Nigerian Journal of Technology (NIJOTECH)
Vol. 44, No. 1, March, 2025, pp.114 - 122
www.nijotech.com

Print ISSN: 0331-8443
Electronic ISSN: 2467-8821
https://doi.org/10.4314/njt.v44i1.13

APPLICATION OF RANDOM FOREST AND HIERARCHICAL CLUSTERING MODELS FOR CROP AND FERTILIZER RECOMMENDATION TO FARMERS

AUTHORS:
S. S. Nandom[1,*], G. T. Abe[2], I. P. Gambo[3]

AFFILIATIONS:
[1,2]Department of Computer Science, Federal University Wukari, Taraba, Nigeria.
[3]Department of Computer Science and Engineering Obafemi Awolowo University, Ile-Ife, Nigeria.

***CORRESPONDING AUTHOR:**
Email: sumayyah@fuwukari.edu.ng

ARTICLE HISTORY:
Received: 31 January, 2024.
Revised: 30 December, 2024.
Accepted: 14 February, 2025.
Published: 14 April, 2025.

KEYWORDS:
Data Analytics, Hierarchical Clustering, Random Forest Algorithm, Fertilizer Recommendation, Crop Recommendation.

ARTICLE INCLUDES:
Peer review

DATA AVAILABILITY:
On request from author(s)

EDITORS:
Ozoemena Anthony Ani

FUNDING:
None

Abstract

Specific recommendations of crop and fertilizer are two critical parts of developing effective agricultural and food policies in Nigeria and other parts of the world. One of the main problems that has negatively affected crop production is the depletion of soil nutrients. Hence maintaining soil nutrients has become a significant concern for farmers. Although fertilizers can be applied manually to increase crop production, it is not optimal since different crops in different fields require different amounts of fertilizer due to soil types, soil fertility levels, and nutrient needs. To effectively and efficiently improve and maintain soil fertility, it is necessary to replace the traditional trial and error method of Nitrogen (N) Potassium (P) and Phosphorus (K) variation at different ratios on untested soils (which most times leads to poor crop yield) with soil testing and fertilizer recommendation using data mining algorithms. This study developed a model to recommend crop and fertilizer using two machine learning algorithms. The RF algorithm, which has shown high level of accuracy in many different agricultural applications, is used for recommending crops, while the hierarchical Clustering algorithm is used for fertilizer recommendation. The models used Crop nutrient requirement and soil sample data for training and testing. The RF and hierarchical algorithm were trained to recommend crop and fertilizer on the basis of multiple biophysical variables and soil nutrients. The system was found effective in recommending crop and fertilizer with an accuracy of 99.70%. The results showed that the model performed effectively and it is versatile machine-learning model for recommending crop and fertilizer due to the high accuracy and precision values. This research pointed out various steps in which a crop and fertilizer recommendation system was achieved using a random forest and hierarchical Clustering algorithms.

HOW TO CITE:
Nandom, S. S., Abe, G. T., and Gambo, I. P. “Application of Random Forest And Hierarchical Clustering Models for Crop and Fertilizer Recommendation to Farmers”, *Nigerian Journal of Technology*, 2025; 44(1), pp. 114 – 122; *https://doi.org/10.4314/njt.v44i1.13*

1.0 INTRODUCTION

To maximize farm yields, farmers have applied different methods at different times. In the early ages of agriculture and crop production, in situations where the farm land lacks soil nutrients, the easiest method farmers applied to help the soil regain its nutrients was “bush fallowing” which involves not cultivating a farmland for a certain number of years such that the soil can regain its nutrients or natural fertility [1]. However, population growth has put some restrictions on how traditional bush fallowing can be used to maintain soil fertility. The use of improved fallows to hasten the natural retention of nitrogen and phosphorus by roots and leaves of leguminous trees and shrubs or herbaceous cover crops planted on fallow lands also faded with population increase [2].

Consequently, farmyard manure was also used to maintain soil fertility in the context of traditional farming. This practice is faced with a lot of limitations including the decreasing population of animals in the country, specifically since the drought of the 1970s and 1980s that led to drop in the number of cattle per head to its lowest level [3]. Over time, the application of mineral or inorganic fertilizer became fundamental for high-input agriculture as a substitute to organic fertilizer when an enhanced variety of crops were introduced. Thus, mineral fertilizer application is a major aspect of the green revolution, which goes along with some other inputs as requisites to ascertain the output potential of hybrid or other enhanced varieties generally.

Recently, farmers have attempted to use fertilizer recommendation systems to determine specific soil nutrients even before planting crops. In that wise, the approach of fertilizer recommendation is most beneficial given the outcome of soil testing as its input. For fertilizer recommendation systems, the accuracy depends on the quality of the soil sample used to suggest. It denotes the area in which the suggestion will be implemented. Because of natural forces and the management to which the land has been subjected, the physical and chemical qualities of soil in an area might vary significantly from place to place. Natural variation is caused by soil-forming processes (such as minerals, weathering, and erosion) that cause nutrient build up or loss at different locations.

To maintain the quality of soil in order to attain crop yield, it is paramount to use the right crop and the right amount of fertilizers which will bring the misuse of soil resources to its minimum. This can be achieved by having a very good knowledge of the exact condition of the soil, this include physical, biological and chemical condition of the soil through observing, investigating and testing the soil. A soil test is an essential tool used to evaluate fertilizer requirements for maintaining crop production and sustainable soil fertility. Soil fertility is a quality of soil that enables it to sustain plant growth in agriculture. Furthermore, soil fertility is connected to the order and level of organic matter, the capacity to store water, air, and nutrients is dependent upon soil fertility. Thus, the low productivity of farm produce is due to low soil fertility, which is as a result of poor fertilizer applications, particularly in high-intensity systems.

[3] proposed a policy reform to expand access to soil based fertilizer, e.g. NPK formula modified to suit soil fertility and crops through enabling programs such as: making soil map information accessible to farmers, enhancing soil testing and mapping, provision of information and the enhancement of creating awareness to farmers on the effective application of fertilizer, soil fertility management, fertilizer quality control and provision as well as the encouragement of targeted soil/ crop based fertilizer formulation by agribusiness. All of these are attempts made towards improving soil fertility for increased food production.

The work presented in this paper seek to increase farmers' crop yields. A framework was developed for crop-fertilizer recommendation using real life soil test results as a case study. A primary consideration for adopting this case study were because of the variety of crops being grown in the areas covered in the case study.

The remaining part of this paper has been systematically set up as follows. The context and inspiration for the research work is presented in Section 2. In Section 3, the deficiencies in crop-fertilizer recommender systems literature addressing different algorithm was examined and pointed out. The research methodology adopts the process of data collection for the algorithm, and the framework design for crop-fertilizer recommendation were discussed in Section 4. The result of the crop-fertilizer adoption recommendation system executed, serves as a tool for putting the framework designed into practice, all these were considered in Section 5. In Section 6, the validation of the framework is in detail, and Sections 7 discusses the findings. In Section 8, the conclusion and future research are described.

2.0 BACKGROUND AND MOTIVATION OF THE STUDY

Food security is the ability for a least earning individual to have access to food. It is a multifaceted idea that considers various aspects of the food system, such as production, processing, distribution, and consumption. Physical, economic, and social access to food supplies are all included in the concept of food security, to obtain the highest level of food security there must be adequate food suppliers, consistent flow of food and ensuring individuals in need can easily access food supplies. In Nigeria, there are many issues with food security one of which is soil fertility as described in [4, 1] “the depletion of soil fertility is one of the top problems that has constrained food security in West Africa”. Among many other challenges agriculturists and farmers face today, maintaining soil fertility is the most challenging [5, 6]. Although fertilizers can be applied manually to increase crop production, it is not optimal since different crops in different fields require different amounts of fertilizer due to soil types, soil fertility levels, and nutrient needs. To effectively and efficiently improve and maintain soil fertility, it is necessary to replace the traditional trial and error method of Nitrogen (N) Potassium (P) and Phosphorus (K) variation at different ratios on untested soils (which most times leads to poor crop yield) with soil testing and crop - fertilizer recommendation systems.

3.0 RELATED WORKS

Different researchers have proposed various approaches using different datasets mostly carried out on Indian soil nutrient dataset and different machine learning algorithms, such as J48, random Forest, K-means etc. to recommend crop and fertilizer, most of the literature reviewed discussed on the importance of having a crop and fertilizer recommendation system, others focus on developing a crop and fertilizer recommendation system using only Nitrogen N, Phosphorus P and Potassium K as the basis for recommendation while others are on the basis of region and soil type, which will not give a good percentage of accuracy.

Data analytics is a key solution for the challenges of food security and quality in Nigeria, as [7] and [3] suggested. Data analytics can help improve food productivity by using data-driven methods in agriculture. Data analytics can also help farmers cope with climate change and other environmental issues, as [7] noted. Data analytics can enable more innovative resource management and adaptation to changing conditions. Data analytics is also a vital tool for precision agriculture, which can enhance the agricultural supply chain's efficiency and responsiveness. Farmers can track their produce better, while distributors, retailers, and others can tailor their products and services to the market demand.

A fertilizer recommendation system is a useful tool for crop producers to apply the available nutrient resources rationally and economically [3]. Many studies have developed such systems with different levels of accuracy. Some of these studies have focused on recommending crops and fertilizers based on soil test data, such as NPK values. For example, [3] used the Random Forest algorithm, [7] used the nutritional balance method, and [7] used the Random Forest and K-means Clustering algorithms. Other studies have focused on recommending crops based on geographic and climatic parameters, such as temperature, rainfall, and location. For example, [7] used the Naive Bayes and other classifier algorithms, and [3] used the Support Vector Machine algorithm. Some studies have also used clustering methods to analyse the variability and classification of soil nutrients and crop traits. For example, [3] used the Principal Components Analysis and Agglomerative Hierarchical Clustering to determine soil nutrients and toxic substances in coffee farms in Brazil, and [8] used the Hierarchical Clustering to evaluate the morphological and physiological parameters of maize hybrids. According to [3], many factors influence crop yields, such as soil nature, soil moisture level, soil pH, soil temperature, and soil type.

Most of the literatures reviewed in this paper have some constraints. They either use only NPK values as the basis for recommendation, or they use only region and soil type as the inputs. They also use mostly Indian soil nutrient data sets, which may not be applicable to other regions. These limitations may affect the accuracy and effectiveness of the recommendation systems.

A summary of five other related papers are shown in Table 1 and some of their strengths and weaknesses has been identified on the basis of four factors that affects crop yield and the algorithm: (i) Covered major attributes of crop-fertilizer recommendation., (ii) Application of real life test data (primary data), (iii) Application of a hybrid model., (iv) Scope (Areas within Nigeria).

In this research, a novel approach was developed for crop and fertilizer recommendation using both Random Forest and Hierarchical Clustering algorithms. Local soil data sets from some local government areas of Plateau state was used, which have not been used before for this purpose. Other attributes besides NPK values, such as humidity, rainfall, pH level, and temperature was also used. The aim is to bridge the gap in the literature by increasing crop yield and accuracy of the recommendation systems.

Table 1: Summary of related works

Criteria for Recommendation	[3]	[7]	[10]	[9]	[1]	CFRM(Our Model)
Covered major attributes of crop-fertilizer recommendation	✓	✓	×	✓	✓	✓
Application of real life test data (primary data)	×	×	×	×	✓	✓
Application of a hybrid model.	×	✓	✓	×	✓	✓
Scope (Areas within Nigeria)	✓	×	×	×	×	✓

4.0 METHODOLOGICAL APPROACH

Data collection is the process of gathering and measuring information on the data attributes of interest in a systematic and reliable way that answers the stated research questions, tests the hypotheses, and evaluates the possible outcomes [7]. This study adopted the quantitative approach, which involves using numerical data and statistical methods to analyze the research problem. The research dataset was collected from three sources: URBA Ventures, which is a platform that provides data-driven solutions for sustainable agriculture; Kaggle, which is an online community that hosts data science competitions and datasets; and GitHub, which is a web-based platform that hosts open-source software projects and datasets. Two types of dataset were used one for crop recommendation and the other for fertilizer recommendation, the data used in training the model for crop recommendation is soil nutrient dataset which consists of six (6) parameters: Nitrogen N, Potassium P, Phosphorus K, temperature, Humidity, and pH level, While the data used in training the model for fertilizer recommendation is crop nutrient dataset which consist of the N, P, K and crop type.

The entire dataset for crop recommendation consists of 2000 samples of soil nutrients. While that of fertilizer recommendation consist of 1600 samples of crop type and nutrient. Each of these dataset is divided into two: 70% for training and 30% for testing and validation, using 70% of the dataset for training the models and 30% for testing, the dataset consists of crop and fertilizer recommendations for different types of soil and weather conditions. Table 2 shows the pre-processed data for crop recommendation, the pre-processing involves extracting and arranging the columns for the equivalent values of Nitrogen, Phosphorus, potassium, Temperature, Humidity and the PH Level. The pre-processing dataset for fertilizer recommendation involves identifying the NPK nutrient requirement for various crop, Table 3 shows the pre-processed data for fertilizer recommendation.

Table 2: Pre-processed data for crop recommendation

N	P	K	Temperature	Humidity	PH
90	42	43	20.88	82.00	6.50
85	58	41	21.77	80.32	7.04
60	55	44	23.00	82.32	7.84
74	35	40	26.49	80.16	6.98
78	42	42	20.13	81.60	7.63
69	37	42	23.06	83.37	7.07
69	55	38	22.71	82.64	5.70
94	53	40	20.28	82.89	5.72
89	54	38	24.52	83.54	6.68

Table 3: Pre-processed data for fertilizer recommendation

Crop	N	P	K
Rice	80	40	40
Maize	80	40	20
Chickpea	40	60	80
Kidneybeans	20	60	20
Pigeonpeas	20	60	20
Mothbeans	20	40	20
Mungbean	20	40	20
Blackgram	40	60	20
Lentil	20	60	20
Pomegranate	20	10	40
Banana	100	75	50
Mango	20	20	30
Grapes	20	125	200
Watermelon	100	10	50
Muskmelon	100	10	50
Apple	20	125	200
Orange	20	10	10
Papaya	50	50	50
Coconut	20	10	30
Cotton	120	40	20

The classification tool is designed using two algorithms: Hierarchical Clustering algorithm with Random Forest algorithm. The Hierarchical Clustering algorithm is used to group the crops into different clusters based on their nutrient requirements. The Random Forest algorithm is used to classify the soil samples into one of the clusters based on their nutrient levels. The data of the soil nutrients collected from the field is pre-processed to remove outliers and missing values and then split into training and testing sets. The training set is used to train the Random Forest algorithm and the testing set is used to check the performance of the RFA on the test data. The output of the algorithm is the most suitable crop cluster for each soil sample based on the available soil nutrients.

Random Forest algorithm (RFA)

The RF algorithm is trained with the soil data, which contains the values of nitrogen (N), phosphorus (P), and potassium (K) for each soil sample. The algorithm then splits the data into multiple decision trees, each representing a different combination of N, P, and K levels. For each tree, the algorithm follows these steps: (1) randomly select a subset of features (N, P, or K) from the data; (2) calculate the best split point for each feature based on a criterion such as entropy or gini index; (3) repeat this process until all the nodes in the tree are either pure or reach a predefined limit. The final prediction for the best crop to plant for each soil sample is obtained by taking the majority vote of all the trees, i.e., choosing the crop cluster that has the highest frequency among the trees.

Hierarchical Clustering Algorithm (HCA)

The HCA is used to recommend the appropriate fertilizer to be used to boost crop yield based on the soil nutrient levels. The algorithm creates a hierarchy of clusters from the data in, which contains the values of N, P, and K for each soil sample. The algorithm follows these steps: (1) Assigning each soil sample to a cluster, such that there are N clusters initially, each containing one soil sample; (2) Finding the closest pair of clusters based on a distance measure such as Euclidean distance or cosine similarity and merging them into a single cluster, such that the clusters are grouped based on their similarity or consistency of the nutrient patterns; (3) Computing the distances between the newly formed cluster and the remaining clusters; (4) Repeating steps 2 and 3 until all soil samples are clustered into a single cluster of size N. The hierarchy of clusters can be represented by a tree-like structure called a dendrogram, which shows the order and level of merging of the clusters. The algorithm then cuts the dendrogram at a certain level to obtain the optimal number of clusters for the data. Each cluster represents a different fertilizer recommendation based on the nutrient levels of the soil samples in that cluster.

4.1 Model Design

The design model for the problem is achieved by applying the classification methods of data mining techniques. This means that the design of the model will be as follows:

i. Collect soil nutrient data
ii. Prepare the Data in (i)
iii. Formulate the model
   a) RFA for crop prediction based on available soil nutrients
   b) Hierarchical Clustering for fertilizer recommendation
iv. Evaluate the model in (iii)
v. Deploy the model in (iii)

In developing the model, data were classified into clusters and trees at different levels. These data grouping or classification levels included a series of data training and validations from the pre-processed dataset.

How the Algorithm Works

The Voting Ensembles combines the hierarchical Clustering algorithm with the random forest algorithm into a model. The Soft Voting approach is used against Hard Voting because Soft voting can accurately and efficiently predict the class with the most considerable probability based on summing from the models. Hence, the reason for choosing and deploying this approach. Figure 1 shows the flowchart of the system.

Random Forest Algorithm for Crop Recommendation

Input: N = Size of training dataset, C = Soil attributes, d = crop
Output: Recommended Crop
1. generate aggregated training (St) and testing (Se) set.
2. i=0
3. while (i<N) do
4. randomize C
5. RTi = buildRandomTree(St,C,d)
6. i = i +1
7. end while
8. RF = {RT0,RT1,.......,RTN-1}
9. return RF

The algorithm starts with an "active set" of clusters S, which has two subsets: St for the training data and Se for the testing data. The algorithm builds several decision trees using the data points from St, c, and d. The algorithm iterates until the counter "i" reaches the size of the training set "N". The forest is formed by combining the decision trees.

Hierarchical Clustering Algorithm for Fertilizer Recommendation

Input: Input: Data vectors {xn}Nn=1, group-wise distance D (G, G′)
Output: Recommend Fertilizer
Method: Build Hierarchical Clustering
1. $A \leftarrow \emptyset$
2. for $n \leftarrow 1 \ldots N$ do
3. $A \leftarrow A \cup \{\{x_n\}\}$
4. end for
5. $T \leftarrow A$
6. while $|A| > 1$ do
7. $G_1, G_2 \leftarrow \arg\min \mathrm{DIST}(G_1, G_2)$
8. $A \leftarrow (A \backslash \{G_1\}) \backslash \{G_2\}$
9. $A \leftarrow A \cup \{G_1 \cup G_2\}$
10. $T \leftarrow T \cup \{G_1 \cup G_2\}$
11. end while
12. Return: Tree T.

The algorithm above contains an “active set” of clusters A, which starts as an empty set. It then loops through the data and adds each datum as a cluster in the Active set A, and at every stage, it chooses which two clusters to merge, after which, they are removed from the active set A, and the union of the merged clusters is added to the active set. This process will continue to iterate until there is only one cluster in the active set, then the tree is formed by keeping track of the merged clusters.

4.2 Model Evaluation

The proposed system takes the input values of nitrogen (N), phosphorus (P), potassium (K), humidity, rainfall, pH level, and temperature to recommend the best crops for a given soil sample. The system uses Hierarchical Clustering algorithm, which is an unsupervised clustering algorithm that creates clusters in a hierarchical presentation from top to bottom. The algorithm clusters the data into three classes based on the nitrogen (N), phosphorus (P), and potassium (K), concentration in the soil. The mathematical representation for Hierarchical Clustering algorithm is shown in Equation 1:

$$D_{(rs)=Trs/(Nr*Ns)} \qquad (1)$$

Where, *Trs* is the sum of all pairwise distances between cluster r and cluster s, *Nr* and *Ns* are the clusters r and s sizes, respectively.

At each step of the Hierarchical Clustering algorithm, the clusters r and s that have the minimum distance D (r, s) are merged. An Random Forest algorithm is considered a supervised learning algorithm that is based on the idea of building a "forest" using several “decision trees”, the RF is trained using the "bagging" also known as bootstrap aggregation. The algorithm uses mean squared error (MSE) to split the nodes of each tree when solving regression problems. The mathematical formula of the Random Forest algorithm is given by Schott [11] in Equation 2:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2 \qquad (2)$$

Where, N is the number of data points, *fi* is the value returned by the model, *yi* is the actual value for the data point.

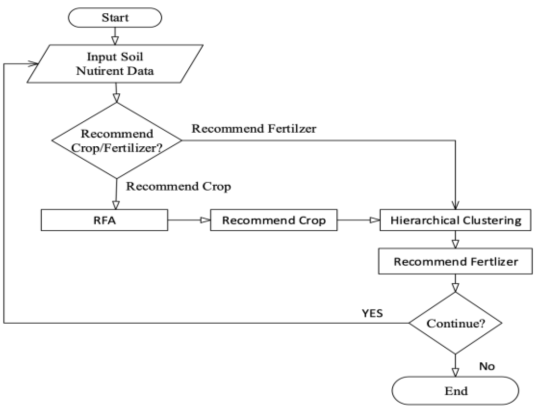

Figure 1: Flow chart of the system

5.0 RESULTS

Figure 2a and 3a shows a user interface with inputted soil data values, after these entries the user clicks on the Recommend button, thus taking the user to the user interface for crop and fertilizer recommendation and displays the recommended crop and fertilizer in Figure 2b and 3b based on the inputted soil data, the system recommends the most suitable crop and most suitable NPK fertilizer for both crop and soil.

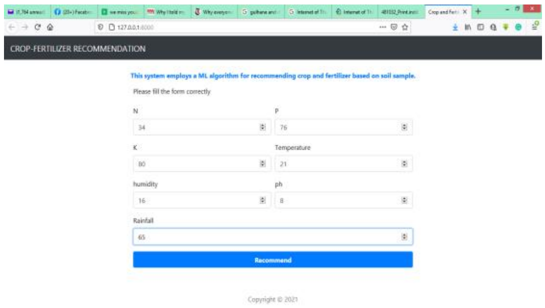

Figure 2a: User interface for entering soil data (A)

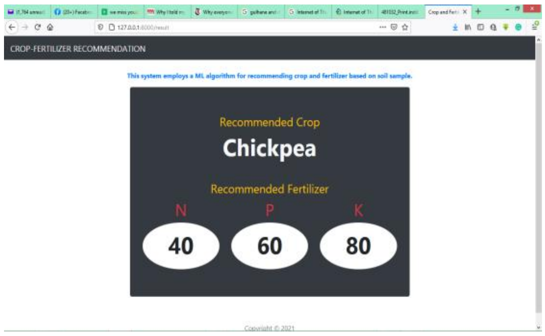

Figure 2b: Interface of the developed system and output screen for recommended crop and the ratio of NPK required for quality yield (A)

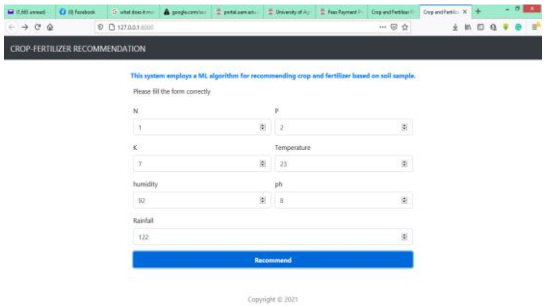

Figure 3a: User interface for entering soil data (C)

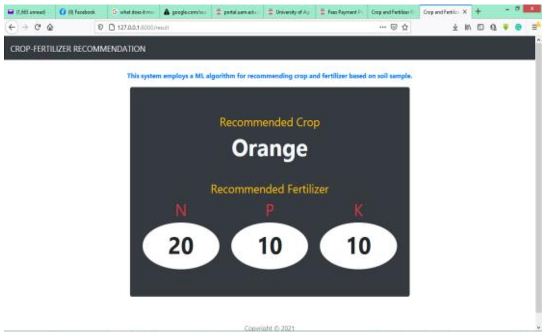

Figure 3b: Interface of the developed system and output screen for recommended crop and the ratio of NPK required for quality yield (C)

6.0 MODEL VALIDATION

Evaluation metrics are used to measure the performance of machine learning models or algorithms. There are different evaluation metrics, including confusion matrix, logarithmic loss, and classification accuracy. The ratio of correct predictions to the total number of input samples. Logarithmic loss, or log loss, penalizes incorrect classifications is considered classification accuracy. Confusion matrix is a matrix that shows the model's performance in terms of true positives, false positives, true negatives, and false negatives, as shown in Figure 4. Therefore, this study uses F1 score, accuracy, precision, and recall as the evaluation metrics for the proposed system. These metrics are suitable for classification and prediction problems such as this system.

TP: The True Positive TP represent the number of soil samples within a class "soil data" that have been correctly classified.
TN: The True Negative TN represents the number of soil samples that is not within a class "soil data" that have been correctly classified.
FN: The False Negative FN represents the number of soil samples that is not within a class "soil data" but has been incorrectly classified.
FP: The False Positive FP gives the number of soil data samples incorrectly classified in another class.

Precision P is calculated for each column (Actual Label) in the model as shown in the confusion matrix using the formula below,

$$P1 = \frac{TP}{TP + FP}$$

$$P = \frac{P1+P2\ldots+Pn}{TNL} \quad (3)$$

Where, P1, P2….Pn represent each actual class in the model, and TNL represents the Total Number of actual Label in the model.

Recall R is calculated for each row (Predicted Label) in the model as shown in the confusion matrix (see Figure 4) using the formula below,

$$R1 = \frac{TP}{TP + FN}$$

$$R = \frac{R1+R2\ldots+Rn}{TNL} \quad (4)$$

Where, R1, R2…. Rn, represent each Predicted class in the model, and TNL represents the Total Number of predicted Label in the model.

F1- Score is a harmonic mean of precision and recall. It is calculated using the formula below,

$$F = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \quad (5)$$

Accuracy A is calculated as the total number of correct predictions divided by the total number of datasets

$$A = \frac{(TP1+TP2\ldots+TPn)}{(TP+TN+FP+FN)} \quad (6)$$

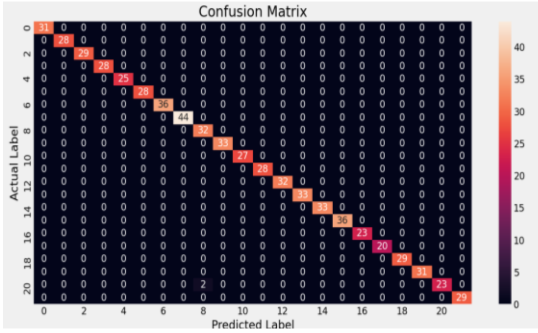

Figure 4: Confusion Matrix for the Fertilizer Recommendation Model

Table 3: Representation of Binary Confusion Matrix

Accuracy = (TP+TN)/TP+TN+FP+FN		Actual Classification	
		Yes	No
Predicted	Yes	TP	FP
Result/Classification	No	FN	TN

Table 4: Performance Evaluation of the Fertilizer Recommendation Model

Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
99.71	99.69	99.67	99.70

Table 5: Qualitative Comparison of the Result

Performance Metric	[9]	[7]	CFRM
Accuracy (%)	90.01	95	99.70
Recall (%)	93.4	95.9	99.67
Precision (%)	95.0	88.3	99.71
F1 score (%)	90.3	86.4	99.67

7.0 DISCUSSION

Determining or recommending the most suitable crop is done by the random forest algorithm. The process involves creating a forest by combining three soil nutrients (NPK) and predicting the best crop with high yield based on the correlation patterns from the dataset. Using the random classifier embedded in the algorithm, the combination of these soil nutrients parameters produces initial crops that may be correct or incorrect. Thus, the averaging or votes for the individual soil parameters are taken by the algorithm to ensure that the final recommendations outputted are correct.

This research paper reviews different literature on crop and fertilizer recommendation systems that use various data mining approaches and data attributes. Some of the previous studies used Random Forest Algorithm (RFA) or Support Vector Machine (SVM) algorithm to recommend crops and fertilizers based on soil nutrient values such as NPK. However, this research paper uses field-based soil nutrient data from three local government areas of Plateau State in Nigeria, and considers other factors such as humidity, rainfall, pH, and temperature. It also uses Hierarchical Clustering algorithm to recommend fertilizers, which is a novel approach in this domain.

The study adopted the system F1 score, Accuracy, Precision and Recall measuring the system's performance. These performance evaluation metrics are best suitable for classifying and predicting scenarios upon which this system is based. The results for this research were based on random forest and hierarchical Clustering algorithms, after evaluating the system using performance metric. The results shows that the proposed system has an accuracy of 99.70%, which is higher than the previous systems reviewed in this paper. This paper demonstrates the potential of Hierarchical Clustering as a data mining technique for fertilizer recommendation. A confusion matrix was used to describe the model's overall performance. When compared with two selected work, this research outperform theirs both quantitatively and qualitatively as shown in Table 5.

The system was implemented using the Python programming language, the algorithms were run, and results were read from a single system unit to avoid inconsistency due to differences in system configuration and settings. A Hp desktop computer with 8GB RAM and 500GB ROM running windows 10 was used for the experiment.

8.0 CONCLUSION

One of the major challenges facing Africa is food insecurity, which can be alleviated by maximizing farm yields. To achieve this, farmers need to know the right crops to plant and the right fertilizers to apply based on the soil quality. Soil quality is being assessed by measuring the biological, chemical, and physical properties of the soil through various methods. This research paper describes a model that uses Random Forest Algorithm (RFA) and Hierarchical Clustering algorithm to recommend the optimal amount of NPK fertilizer for different soil samples. The paper reviews the existing literature on crop recommendation systems that use various machine learning algorithms such as neural networks and decision trees. It also highlights the novelty of using Hierarchical Clustering algorithm for fertilizer recommendation, which is a data mining technique that has not been explored in this domain. The paper describes the steps involved in developing and evaluating the proposed model using a field-based soil nutrient data set from Nigeria.

Researchers have used different machine learning algorithms for crop recommendation systems though not hierarchical Clustering. this research has explored the use of both random forest algorithm and hierarchical Clustering algorithm to recommend crop and fertilizer, this research bridges the gap between the traditional method of planting, fertilizing and soil test based planting, fertilizing, where the former leads to low crop yield and nutrient while the later leads to high crop yield and nutrient, thus addresses the concern for food security in the nearest future. this research has provided a model that has a two-in-one system for recommending both crop and fertilizer. Thus, optimizing crop production by utilization of the developed system which promises a boost in crop yields.

This research can be further extended to include functionalities such as IoT to detect soil nutrients, image processing to detect Crop diseases such that the user gets pesticides recommendation based on the crop disease images, crop yield prediction and to Implement Smart Irrigation System for farms to get better yield. Farmers may not always have access to computers before they can recommend crops or fertilizers; thus, the system can be made available for distribution on mobile devices.

9.0 ACKNOWLEDGMENT

The authors would like to thank all the experts at URBA Ventures who carried out the soil test that provide this research with field-based test data.

REFERENCE

[1] Yemefack, M. “Preserving Soil Fertility for Ensuring Food Security in Africa”, *in proceedings of the American Association for the Advancement of Science, Chicago*, 2014. https://www.researchgate.net/publication/ 268445205_Preserving_Soil_Fertility-Ensuring-Food_Security_in_Africa.

[2] Nair, P. K. R., Kumar, B. M., Nair, V. D. “Soil Organic Matter (SOM) and Nutrient Cycling. In: An Introduction to Agroforestry”, *Springer, Cham*. 2021. https://doi.org/10.1007/978-3-030-75358-0_16

[3] Yange, T. S., Egbunu, C. O., Rufai, M. A., Onyekwere, O., Abdulrahman, A. A., Abdulkadri, I. “Using Prescriptive Analytics for the Determination of Optimal Crop Yield”, In: *International Journal of Data Science and Analysis*, 6(3), 72; 2020.

[4] Dimkpa, C., Adzawla, W., Pandey, R., Atakora, W. K., Kouame, A. K., Jemo, M., and Bindraban, P. S. "Fertilizers for food and nutrition security in sub-Saharan Africa: An overview of soil health implications", *Frontiers*

in Soil Science, vol. 3, Article 1123931, 2023. doi: 10.3389/fsoil.2023.1123931.

[5] Gamage, A., Gangahagedara, R., Gamage, J., Jayasinghe, N., Kodikara, N., Suraweera, P., and Merah, O. "Role of organic farming for achieving sustainability in agriculture", *Farming System*, vol. 1, no. 1, p. 100005, 2023. doi: 10.1016/j.farsys.2023.100005.

[6] Khatri, P., Kumar, P., Shakya, K. S., Kirlas, M. C., and Tiwari, K. K., "Understanding the intertwined nature of rising multiple risks in modern agriculture and food system," *Environment, Development and Sustainability*, vol. 26, pp. 24107–24150, 2024. doi: 10.1007/s10668-023-03638-7.

[7] Viviliya, B., and Vaidhehi, V. "The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms", In: *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*. 9(2), 4305-4311; 2019.

[8] Kaur, J., Ghai, N., Chahal, G. K., and Sandhu, S. "Characterization of maize inbred lines using principal component and cluster analysis for heat tolerance at the seedling stage under in vitro conditions", *Cereal Research Communications*, vol. 52, no. 4, pp. 1431–1442, 2024.

[9] Palaniraj, A., Balamurugan A.S., Durga, P. R., Pradeep, P. "Crop and Fertilizer Recommendation System using Machine Learning", In: *International Research Journal of Engineering and Technology (IRJET)*, 8(4): 319- 323, 2021.

[10] Chougule, A., Jha, V. J., Mukhnopadhyay, D. "Crop Suitability and Fertilizers Recommendation using Data Mining Techniques", In: Progress in Advanced Computing and Intelligent Engineering, 714, 25–35, 2019; https://doi.org/10.1007/978-981-13-0224-4

[11] Schott, M. "Random Forest Algorithm for Machine Learning", *Medium, Capital One Tech*, 2019; Retrieved from https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb.