

Improved Predictive Model for Crop Recommendation and Crop Yield in Rural Areas using K-Nearest Neighbour (KNN) and Decision Tree (DT) Machine Learning Classification Algorithms

JOSHUA CHINEMEREM CHIBUEZE[1], AGBAKWURU ALPHONSUS ONYEKACHI[2]
[1]*Department of Computer Engineering, Ogbonnaya Onu Polytechnic, Aba Abia State, Nigeria*
[2]*Department of Computer Science Imo State University of Owerri, Imo State, Nigeria*

Abstract - The aim of this paper is to develop an improved predictive model for crop recommendation and crop yield in rural areas using K-Nearest Neighbour (KNN) and Decision Tree (DT) machine learning classification algorithms, to improve crop growth and yield through the analysis of the soil ph value as the target towards effective prediction and recommendation of crops using variables such as: crop types, rainfall, soil humidity and temperature are the major objectives of this paper. The study employed two machine learning classification algorithms namely: K-Nearest Neighbour (KNN) and Decision Tree (DT). The data was analyzed with JASP machine learning platform while the experiments are done using a dataset containing 2200 data's sourced from Kaggle machine learning repository and was named as Hybrid_Agro_Crop_Recomender. During the experiment on the 2200 data's, as contained in the dataset, 20% was split for test while 80% was split for train making up to a total of 100% after preprocessing was concluded. The result of the experiment showed a very high performance on both applied algorithms where (DT = 95% and KNN =94%). Decision Tree (DT) and K-Nearest Neighbors (KNN) produced F1 Score result for the following variables (apple 100%, banana 100%, blackgram 83%, chickpea 100%, coconut 97%, coffee 97%, cotton 97%, grapes 100%, jute 89% and kidneybeans 100%) accuracy for perfect soil compatibility with high percent nutrient to grow such recommended crops and on the result as produced by ROC curve, the accuracy shows that apple, banana and blackgram has a predictive accuracy from 80% to 100% while crops like maize, mango, mothbeans, mungbeans, mushmelon and orange has a predicted accuracy between 81% to 96%. From the result performance on both algorithms, the experiment shows that the use of a hybrid approach involving two or more classification algorithms in object classification is very much essential for effective decision making. Therefore the developed model was called Hybrid_Agro_Crop_Recomender as its application in decision making produced an excellent outcome to improve crop produce and health of the applied crops.

Subject Area
Artificial Intelligence, Intelligent Agriculture, Machine Learning, Decision Tree

Keywords: Artificial Intelligence, Machine Learning, Agriculture, Classification Model algorithms, and Crop recommendation model, Hybrid_Agro_Crop_Recomender, Crop Recommender Model

I. INTRODUCTION

The agriculture sector has witnessed numerous changes for improving crop production. Several standards have been set to promote agricultural businesses, helping farmers improve their operational efficiency, reduce cost, provide quality food, and ensure their food hygiene and safety. Soil productivity closely depends on the available nutrients that result in a good yield of crops depending the soil. The availability of nutrients in the soil is monitored using a specific system to determine the fertility of that specific area. An analysis is done to decide on soil fertility and recommendations to strengthen crop growth. Due to the adoption of synthetic or chemical-based fertilizers by most farmers in the twentieth century, there had been a 50% increase in the overall yield from the field. Still, it has led to the major issue of Soil infertility or unavailability of major natural elements [1]. The climatic effects and environmental conditions should not degrade the yield. Farmers require data-driven or service-based techniques to enhance crop yield with the available field and other resources to meet all these needs. In this regard, precision agriculture has evolved with several tools and techniques that are being formulated, including automated harvesters, robot-weeders, Smartphone-based monitoring, UAVs, computer vision, pervasive computing, wireless ad-hoc sensor networks, Radio Frequency Identifier (RFID), cloud computing based data storage, Machine Learning models, IoT based devices combined with Deep Learning, satellite monitoring, remote sensing, context-aware computing, etc., which are becoming increasingly popular and beneficial to the farmers for monitoring the crop stress which limits the output. With regard to precision agriculture, many areas of scope or use case models shall be explored. They are as follows: Crop health monitoring for deficiencies and diseases, Soil Nutrient management, Monitoring of climate conditions, Farm land monitoring and mapping for predictive Analytics, Greenhouse automation, Automated irrigation scheduling and optimization, Production and yield management, Livestock monitoring, Farm Inventory management systems, Crop security and sorting, livestock monitoring, identification of diseases and nutrition deficiency in plants/crops.

Furthermore, reliable diagnosis of the nutritional status of crops is an essential part of the management of a farm, since both excess and deficiency of nutrients can cause severe damage and yield loss. Accurate determination of the nutritional status can not only prevent those losses, but also serve as basis for the rational use of nutritional supplements, as preconized by precision agriculture principles. As a result, waste of financial resources is avoided and environmental impacts are reduced. Moreover, computational tools for nutrition monitoring can be made available as part of decision support and farm management tools, which can be particularly valuable for farmers that do not have access to expert advice. Currently, the most common way to determine the nutritional status is visually, by means of plant color guides that do not allow quantitatively rigorous assessments [2]. More accurate evaluations require laboratorial leaf analyses, which are time consuming and require the application of specific methods for correct interpretation of the data [3], [4]. This paper aims to develop an improved predictive model for crop recommendation and crop yield in rural areas using K-Nearest Neighbour (KNN) and Decision Tree (DT) machine learning classification algorithms. There are other arias of precision agriculture as stated above which this paper did not cover rather concentrated on the predictive crop recommendation base on soil ph value to improve crop yield and ensure that farmers produce are improved by the recommendation model produced by the application of the two machine learning algorithms for better result classification of the crops and soil type. In other to plant a healthy plant, there is need to know each soil that best grow a particular type of crop, which motivated this study to ensure that the right crop are planted on the correct soil with adequate nutritional contents to support the crop. These model will help improve agricultural activity especially crops of different types and motivate famers to put more energy in farming which will in turn improve agricultural produce.

II. RELATED LITERATURE REVIEW

Nutrient Deficiency Detection in Leaves using Deep Learning indicates that Nutrient deficiency is major problem in the agricultural field and solutions available for this problem are not sufficient. Concept of Neural Network and Deep Learning can be used to solve such problem more efficiently. The nutrient deficiency detector is used for tracking the health of leaves [5]. A data set was created to check the different features of leaves through deep learning modules and techniques. The user will have to input a leaf image. This image passes through various neural networks in order to look for the different deficiency features present in the leaf so as to determine the type of deficiency. The neural network will classify the image into its respective deficiency class. Once the state of the leaf was identified by the model, it let the user know about the nutrient shortage in the plant. According to [6] researched on a study called Plant disease detection using machine learning approaches, stating that Plant health care is the science of anticipating and diagnosing the advent of life-threatening diseases in plants. The fatality rate of plants can be reduced by diagnosing them for any signs early on. The early detection of such diseases is one possibility for lowering plant mortality rates. Machine learning (ML), a type of artificial intelligence technology that allows researchers to enhance and develop without being explicitly programmed, is used in this study to build early prediction models for plant disease diagnosis. Due to the similarities of crops throughout the early phonological phases, crop classification has proved problematic. ML can be applied to a variety of tasks recognize different types of crops at low altitude platforms with the help of drones that provide high-resolution optical imagery. The drones are employed to photograph phonological stages, and these greyscale photographs are then utilized to develop grey level co-occurrence matrices-based characteristics. In this article, the proposed plant disease detection models are developed using ML approaches such as random forest-nearest neighbours, linear regression, Naive Bayes, neural networks, and support vector machine. The performance of the generated plants disease risk evaluation model is calculated using unbiased metrics such as true positive rate, true negative rate, precision, recall, and $F1$-score method are all factors to consider. The results revealed that the ensemble plants disease model outperforms the other proposed and developed plant disease detection models [7].

[8] made a review on Plant nutritional deficiency detection: a survey of predictive analytics approaches, stating that detecting plant nutritional deficiencies is crucial in agriculture, as these deficiencies directly impact productivity, food security, and the environment. Conventional methods for assessing plant nutritional content, such as soil testing and leaf tissue analysis, are time-consuming and expensive. Over the past decade, researchers have been working on automating this process using predictive analytics. This study discusses the importance of essential nutrients in plants, their visual symptoms, and various methods for assessing nutritional deficiencies. This study categorizes current research into two categories based on the type of data used for prediction: visible spectral images and multispectral/hyperspectral images [9]. This classification offers valuable insight into the strengths and limitations of each, thereby shedding light on their potential applications. This research examined the challenges and possible solutions in automating the detection of nutritional deficiencies. It highlights the need for scalable and accessible solutions and emphasizes human–machine collaboration for precision and interpretability.

According to [10] made a comprehensive review on detection of plant disease using machine learning and deep learning approaches, Also agriculture plays a significant part in India due to their population growth and increased food demands. Hence, there is a need to enhance the yield of crop. One of these important effects on low crop yields is diseases caused by bacteria, fungi and viruses. Plant nutrition deficiency and disease classification using graph convolutional network, Agricultural production plays a crucial role in the sustainable economic and societal growth of a country. High-quality crop yield production is essential for satisfying global food demands and better health [11]. [12] carried a study on investigating plant Disease Prediction System Using Machine Learning Plant diseases. Their study used a dataset of plant images infected with various diseases, which will pre-process and classified using advanced algorithms like convolutional neural network and their study aimed at recognizing disease and pattern of leaves, steam, and fruits. A result of their study was able to demonstrate impressive testing accuracies of 96.63% respectively. As stated by [13] proposed for the measurement of disease severity of rice crop using machine learning and computational intelligence. The paper introduces Fuzzy Logic with K-Means segmentation technique to compute the degree of disease severity of leaves in rice crop. Fuzzy system is used here because of its flexible nature and conceptually easy to understand according to the writer.

[14] found a model meant to transfer Learning, being able to achieve respectable accuracy scores within a short space of time and with limited computational power. The study addressed different aspects of Machine Learning and explained the principals behind the Convolutional Neural Network architecture. They were able to find a suitable architecture that allows image classification through Transfer Learning, this came in the form of Inception-v3. [15] proposed the semi-supervised few-shot learning scheme, which can improve the average accuracy of few-shot classification by adaptively selecting the pseudo-labeled samples to help fine-tune the model. Through literature research, they carried out the first semi-supervised work in the field of few-shot plant diseases classification. The PlantVillage dataset was divided into three split modes, and extensive comparison experiments were executed to prove the correctness and generalization of proposed methods. Considering all the different domain splits and k-shot, the average improvement by the proposed single semi-supervised method is 2.8%, and that by the iterative semi-supervised method is 4.6%. [16] designed and developed real-time decision support hardware for the identification of healthy leaf and diseased leaf and to analyze the performance of the adopted machine learning classifiers such as extreme learning machine and Support Vector

Machine with linear and polynomial kernels. In this work, a real-time decision support system integrated with a camera sensor module was designed and developed for identification of plant disease. Furthermore, the performance of three machine learning algorithms, such as Extreme Learning Machine (ELM) and Support Vector Machine (SVM) with linear and polynomial kernels was analyzed. Results demonstrate that the performance of the extreme learning machine is better when compared to the adopted support vector machine classifier. It is also observed that the sensitivity of the support vector machine with a polynomial kernel is better when compared to the other classifiers.

III. METHODOLOGY

The machine learning model development life-cycle was used to analyze the proposed System. This approach helped the researchers to adopt two machine learning classification algorithms namely: Decision Tree (DT) and K-Nearest Neighbor (KNN) algorithms because of their ability to uncover or translate hidden pattern from a model and also accuracy in data prediction for effective decision making and in handing sequential data. These two ML classification algorithms were chosen so that adequate data accuracy can be compared between them and to ensure unique comparison of results and also to use the produced model to predict and recommend crop yield base on different plants and soil ph value.

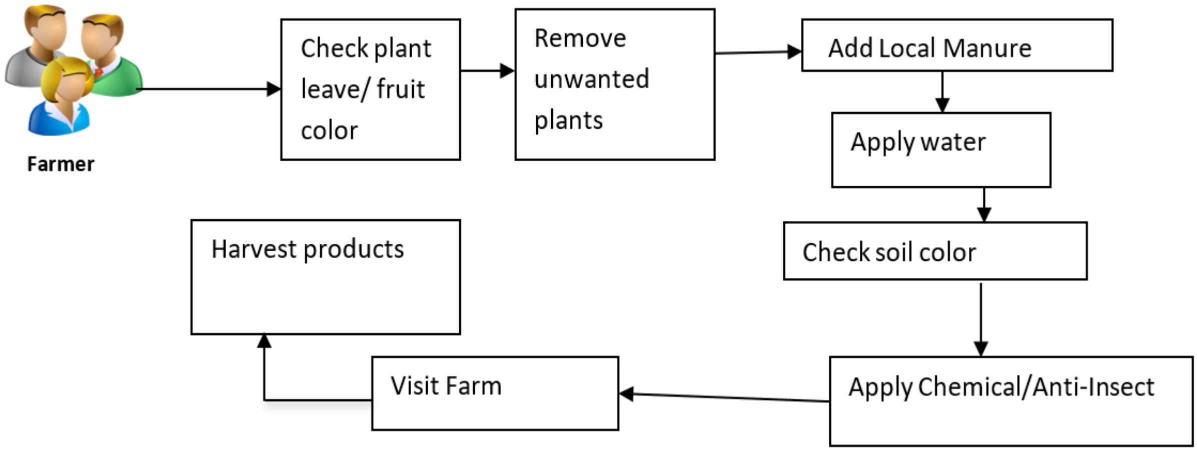


Figure 1: Analysis of the existing traditional approach to treat disease on plants

On the existing traditional approach to manage or treat plants, the famers first visit their farm to check the color of the plants or leave, it is from the coloring of the leaves or fruits that could inform the farmer if the plant is healthy or if the soil has adequate nutrients or not. The process also allow the famer to remove unwanted plants which takes the nutrients meant for the plants to grow well by removing them and by applying either manure or fertilizer for effective nutritional effects on the plants. After adding the required nutrients, the next is to ensure that there is adequate water for the plants to survive then lastly to harvest the products.

THE PROPOSED SYSTEM DIAGRAM

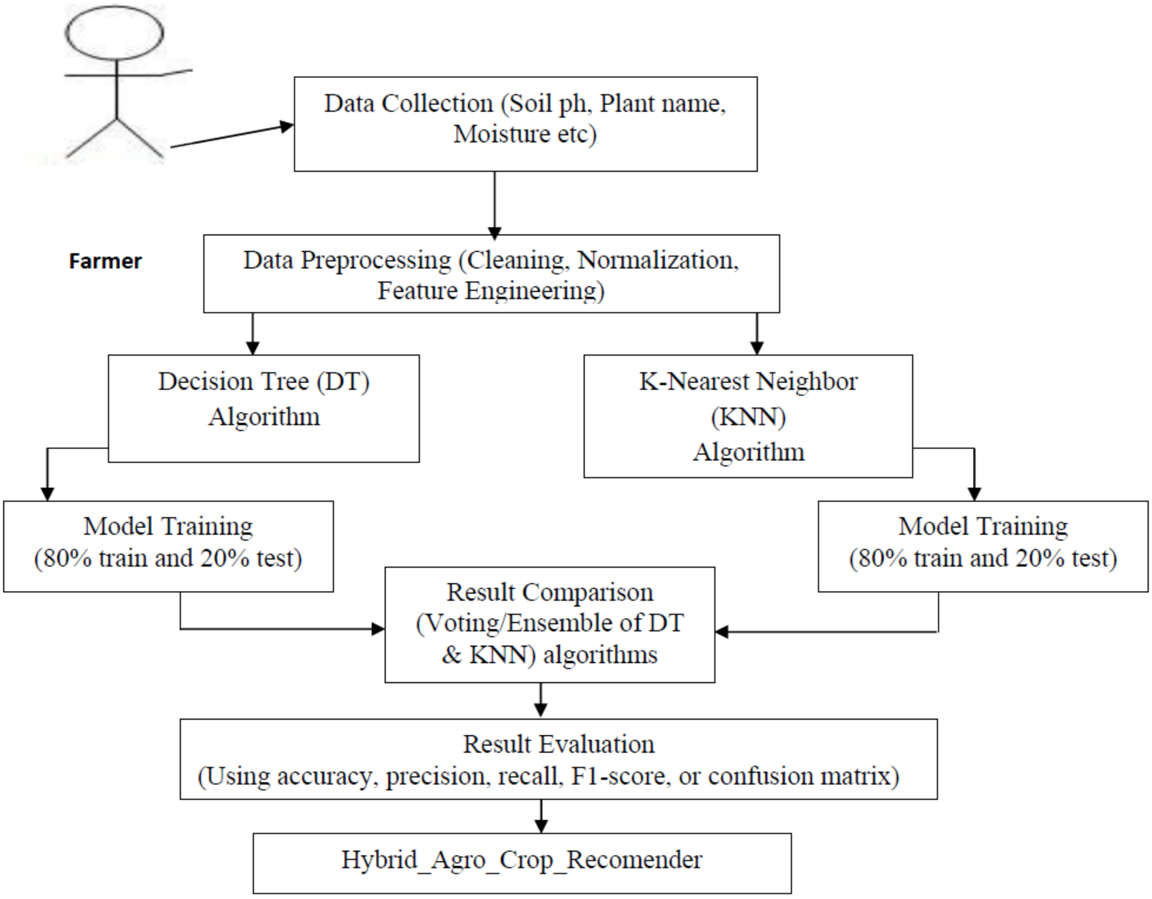

Data Collection (Soil ph, Plant name, Moisture etc)
Farmer
Data Preprocessing (Cleaning, Normalization, Feature Engineering)
Decision Tree (DT) Algorithm
K-Nearest Neighbor (KNN) Algorithm
Model Training (80% train and 20% test)
Model Training (80% train and 20% test)
Result Comparison (Voting/Ensemble of DT & KNN) algorithms
Result Evaluation (Using accuracy, precision, recall, F1-score, or confusion matrix)
Hybrid_Agro_Crop_Recomender

Figure 2: Diagram of the proposed model

Analysis of the proposed system as shown in figure 2 above allow a farmer to collect data from collection point sourced from Kaggle repository which contains (plant type, disease types, ph value, nutrients, soil ph value, humidity, soil type etc.). These variable/features are some of the contains used to build the dataset for the prediction of crop recommendation model. This datasets are collected and Data Preprocessing which includes (Cleaning, Normalization, Feature Engineering) was done on them to ensure a meaningful assessment, verification of variables in the dataset are achieved out for a more accurate experiment, then after which Decision Tree (DT) Algorithm and K-Nearest Neighbor (KNN) Algorithms was applied to train and test the performance of the developed model to detect disease per plant and recommend plants. Before the experiment was done on the two algorithms, a division of (80% train and 20% test) was made to ensure a more accurate evaluation. After the application of the two different algorithms, Model Evaluation was performed on DT and KNN by monitoring the Accuracy, precision, recall, F1-score, Roc-Curve or confusion matrix of the both models produced by the two different algorithms which now formed as the Final Prediction of the proposed model named Hybrid_Agro_Recommender.

Table 1: SYSTEM ALGORITHM

INPUT	Crop_recommendation from Kaggle machine learning repository 2200 dataset features
OUTPUT	Improved Predictive Model for Crop Recommendation and Crop Yield in Rural Areas using K-Nearest Neighbour (KNN) and Decision Tree (DT) Machine Learning Classification Algorithms (Hybrid_Agro_Crop_Recomender)

IV. RESULTS

EXPERIMENTS ON THE DATASET USING JASP ML PLATFORM

The first process was launching of the JASP PLATFORM after a successive launching, the following steps were done to build the model.

Step 1: Loading the dataset from the location Crop_dataset (Crop_recommendation.csv)

Step 2: Select machine learning packages

Step 3: Select first classification algorithm (Decision Tree)

Step 4: Set the target and features (Class: soil ph)

Step 5: Click to start the analysis on the dataset

Step 6: Click on Data split

Step 7: Click Confusion Matrix

Step 8: Class Proportions

Step 9: Click Evaluation Metrics

Step 10: Click ROC Curve Plot

Step 11: click Andrews Plot

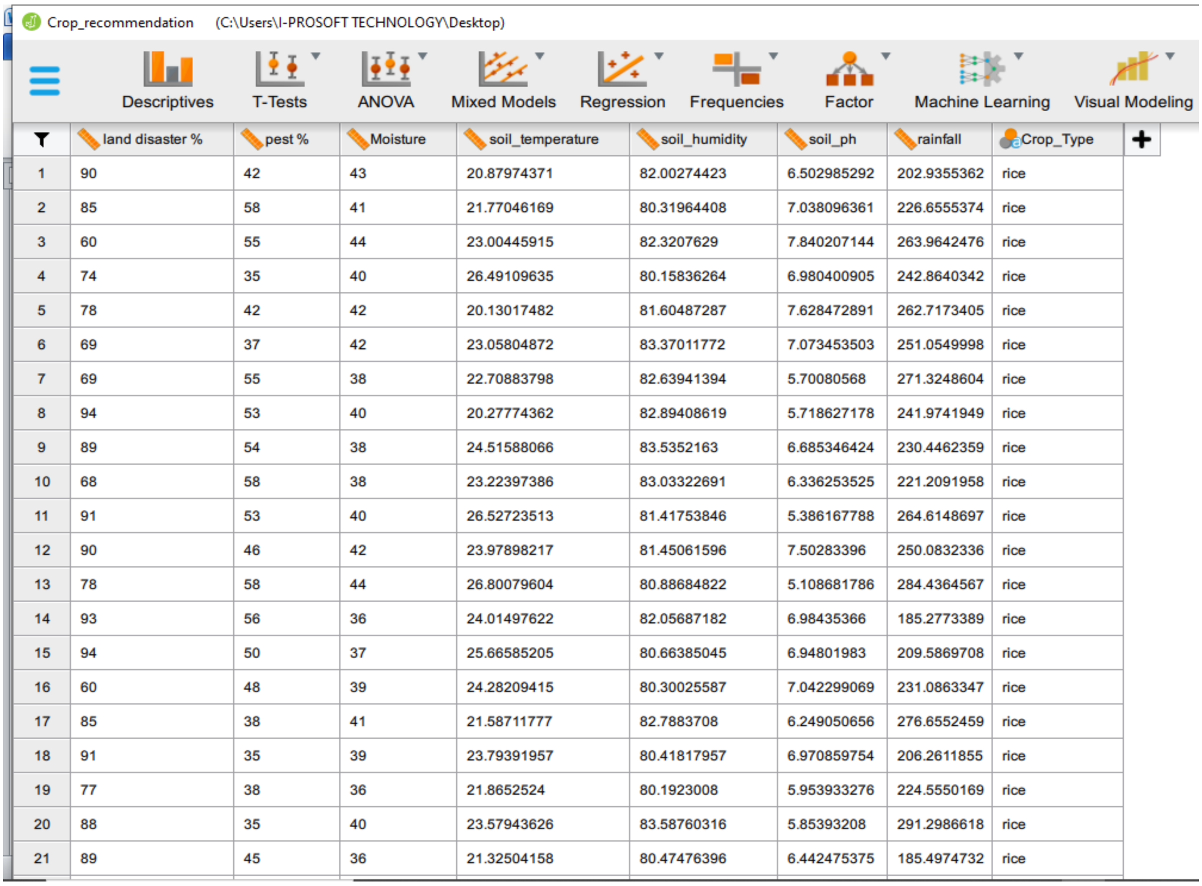
Step 12: click decision tree plot

Download visualizations results

Step 13: Start prediction accuracy by checking (F1 score, confusion matric, and Roc Curve and precision (positive predictive value) results).

Step 3: was carried out again to use K-Nearest Neighbor (K-NN) algorithms on the same dataset in other to complete the hybrid use of the two classification algorithms and it is done following other steps below for a more accurate prediction of the model produced which are predicted by looking at the output F1 score, ROC curve, confusion matric and decision tree models built.

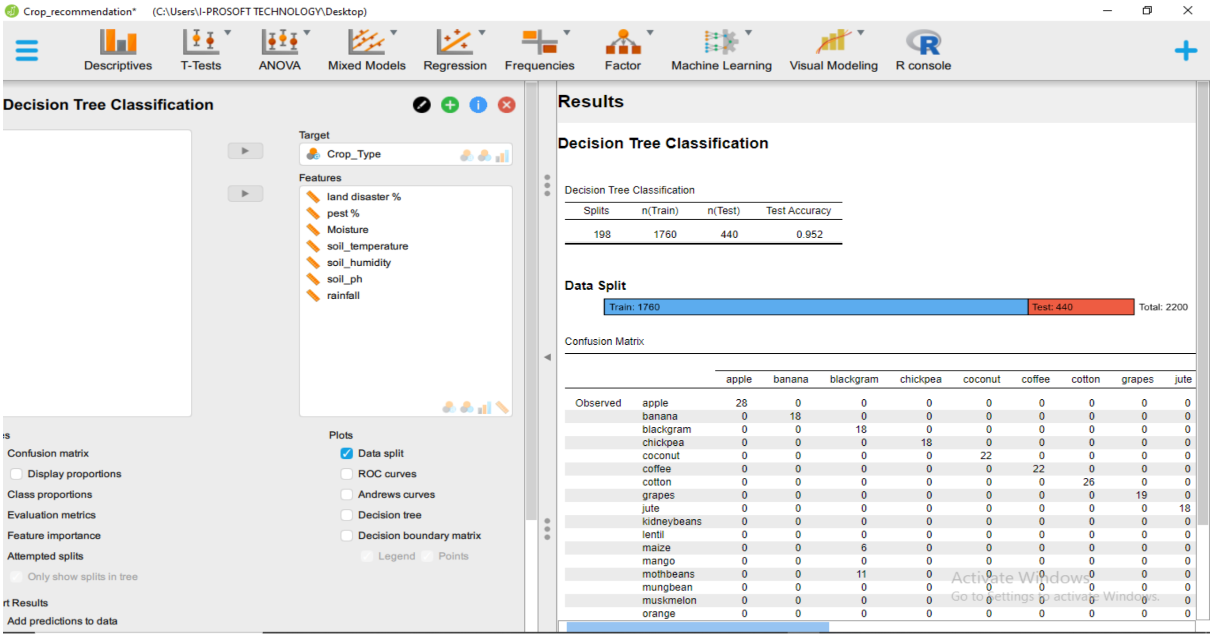
EXPERIMENT OUTPUT



	land disaster %	pest %	Moisture	soil_temperature	soil_humidity	soil_ph	rainfall	Crop_Type
1	90	42	43	20.87974371	82.00274423	6.502985292	202.9355362	rice
2	85	58	41	21.77046169	80.31964408	7.038096361	226.6555374	rice
3	60	55	44	23.00445915	82.3207629	7.840207144	263.9642476	rice
4	74	35	40	26.49109635	80.15836264	6.980400905	242.8640342	rice
5	78	42	42	20.13017482	81.60487287	7.628472891	262.7173405	rice
6	69	37	42	23.05804872	83.37011772	7.073453503	251.0549998	rice
7	69	55	38	22.70883798	82.63941394	5.70080568	271.3248604	rice
8	94	53	40	20.27774362	82.89408619	5.718627178	241.9741949	rice
9	89	54	38	24.51588066	83.5352163	6.685346424	230.4462359	rice
10	68	58	38	23.22397386	83.03322691	6.336253525	221.2091958	rice
11	91	53	40	26.52723513	81.41753846	5.386167788	264.6148697	rice
12	90	46	42	23.97898217	81.45061596	7.50283396	250.0832336	rice
13	78	58	44	26.80079604	80.88684822	5.108681786	284.4364567	rice
14	93	56	36	24.01497622	82.05687182	6.98435366	185.2773389	rice
15	94	50	37	25.66585205	80.66385045	6.94801983	209.5869708	rice
16	60	48	39	24.28209415	80.30025587	7.042299069	231.0863347	rice
17	85	38	41	21.58711777	82.7883708	6.249050656	276.6552459	rice
18	91	35	39	23.79391957	80.41817957	6.970859754	206.2611855	rice
19	77	38	36	21.8652524	80.1923008	5.953933276	224.5550169	rice
20	88	35	40	23.57943626	83.58760316	5.85393208	291.2986618	rice
21	89	45	36	21.32504158	80.47476396	6.442475375	185.4974732	rice

Figure 3: JASP View on the first 20 Crop Recommender dataset

Experiment on dataset using Decision Tree Classification



Decision Tree Classification

Target
Crop_Type

Features
land disaster %
pest %
Moisture
soil_temperature
soil_humidity
soil_ph
rainfall

s
Confusion matrix
□ Display proportions
Class proportions
Evaluation metrics
Feature importance
Attempted splits
☑ Only show splits in tree

rt Results
Add predictions to data

Plots
☑ Data split
□ ROC curves
□ Andrews curves
□ Decision tree
□ Decision boundary matrix
☑ Legend ☑ Points

Results

Decision Tree Classification

Decision Tree Classification

Splits	n(Train)	n(Test)	Test Accuracy
198	1760	440	0.952

Data Split

Train: 1760 Test: 440 Total: 2200

Confusion Matrix

		apple	banana	blackgram	chickpea	coconut	coffee	cotton	grapes	jute
Observed	apple	28	0	0	0	0	0	0	0	0
	banana	0	18	0	0	0	0	0	0	0
	blackgram	0	0	18	0	0	0	0	0	0
	chickpea	0	0	0	18	0	0	0	0	0
	coconut	0	0	0	0	22	0	0	0	0
	coffee	0	0	0	0	0	22	0	0	0
	cotton	0	0	0	0	0	0	26	0	0
	grapes	0	0	0	0	0	0	0	19	0
	jute	0	0	0	0	0	0	0	0	18
	kidneybeans	0	0	0	0	0	0	0	0	0
	lentil	0	0	0	0	0	0	0	0	0
	maize	0	0	6	0	0	0	0	0	0
	mango	0	0	0	0	0	0	0	0	0
	mothbeans	0	0	11	0	0	0	0	0	0
	mungbean	0	0	0	0	0	0	0	0	0
	muskmelon	0	0	0	0	0	0	0	0	0
	orange	0	0	0	0	0	0	0	0	0

Figure 4: JASP ML-Platform View on the prediction of dataset using Decision Tree Classification Algorithm

Table 2: Decision Tree (DT) Classification

Splits	n(Train)	n(Test)	Test Accuracy
198	1760	440	0.952

The accuracy result as predicted by DT was 0.952 which is 95% accuracy recommended for adequate use for crop recommendation and faster decision making for farmers.

Data Split

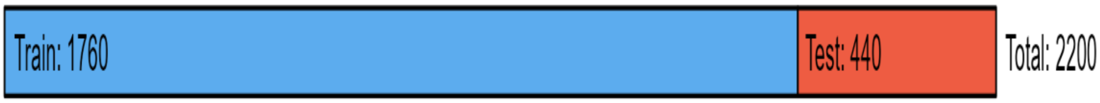
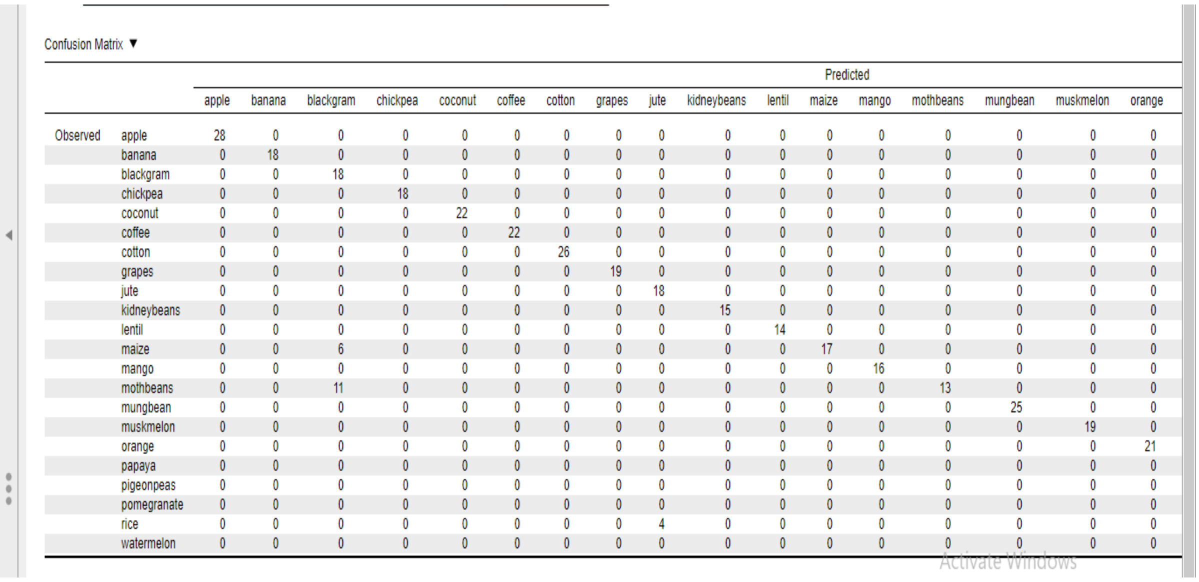


Figure 5: Dataset Split

Table 3: Confusion Matrix



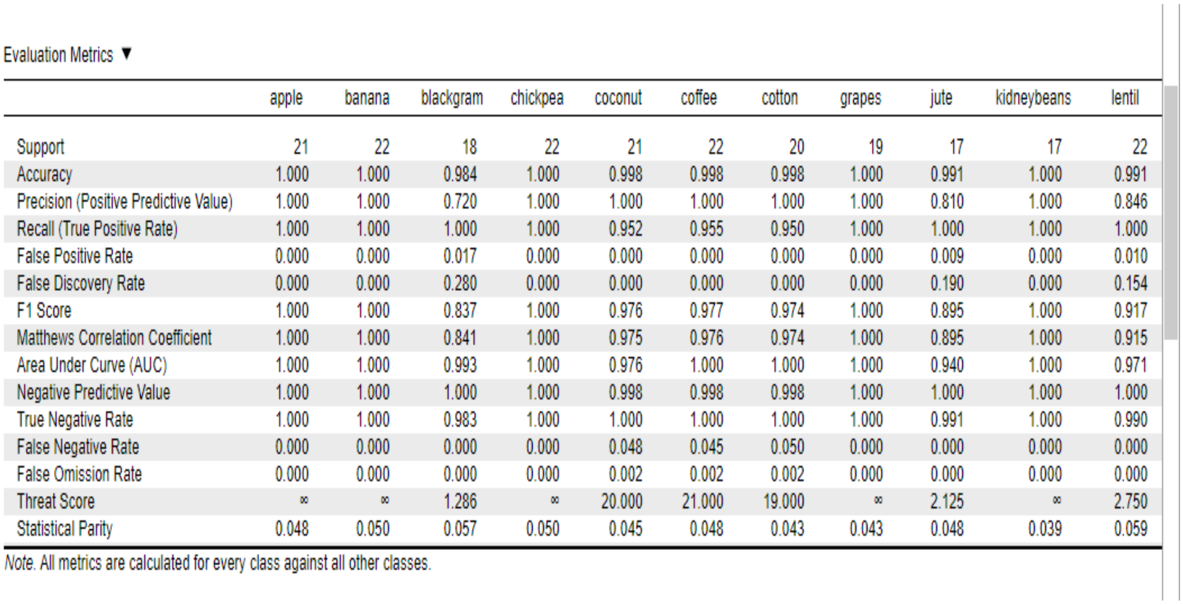
Confusion Matrix ▼

		Predicted																
		apple	banana	blackgram	chickpea	coconut	coffee	cotton	grapes	jute	kidneybeans	lentil	maize	mango	mothbeans	mungbean	muskmelon	orange
Observed	apple	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	banana	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	blackgram	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	chickpea	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0
	coconut	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0
	coffee	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0
	cotton	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0
	grapes	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0
	jute	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0
	kidneybeans	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0
	lentil	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0
	maize	0	0	6	0	0	0	0	0	0	0	0	17	0	0	0	0	0
	mango	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0
	mothbeans	0	0	11	0	0	0	0	0	0	0	0	0	0	13	0	0	0
	mungbean	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0
	muskmelon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0
	orange	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21
	papaya	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	pigeonpeas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	pomegranate	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	rice	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
	watermelon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 4: Class Proportions using DT algorithm

	Data Set	Training Set	Test Set
apple	0.045	0.045	0.048
banana	0.045	0.044	0.050
blackgram	0.045	0.047	0.041
chickpea	0.045	0.044	0.050
coconut	0.045	0.045	0.048
coffee	0.045	0.044	0.050
cotton	0.045	0.045	0.045
grapes	0.045	0.046	0.043
jute	0.045	0.047	0.039
kidneybeans	0.045	0.047	0.039
lentil	0.045	0.044	0.050
maize	0.045	0.043	0.055
mango	0.045	0.047	0.041
mothbeans	0.045	0.043	0.057
mungbean	0.045	0.046	0.043
muskmelon	0.045	0.046	0.043
orange	0.045	0.047	0.039
papaya	0.045	0.047	0.039
pigeonpeas	0.045	0.044	0.052
pomegranate	0.045	0.048	0.034
rice	0.045	0.045	0.045
watermelon	0.045	0.044	0.050

Table 5: Evaluation Matrix using DT algorithm



Evaluation Metrics ▼

	apple	banana	blackgram	chickpea	coconut	coffee	cotton	grapes	jute	kidneybeans	lentil
Support	21	22	18	22	21	22	20	19	17	17	22
Accuracy	1.000	1.000	0.984	1.000	0.998	0.998	0.998	1.000	0.991	1.000	0.991
Precision (Positive Predictive Value)	1.000	1.000	0.720	1.000	1.000	1.000	1.000	1.000	0.810	1.000	0.846
Recall (True Positive Rate)	1.000	1.000	1.000	1.000	0.952	0.955	0.950	1.000	1.000	1.000	1.000
False Positive Rate	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.010
False Discovery Rate	0.000	0.000	0.280	0.000	0.000	0.000	0.000	0.000	0.190	0.000	0.154
F1 Score	1.000	1.000	0.837	1.000	0.976	0.977	0.974	1.000	0.895	1.000	0.917
Matthews Correlation Coefficient	1.000	1.000	0.841	1.000	0.975	0.976	0.974	1.000	0.895	1.000	0.915
Area Under Curve (AUC)	1.000	1.000	0.993	1.000	0.976	1.000	1.000	1.000	0.940	1.000	0.971
Negative Predictive Value	1.000	1.000	1.000	1.000	0.998	0.998	0.998	1.000	1.000	1.000	1.000
True Negative Rate	1.000	1.000	0.983	1.000	1.000	1.000	1.000	1.000	0.991	1.000	0.990
False Negative Rate	0.000	0.000	0.000	0.000	0.048	0.045	0.050	0.000	0.000	0.000	0.000
False Omission Rate	0.000	0.000	0.000	0.000	0.002	0.002	0.002	0.000	0.000	0.000	0.000
Threat Score	∞	∞	1.286	∞	20.000	21.000	19.000	∞	2.125	∞	2.750
Statistical Parity	0.048	0.050	0.057	0.050	0.045	0.048	0.043	0.043	0.048	0.039	0.059

Note. All metrics are calculated for every class against all other classes.

ROC Curves Plot

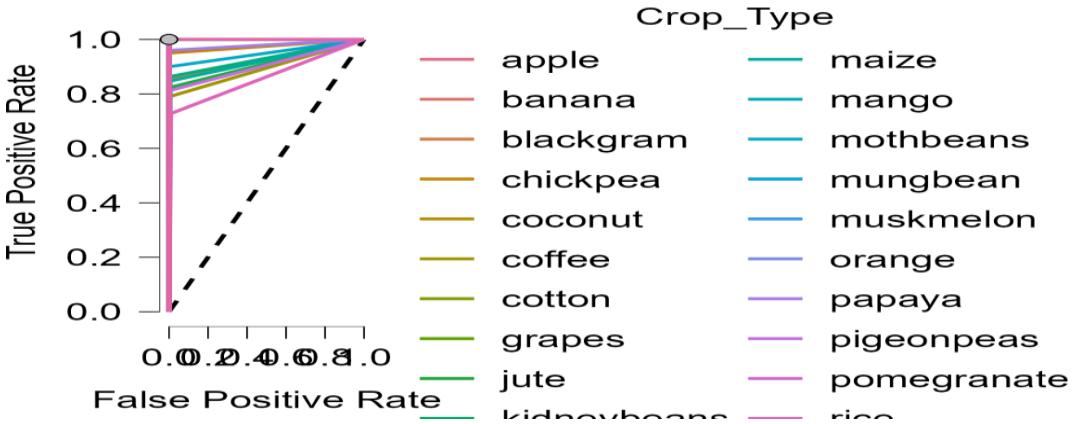


Figure 6: DT ROC Curve Plot

ROC curve plot shown in figure 6 above gave an analyzed chart of the DT algorithm on crop recommender with True Positive Rate against False Positive Rate. From the prediction given by ROC curve, the accuracy shows that apple, banana and blackgram has a predictive accuracy from 80% to 100% while crops like maize, mango, mothbeans, mungbeans, mushmelon and orange has a predicted accuracy between 81% to 96%. Each predicted a result on each crop/plant was based on the available datasets.

Andrews Curves Plot

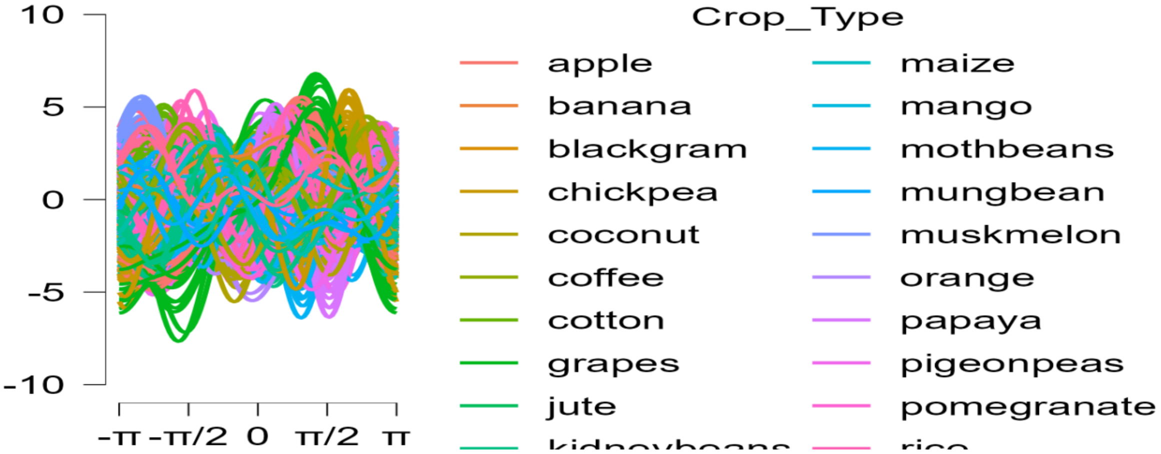


Figure 7: Andrews Curves Plot

This plot presented the class as analyzed by the DT algorithm with class of the different plant types. From the predicted results shown in figure 7 above, Andrews curve was able to predict 65% accuracy on apple crop, 52% accuracy on banana crop, 50% accuracy recommendation for maize, mango and mothbeans while -50% for grapes despite challenges of the soil type base on the nutrient deficiency and dataset collected for the prediction.

Decision Tree Plot Result

From the result shown in Figure 8 below using the DT classification algorithm presents the various variables/features organized in a tree structure showing various percentage rate of accuracy base on each crop/plant type and its soil and nutrient efficient before a farmer can venture into planting on the soil. This predicted results shows how decision tree algorithm uses its logical machine learning structure to predict the crops for each variables against number of other features and number of occurrence.

Decision Tree Plot

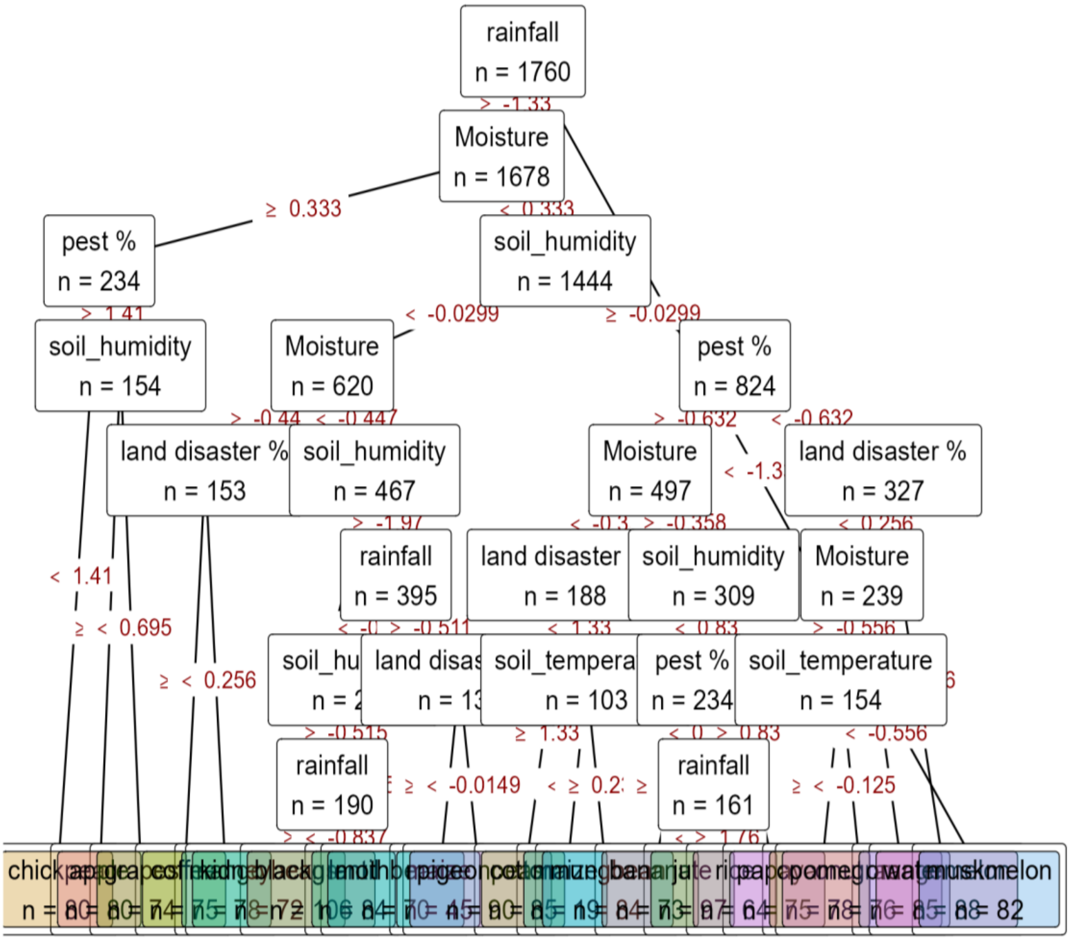
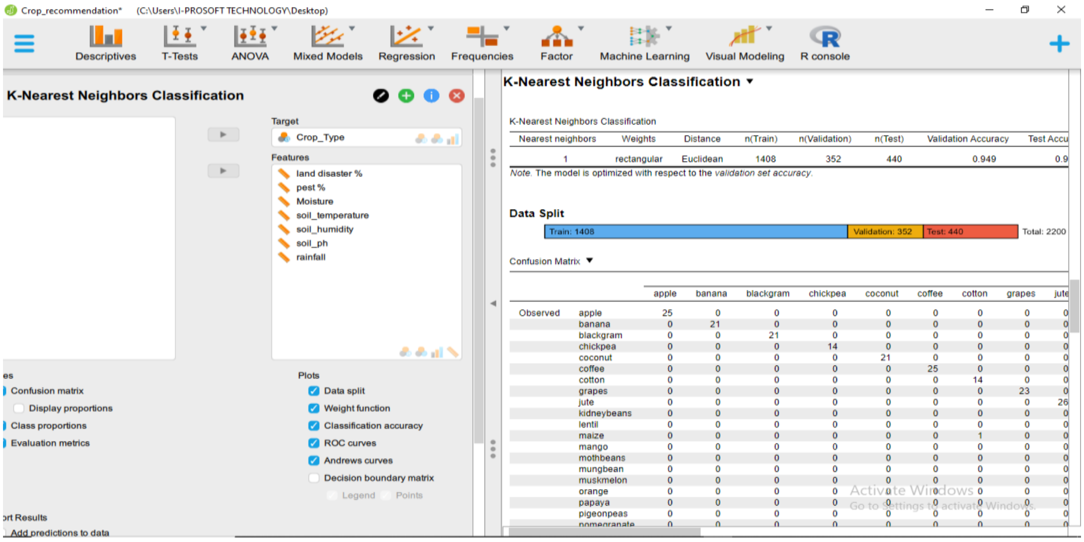


Figure 8: Decision Tree (DT) classification model

The major target is the Crop_Type, to recommend each crop that will grow or perform better in a particular soil. This is one of the best machine learning classification algorithms for classifying features for better placement or organization.

Experiment using K-Nearest Neighbor (K-NN) algorithms Classification



K-Nearest Neighbors Classification

Target
Crop_Type

Features
land disaster %
pest %
Moisture
soil_temperature
soil_humidity
soil_ph
rainfall

es
Confusion matrix
□ Display proportions
Class proportions
Evaluation metrics

Plots
☑ Data split
☑ Weight function
☑ Classification accuracy
☑ ROC curves
☑ Andrews curves
□ Decision boundary matrix
Legend Points

ort Results
Add predictions to data

K-Nearest Neighbors Classification ▼

K-Nearest Neighbors Classification

Nearest neighbors	Weights	Distance	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accu
1	rectangular	Euclidean	1408	352	440	0.949	0.9

Note. The model is optimized with respect to the *validation set accuracy*.

Data Split

Train: 1408 Validation: 352 Test: 440 Total: 2200

Confusion Matrix ▼

		apple	banana	blackgram	chickpea	coconut	coffee	cotton	grapes	jute
Observed	apple	25	0	0	0	0	0	0	0	0
	banana	0	21	0	0	0	0	0	0	0
	blackgram	0	0	21	0	0	0	0	0	0
	chickpea	0	0	0	14	0	0	0	0	0
	coconut	0	0	0	0	21	0	0	0	0
	coffee	0	0	0	0	0	25	0	0	0
	cotton	0	0	0	0	0	0	14	0	0
	grapes	0	0	0	0	0	0	0	23	0
	jute	0	0	0	0	0	0	0	0	26
	kidneybeans	0	0	0	0	0	0	0	0	0
	lentil	0	0	0	0	0	0	0	0	0
	maize	0	0	0	0	0	0	1	0	0
	mango	0	0	0	0	0	0	0	0	0
	mothbeans	0	0	0	0	0	0	0	0	0
	mungbean	0	0	0	0	0	0	0	0	0
	muskmelon	0	0	0	0	0	0	0	0	0
	orange	0	0	0	0	0	0	0	0	0
	papaya	0	0	0	0	0	0	0	0	0
	pigeonpeas	0	0	0	0	0	0	0	0	0
	pomegranate	0	0	0	0	0	0	0	0	0

Figure 9: JASP ML-Platform View on the prediction of dataset using K-Nearest Neighbor (KNN) Classification Algorithm

Data Split

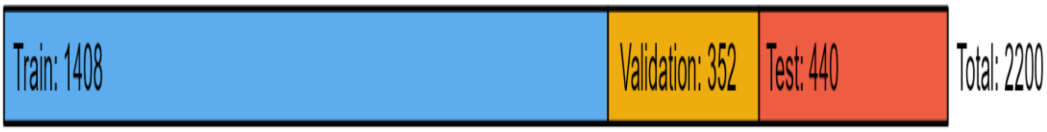


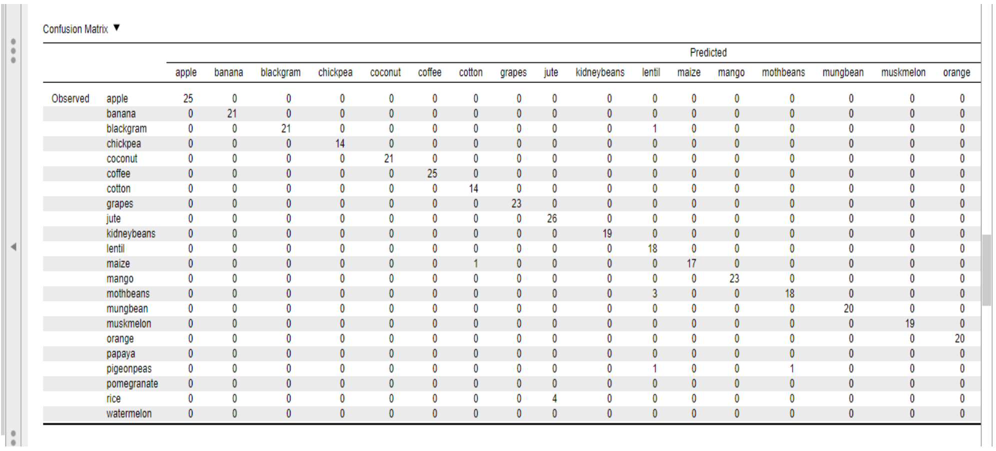
Figure 10: K-Nearest Neighbor (KNN) Data split

Table 6: K-Nearest Neighbors Classification

Nearest neighbors	Weights	Distance	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accuracy
1	rectangular	Euclidean	1408	352	440	0.949	0.970

Note. The model is optimized with respect to the *validation set accuracy*.

From the experiment conducted using KNN algorithm, table 6 above present the summary of the predicted accuracy between the Test and Train split on the dataset. KNN predicted an accuracy of 0.949 which is 95% of the validation set accuracy.



Confusion Matrix ▼

		Predicted																
		apple	banana	blackgram	chickpea	coconut	coffee	cotton	grapes	jute	kidneybeans	lentil	maize	mango	mothbeans	mungbean	muskmelon	orange
Observed	apple	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	banana	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	blackgram	0	0	21	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	chickpea	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0
	coconut	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0
	coffee	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0
	cotton	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0
	grapes	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0
	jute	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0
	kidneybeans	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0
	lentil	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0
	maize	0	0	0	0	0	0	1	0	0	0	0	17	0	0	0	0	0
	mango	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0
	mothbeans	0	0	0	0	0	0	0	0	0	0	3	0	0	18	0	0	0
	mungbean	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0
	muskmelon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0
	orange	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
	papaya	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	pigeonpeas	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
	pomegranate	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	rice	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
	watermelon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

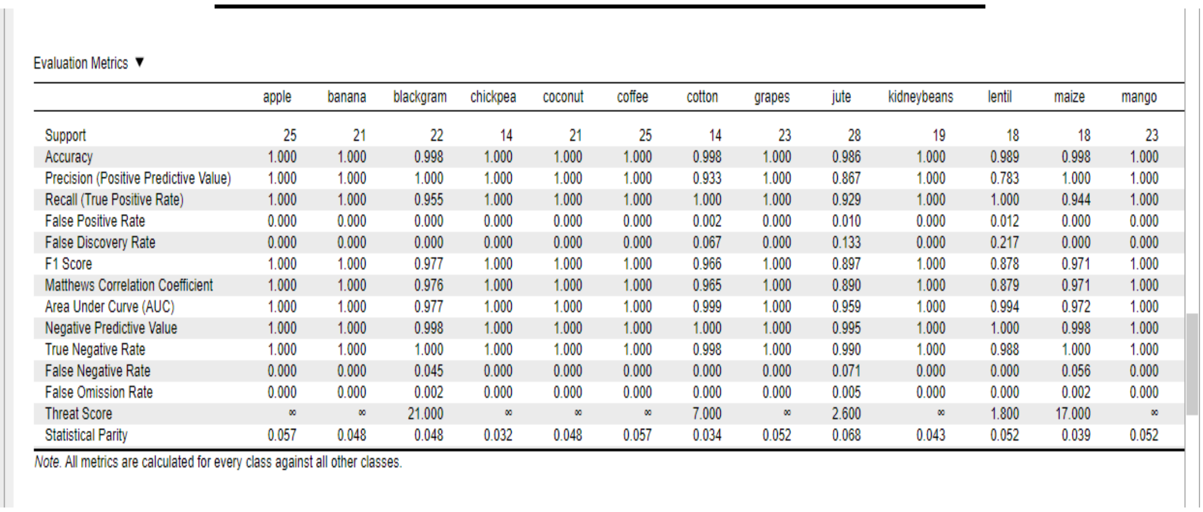
Figure 11: Confusion Matrix for KNN algorithm

Table 7: Class Proportions (KNN)

	Data Set	Training Set	Validation Set	Test Set
apple	0.045	0.048	0.023	0.057
banana	0.045	0.045	0.045	0.048
blackgram	0.045	0.050	0.023	0.050
chickpea	0.045	0.053	0.034	0.032
coconut	0.045	0.043	0.051	0.048
coffee	0.045	0.042	0.045	0.057
cotton	0.045	0.048	0.054	0.032
grapes	0.045	0.043	0.048	0.052
jute	0.045	0.042	0.037	0.064
kidneybeans	0.045	0.043	0.060	0.043
lentil	0.045	0.050	0.031	0.041
maize	0.045	0.041	0.068	0.041
mango	0.045	0.044	0.043	0.052
mothbeans	0.045	0.043	0.054	0.048
mungbean	0.045	0.048	0.037	0.045
muskmelon	0.045	0.044	0.054	0.043
orange	0.045	0.043	0.054	0.045
papaya	0.045	0.048	0.057	0.030

Table 7: Class Proportions (KNN)

	Data Set	Training Set	Validation Set	Test Set
pigeonpeas	0.045	0.046	0.034	0.052
pomegranate	0.045	0.046	0.043	0.045
rice	0.045	0.045	0.040	0.050
watermelon	0.045	0.047	0.065	0.025



Evaluation Metrics ▼

	apple	banana	blackgram	chickpea	coconut	coffee	cotton	grapes	jute	kidneybeans	lentil	maize	mango
Support	25	21	22	14	21	25	14	23	28	19	18	18	23
Accuracy	1.000	1.000	0.998	1.000	1.000	1.000	0.998	1.000	0.986	1.000	0.989	0.998	1.000
Precision (Positive Predictive Value)	1.000	1.000	1.000	1.000	1.000	1.000	0.933	1.000	0.867	1.000	0.783	1.000	1.000
Recall (True Positive Rate)	1.000	1.000	0.955	1.000	1.000	1.000	1.000	1.000	0.929	1.000	1.000	0.944	1.000
False Positive Rate	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.010	0.000	0.012	0.000	0.000
False Discovery Rate	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.133	0.000	0.217	0.000	0.000
F1 Score	1.000	1.000	0.977	1.000	1.000	1.000	0.966	1.000	0.897	1.000	0.878	0.971	1.000
Matthews Correlation Coefficient	1.000	1.000	0.976	1.000	1.000	1.000	0.965	1.000	0.890	1.000	0.879	0.971	1.000
Area Under Curve (AUC)	1.000	1.000	0.977	1.000	1.000	1.000	0.999	1.000	0.959	1.000	0.994	0.972	1.000
Negative Predictive Value	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000	0.998	1.000
True Negative Rate	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	0.990	1.000	0.988	1.000	1.000
False Negative Rate	0.000	0.000	0.045	0.000	0.000	0.000	0.000	0.000	0.071	0.000	0.000	0.056	0.000
False Omission Rate	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.002	0.000
Threat Score	∞	∞	21.000	∞	∞	∞	7.000	∞	2.600	∞	1.800	17.000	∞
Statistical Parity	0.057	0.048	0.048	0.032	0.048	0.057	0.034	0.052	0.068	0.043	0.052	0.039	0.052

Note. All metrics are calculated for every class against all other classes.

Figure 12: Evaluation Matrix for KNN algorithm

Rectangular Weight Function

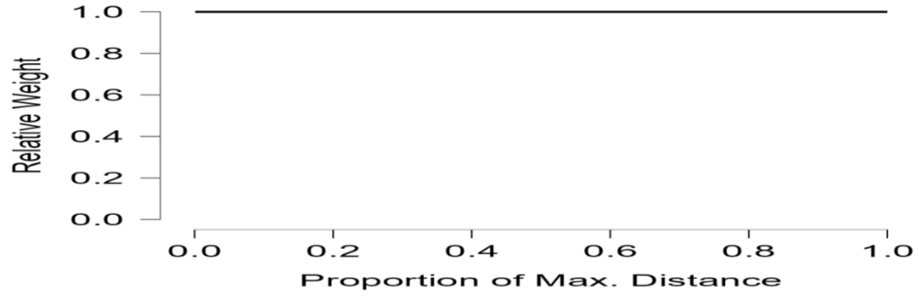


Figure 13: Rectangular Weight Function Plot (KNN)

Rectangular weight function as produced by KNN algorithm compared the distance between the propotion of MAX and relative Weight for ach feature variable.

Classification Accuracy Plot

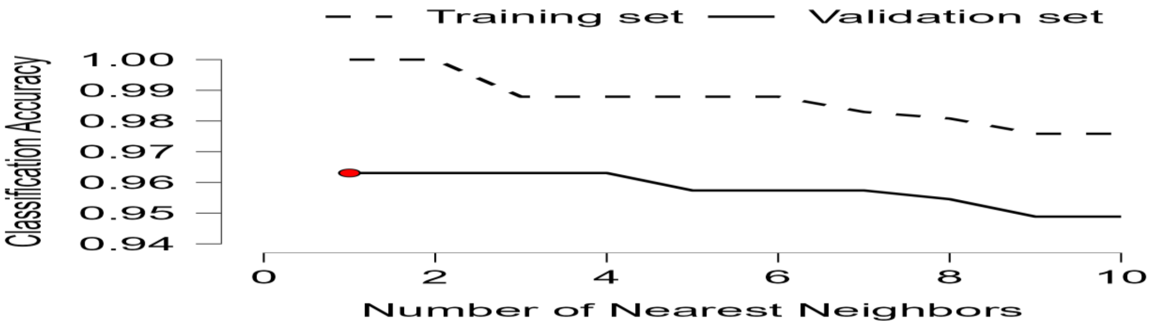


Figure 14: Classification Accuracy Plot (KNN)

This classification accuracy plot as produced by KNN shows the difference between the training set and its validation set.

ROC Curves Plot

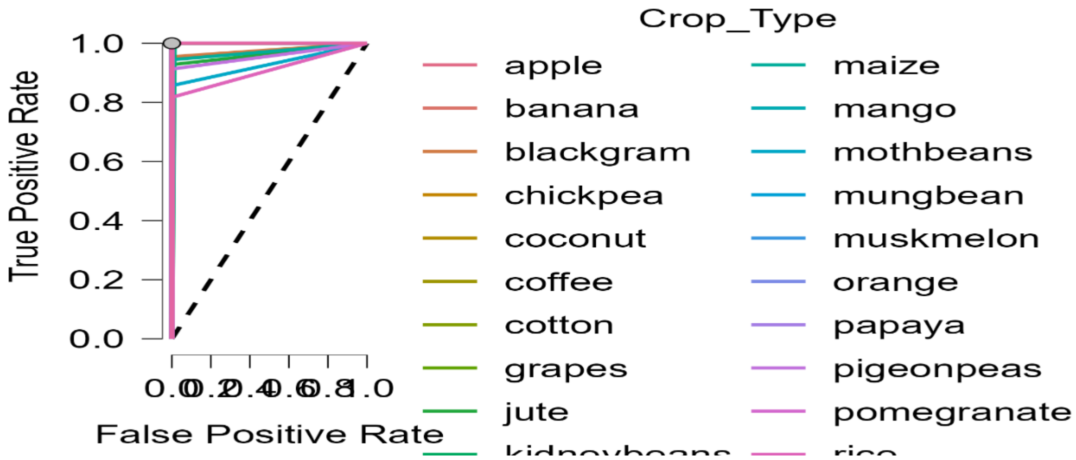


Figure 15: Classification Accuracy Plot (KNN)

Result as predicted by KNN algorithm shows apple, banana 95% accuracy for crop recommendation base on soil ph value, maize, mango, mothbeans, mungbean and muskmelon shows 86% accuracy for soil recommendation for each crop.

Andrews Curves Plot

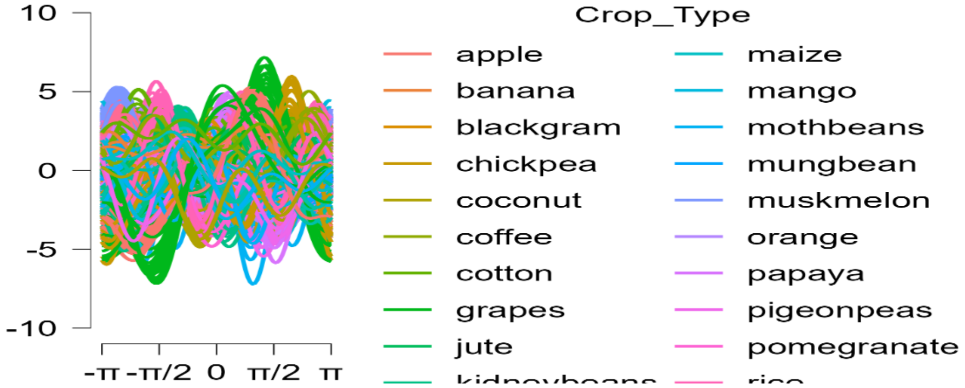


Figure 16: Andrews Curves Plot (KNN)

V. RESULT COMPARISON ON THE TWO DEVELOPED MODELS

At this point, the two results was compared by looking at the produced Evaluation metrics and classification results for DT and KNN shown in table 8, table 9, table 10 and 11 below.

Table 8: Decision Tree (DT) Classification

Splits	n(Train)	n(Test)	Test Accuracy
198	1760	440	0.952

The accuracy result as predicted by DT was 0.952 which is 95% accuracy recommended for adequate use for crop recommendation and faster decision making for farmers.

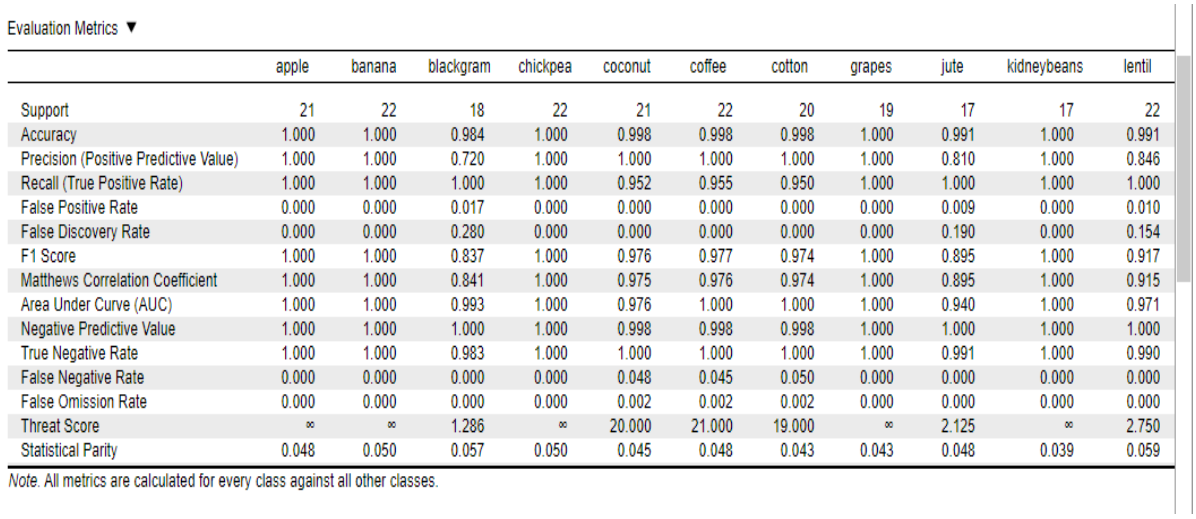
Table 9: K-Nearest Neighbors Classification

Nearest neighbors	Weights	Distance	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accuracy
1	rectangular	Euclidean	1408	352	440	0.949	0.970

Note. The model is optimized with respect to the *validation set accuracy*.

From the experiment conducted using KNN algorithm, table 9 above present the summary of the predicted accuracy between the Test and Train split on the dataset. KNN predicted an accuracy of 0.949 which is 94% of the validation set accuracy.

Table 10: Evaluation Result Metrics for DT

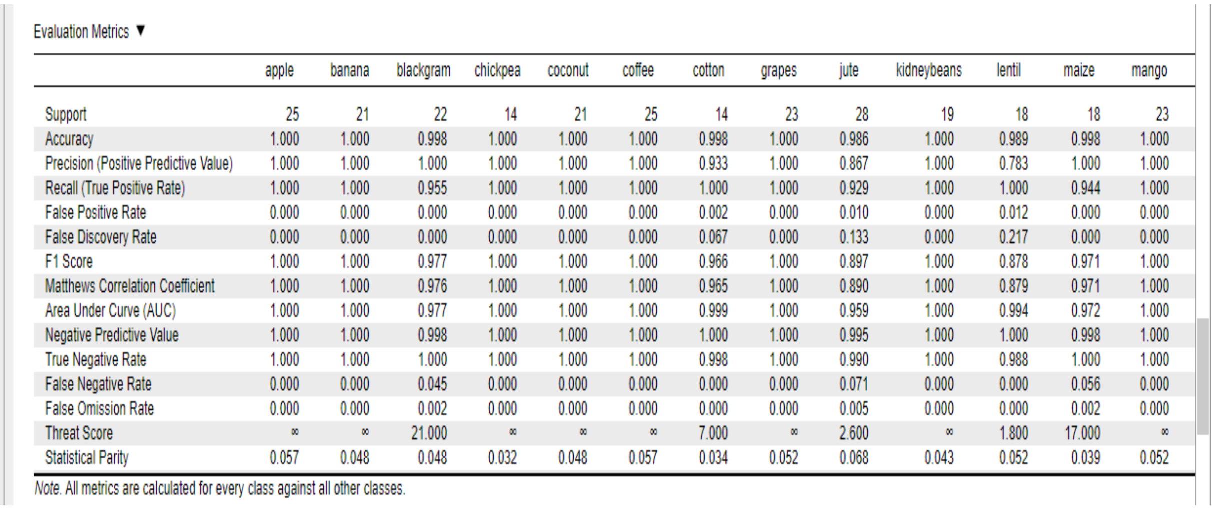


Evaluation Metrics ▼

	apple	banana	blackgram	chickpea	coconut	coffee	cotton	grapes	jute	kidneybeans	lentil
Support	21	22	18	22	21	22	20	19	17	17	22
Accuracy	1.000	1.000	0.984	1.000	0.998	0.998	0.998	1.000	0.991	1.000	0.991
Precision (Positive Predictive Value)	1.000	1.000	0.720	1.000	1.000	1.000	1.000	1.000	0.810	1.000	0.846
Recall (True Positive Rate)	1.000	1.000	1.000	1.000	0.952	0.955	0.950	1.000	1.000	1.000	1.000
False Positive Rate	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.010
False Discovery Rate	0.000	0.000	0.280	0.000	0.000	0.000	0.000	0.000	0.190	0.000	0.154
F1 Score	1.000	1.000	0.837	1.000	0.976	0.977	0.974	1.000	0.895	1.000	0.917
Matthews Correlation Coefficient	1.000	1.000	0.841	1.000	0.975	0.976	0.974	1.000	0.895	1.000	0.915
Area Under Curve (AUC)	1.000	1.000	0.993	1.000	0.976	1.000	1.000	1.000	0.940	1.000	0.971
Negative Predictive Value	1.000	1.000	1.000	1.000	0.998	0.998	0.998	1.000	1.000	1.000	1.000
True Negative Rate	1.000	1.000	0.983	1.000	1.000	1.000	1.000	1.000	0.991	1.000	0.990
False Negative Rate	0.000	0.000	0.000	0.000	0.048	0.045	0.050	0.000	0.000	0.000	0.000
False Omission Rate	0.000	0.000	0.000	0.000	0.002	0.002	0.002	0.000	0.000	0.000	0.000
Threat Score	∞	∞	1.286	∞	20.000	21.000	19.000	∞	2.125	∞	2.750
Statistical Parity	0.048	0.050	0.057	0.050	0.045	0.048	0.043	0.043	0.048	0.039	0.059

Note. All metrics are calculated for every class against all other classes.

Table 11: Evaluation Result Metrics for KNN



Evaluation Metrics ▼

	apple	banana	blackgram	chickpea	coconut	coffee	cotton	grapes	jute	kidneybeans	lentil	maize	mango
Support	25	21	22	14	21	25	14	23	28	19	18	18	23
Accuracy	1.000	1.000	0.998	1.000	1.000	1.000	0.998	1.000	0.986	1.000	0.989	0.998	1.000
Precision (Positive Predictive Value)	1.000	1.000	1.000	1.000	1.000	1.000	0.933	1.000	0.867	1.000	0.783	1.000	1.000
Recall (True Positive Rate)	1.000	1.000	0.955	1.000	1.000	1.000	1.000	1.000	0.929	1.000	1.000	0.944	1.000
False Positive Rate	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.010	0.000	0.012	0.000	0.000
False Discovery Rate	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.133	0.000	0.217	0.000	0.000
F1 Score	1.000	1.000	0.977	1.000	1.000	1.000	0.966	1.000	0.897	1.000	0.878	0.971	1.000
Matthews Correlation Coefficient	1.000	1.000	0.976	1.000	1.000	1.000	0.965	1.000	0.890	1.000	0.879	0.971	1.000
Area Under Curve (AUC)	1.000	1.000	0.977	1.000	1.000	1.000	0.999	1.000	0.959	1.000	0.994	0.972	1.000
Negative Predictive Value	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000	0.998	1.000
True Negative Rate	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	0.990	1.000	0.988	1.000	1.000
False Negative Rate	0.000	0.000	0.045	0.000	0.000	0.000	0.000	0.000	0.071	0.000	0.000	0.056	0.000
False Omission Rate	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.002	0.000
Threat Score	∞	∞	21.000	∞	∞	∞	7.000	∞	2.600	∞	1.800	17.000	∞
Statistical Parity	0.057	0.048	0.048	0.032	0.048	0.057	0.034	0.052	0.068	0.043	0.052	0.039	0.052

Note. All metrics are calculated for every class against all other classes.

In summary on the two compared developed model, the two algorithms on the machine learning classification model both for (DT = 95% and KNN =94%) developed an accurate results respectively showing a very improved model for crop recommendation per each crop and soil.

Comparing from the evaluation results shown in table 10 and 11 for each algorithm, Decision Tree (DT) and K-Nearest Neighbors (KNN) produced F1 Score result for the following variables (apple 100%, banana 100%, blackgram 83%, chickpea 100%, coconut 97%, coffee 97%, cotton 97%, grapes 100%, jute 89% and kidneybeans 100%) accuracy for perfect soil compatibility with high percent nutrient to grow such recommended crops. Close observation between results produced by DT on Precision, Recall, and F1 Score as shown in table 10 below predicted a very close outcome with no much different amongst the evaluations.

VI. CONCLUSION

As earlier stated, that the aim of this study was to develop an improved predictive model for crop recommendation and crop yield in rural areas using K-Nearest Neighbour (KNN) and Decision Tree (DT) machine learning classification algorithms. This paper was able to build a model called and referred to as Hybrid_Agro_Crop_Recomender as the developed model produced a very high accuracy of 95% and 94% respectively which could be used to make decision on the type of soil to plant the crops as stated below: (apple 100%, banana 100%, blackgram 83%, chickpea 100%, coconut 97%, coffee 97%, cotton 97%, grapes 100%, jute 89% and kidneybeans 100%). This model also proved that DT and KNN machine learning algorithms is very much good proper classification of objects especially whien it has to deal with sequential data.

VII. RECOMMENDATION

The researcher therefore recommends the following:

1. Other scholars could improve on the study by getting more variables and Crops
2. The Hybrid_Agro_Crop_Recomender accuracy could be compared with other machine learning techniques for enhance accuracy
3. Analysis of the model can be developed using three data analytical programming languages such as Python, R and JASP platform involving the integration of more machine learning or deep learning algorithms
4. A cloud base intelligent software could be developed to enable local farmers diagnose and predict perfect soil for their crops before venturing into planting.

REFERENCE

[1] Kilic, Ijaz (2020.), Plant Disease Management Strategies for sustainable Agriculture through Traditional and Modern Approaches, vol. 13, Springer Nature (2020)

[2] Simone Graeff, Judit Pfenning, Wilhelm Claupein, and Hans-Peter Liebig (2008) Evaluation of Image Analysis to Determine the N-Fertilizer Demand of Broccoli Plants (Brassica oleracea convar. botrytis var. italica), Hindawi Publishing Corporation Advances in Optical Technologies Volume 2008, Article ID 359760, 8 pages doi:10.1155/2008/359760

[3] Dezordi LR, Aquino LA, Aquino RFBA, Clemente JM, Assunção NS. (2016) Diagnostic Methods to Assess the Nutritional Status of the Carrot Crop. Rev Bras Cienc Solo;v40:e0140813

[4] Cunha et al., 2016 Cunha MLP, Aquino LA, Novais RF, Clemente JM, Aquino PR, Oliveira TF (2016) Diagnosis of the Nutritional Status of Garlic Crops. Rev Bras Cienc Solo;v40:e0140771.

[5] Manhas, S. S., Randive, R., Sawant, S., Chimurkar, P., & Haldankar, G. T. (2021, June). Nutrient Deficiency Detection in Leaves using Deep Learning. In 2021 International Conference on Communication information and Computing Technology (ICCICT) (pp. 1-6). IEEE.

[6] Ahmed, I., & Yadav, P. K. (2023). Plant disease detection using machine learning approaches. Expert Systems, 40(5), e13136.

[7] Ahmed, I., & Yadav, P. K. (2023). Plant disease detection using machine learning approaches. Expert Systems, 40(5), e13136.

[8] Nikitha, S., Prabhanjan, S., & Sathyanarayan, A. (2025). Plant nutritional deficiency detection: A survey of predictive analytics approaches. Iran Journal of Computer Science, 8(1), 83-101

[9] Nikitha, S., Prabhanjan, S., & Sathyanarayan, A. (2025). Plant nutritional deficiency detection: A survey of predictive analytics approaches. Iran Journal of Computer Science, 8(1), 83-101

[10] Jackulin, C., & Murugavalli, S. J. M. S. (2022). A comprehensive review on detection of plant disease using machine learning and deep learning approaches. Measurement: Sensors, 24, 100441

[11] Bera, A., Bhattacharjee, D., & Krejcar, O. (2024). PND-Net: plant nutrition deficiency and disease classification using graph convolutional network. Scientific Reports, 14(1), 15537.

[12] Anuja et al., (2024) Anuja Gedam, Ankita Sambre, Jagruti Pawar, Chetesh Yelekar, Ashish Trivedi(2024) Plant Disease Prediction System Using Machine Learning, International Journal for Multidisciplinary Research (IJFMR) E-ISSN: 2582-2160, Volume 6, Issue 6,: Website: www.ijfmr.com

[13] Prabira Kumar Sethy, Baishalee Negi, Nalini Kanta Barpanda, Santi Kumari Behera and Amiya Kumar Rath (2018) Cognitive Science and Artificial Intelligence, SpringerBriefs in Forensic and Medical Bioinformatics, https://doi.org/10.1007/978-981-10-6698-6_1

[14] El-Rashidy, N., Abdelrazik, S., Abuhmed, T., Amer, E., Ali, F., Hu, J. W., & El-Sappagh, S. (2021). Comprehensive survey of using machine learning in the COVID-19 pandemic. *Diagnostics*, *11*(7), 1155

[15] Yang et al., (2021) Yang Li1,2 and Xuewei Chao1 Li and Chao *Plant Methods (2021)*

17:68 https://doi.org/10.1186/s13007-021-00770-1

[16] Paramasivam Alagumariappan, Najumnissa Jamal Dewan, Gughan Narasimhan Muthukrishnan, Bhaskar K. Bojji Raju, Ramzan Ali Arshad Bilal and Vijayalakshmi Sankaran (2020). Convolutional Neural Networks in Agriculture Retrieved, from http://agriculture.iiit.ac.in/esagu/esagu2004/docs/ApeaAgrid04.pdf