engineering proceedings

Proceeding Paper

Enhancing Soil Fertility Prediction Through Federated Learning on IoT-Generated Datasets with a Feature Selection Perspective †

Murali Krishna Senapaty [1], Abhishek Ray [1] and Neelamadhab Padhy [2,*]

[1] School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Pin-751 024, India; muralisenapaty@gmail.com (M.K.S.); arayfcs@kiit.ac.in (A.R.)
[2] School of Engineering, & Technology, GIET University, Gunupur, Pin-765022, India
* Correspondence: dr.neelamadhab@giet.edu
† Presented at The 11th International Electronic Conference on Sensors and Applications (ECSA-11), 26–28 November 2024; Available online: https://sciforum.net/event/ecsa-11.

Abstract: Introduction: The fertile soil has a balanced nutrient value of pH, potassium, phosphorus, nitrogen, water retention capability, and organic substances. A fertile soil allows for better plant growth, leading to better production. The soil fertility requirement varies from crop to crop. So it is essential to identify the soil's fertile level according to the crop type. Objective: The objective of this paper is to develop a robust model that is capable of predicting soil fertility. The model is integrated with the IoT-generated data and federated learning-based feature selection techniques to improve the accuracy of the dataset. Material/Methods: Different feature selection techniques were applied to the dataset. Then we applied machine learning algorithms such as Logistic Regression, Decision Trees, Naive Bayes, and their ensemble to analyze and improve the performance. The federated learning approach is implemented for training the local models using the individual partitioned datasets. Each local model of the client shares the cryptic output weight and bias without sharing raw data. There is a centralized model at the server end that collects these weights and biases by preserving data privacy. These collected data are aggregated and applied to find a least square error (LSE). Then a gradient descent curve (GDC) is applied to identify the optimized weight and bias which shall be fed back again to improve the accuracy of prediction. Result: From our experimental observation, we analyzed the performance metrics of different ML classifiers and it revealed that the ensemble of logistic regression and decision tree has better performance compared to other models. One of our client models generates weight and bias with a precision of 87%, accuracy of 87%, recall of 87%, and F1-Score of 86%. Further, we have collected two of our client system model outcomes at a server model and applied the LSE to identify the optimal W and B. Further, in future work, we shall improve the performance of our model with a recursive approach by verifying the W, and B at the client model in a feedback process.

Keywords: soil fertility; IoT sensors; federated learning; machine learning

Citation: Senapaty, M.K.; Ray, A.; Padhy, N. Enhancing Soil Fertility Prediction Through Federated Learning on IoT-Generated Datasets with a Feature Selection Perspective. *Eng. Proc.* **2024**, *5*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: 26 November 2024

1. Introduction

The crop production is mainly based on soil and weather conditions. In which the soil health plays a very important role in crop cultivation. The soil nutrients need to vary from crop to crop. So it is important to get updated knowledge of soil nutrients so that a suitable crop can be selected and also the nutrient deficiency can be fulfilled.

In this approach, our main goal is to identify whether the soil is fertile or not. This is because fertile soil will have a balanced nutrient. When the soil is fertile, then by proper availability of water, temperature, and organic matters the crop can grow better and lead to better production.

To identify the fertility status of a soil sample we shall prepare a prediction model. Before training a model feature selection is very important in the preprocessing phase. So different selection techniques such as Variance Threshold, Chi-Square, Recursive Feature Elimination (RFE), PCA Loadings, Random Forest (RF) Importance, Mutual Information, and Lasso Coefficients) shall be applied. Once the feature selection and preprocessing are done now the dataset shall be used to train a model. In really many clients may not be interested in sharing their data with others. So, for better security of the data of each client and also to provide a better prediction model a novel approach can be implemented.

Federated learning allows us to maintain data security at the client end and also prepare a model at the server end with high accuracy. So the individual models shall be trained at the client end and then the results shall be shared cryptically to the server. The server shall contain cumulative details of all the predictions of local models and shall be used for further analysis. Here some aggregation and filtering techniques shall be applied to the dataset collection so that the model at the server can improve its performance in predictions.

In this concern, different machine learning techniques and an ensemble of their combinations shall be used for predictions. So a balanced integration of IoT sensors and federated learning approach allows for the maintenance of data security at each client and also improves the performance of the centralized server model. The local clients' training models at their end generate weight and bias. There is a centralized model at the server end that collects these weights and biases, aggregates them, and supplies them back to improve their performance. This approach maintains data privacy and also improves accuracy.

Then the soil mineral such as moisture, pH, nutrient levels, etc. data was collected through the IoT sensors of different agricultural fields. These data are supplied as input to the centralized model which predicts whether the soil is fertile or not.

2. Literature Review

The soil fertility analysis is essential for identifying suitable crops for cultivation. The fertile soil is useful and can produce crops in a better way. But when a soil fertility rate is decreased then by identifying it we shall take necessary steps to make it fertile and ready for crop cultivation. Here we review the relevant research papers so as to identify the work done so far on soil fertility analysis and crop predictions [1]. The author applied an Explainable AI technique based on Random Forest to predict soil fertility. It has been observed that an analysis has been done on a dataset of the European Union with an accuracy of 97% [2]. The author introduced a weighted K-mean algorithm to evaluate soil fertility. An integration of the Analytic Hierarchy Process to obtain the weights of soil nutrient attributes. It is observed that this approach has a better accuracy of 96.91% compared to traditional K-mean clustering [3]. The author analyzed the Indian agriculture dataset for soil fertility using classification techniques. It is found that the Support Vector Machine obtained a higher accuracy of 80% compared to other machine learning classifiers [4]. The author found that prediction was achieved with an accuracy of 78% to analyze the soil fertility using the multiple linear regression (MLR) method. A model proposed which demonstrates an effective analysis and predictions using the MLR technique [5]. The author reviewed 1328 articles out of which a selective 20 articles were chosen based on their efficiency in predictions. The soil was analyzed by many for minerals, water, air, and organic matter using different machine learning techniques such as linear regression, Support vector machines, K-nearest Neighbors, Decision Trees, Artificial Neural Networks, Naive Bayes, etc. The Random forest and deep learning methods have better performance compared to others [6]. The author proposed a framework using ANN for soil nutrient quality analysis. It is observed that ANN with Relu and Tanh gives an average accuracy of 90% in fertility classification. The author analyzed the area of Uttarakhand state, India to identify the soil fertility needs to avoid unnecessary use of fertilizers [7]. The authors studied the nutrient levels of mulberry gardens in Tamil Nadu. He applied an extreme

Learning Machine (ELM) with various activation functions and observed that the soil had normal electric conductivity, abundant in nitrogen, potassium, and sulfur but there was a deficiency in magnesium, copper, zinc, etc. [8] The author proposed a model for testing a soil is fertile or not in terms of 1 or 0. Then the output is compared with actual values. Then the model predicts the crops based on available soil features. Various machine learning techniques were applied for it out of which Random Forest was obtained with 100% accuracy in prediction [9]. The author applies machine learning techniques such as decision trees, K nearest Neighbour, Support Vector Machine, etc. to identify the soil fertility and in turn, improve crop productivity. The performance is measured in terms of cross-validation, absolute error, and accuracy. It is seen that the accuracy rate becomes 99% for the Decision Tree technique [10]. The author applies an extreme learning machine (ELM) along with many activation functions to classify the soil features. Here the Gaussian radial basis function is applied and obtains a higher accuracy of 90%. So it optimizes the use of nutrients and also reduces the fertilizers purchasing cost.

Overall the review, we observed that the soil analysis for fertility has been conducted using different learning techniques and methodologies to measure the fertility levels to identify the suitable crops and also the fertility needs so that there can be use of low fertilizers and maintain the soil health.

3. Materials and Methods

The methods at the client and server are different. At the client, the model shall apply a uniform model to its existing dataset. We used feature selection techniques on a dataset in each client to improve model performance, reduce the complexity, and enhance the data quality. Then we applied a model consisting of techniques such as Logistic regression, Decision Tree, Naïve Bayes, and their ensemble approaches for identifying the best model fits to have an optimized analysis. Then the encrypted results are sent to the server-end model.

3.1. Feature Section Techniques

We applied different features and compared their performance. The Variance Threshold is used for selecting the features whose variance exceeds the threshold value based on the given formula in Equation (1):

$$Var(X_j) > threshold \tag{1}$$

The Chi-Square is calculated to identify the dependencies between a feature and the target variables shown in Equation (2):

$$\chi^2 = \sum (O_i - E_i)^2 / E_i \tag{2}$$

where O_i is the observed frequency and E_i is the expected frequency. The Recursive Feature Elimination is used for eliminating the least important features based on the model coefficient. Also, the PCA Loadings are applied to identify the important features in the dataset. Random Forest applied to identify the importance of certain features based on entropy shown in Equation (3):

$$Importance(X_j) = \sum_{t \in T} \Delta I_t (X_j) \tag{3}$$

where $\Delta I_t(X_j)$ is the decrease in impurity for feature X_j in tree t. Mutual Information is applied to measure the amount of information contained by our target variables. The Lasso Coefficients were obtained by reducing the coefficients to zero using regularization. The Recursive Feature Elimination with Cross-Validation Support is used to identify the best number of features. Boruta Support applied along with Random Forest to identify the relevant features.

3.2. Machine Learning Techniques

Logistic Regression Applied for binary classification that predicts the probability of target variable soil fertility status shown in Equation (4)

$$P(Y=1|X)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\cdots+\beta_pX_p)}} \quad (4)$$

Decision Tree is used to split the whole dataset into subsets based on the features for prediction. The formula shown in Equation (5):

$$y=\sum_{m=1}^{M} c_m 1(x\in R_m) \quad (5)$$

where c_m is the prediction for the region
R_m defined by the tree.

Naive Bayes has been applied based on Bayes Theorem for identifying the independence between the features shown in Equation (6):

$$P(Y|X)=\frac{P(X|Y)P(Y)}{P(X)} \quad (6)$$

where $P(X|Y)$ is modeled as the product of individual feature probabilities.

Ensemble methods applied by combining multiple models such as LR+DT DT + NB LR+DT+NB to improve the overall accuracy. Here the formula is shown in Equation (7):

$$\hat{y}=\sum_{m=1}^{M} w_m\hat{y}_m(x) \quad (7)$$

where w_m is the weight assigned to the prediction of the *mth* model in the ensemble.

3.3. Sensors with Arduino Board

We used IoT sensors along with an Arduino board for the collection of soil nutrient values which shall be used as test samples for analysis.

The Arduino Board is used to connect with sensors collecting the soil nutrient data. We used the NPK sensor for collecting Nitrogen, Phosphorus, and Potassium. The pH Sensor for identifying whether the soil is alkaline or acidic, the EC Sensor to identify the ability to conduct electricity, the NIR Sensor to identify the amount of organic carbon in the soil, the Ion-Selective Electrode for Sulfur, the X-ray fluorescence (XRF) sensor and the Colorimetric RGB Sensor for Zinc, Iron, Copper, Manganese, Boron. These data are fed to the model at the server end for testing and generating the result.

3.4. Research Questions

RQ1: How do the different feature selection techniques impact the performance of the machine learning model?

RQ2: Can a hybrid model of different classifiers improve overall predictive accuracy and robustness for soil fertility prediction?

RQ3: How to generate and share client model results to the central model by maintaining privacy?

RQ4: How the central model can improvise its performance of predictions using optimization techniques?

4. Proposed Model

The proposed model shown in Figure 2 has three phases out of which Phase I comprises of local model and its training using its dataset at individual client systems. Phase II is all about applying a set of techniques to improve the accuracy of the server-end model. Further, Phase III focuses on verifying the centralized model with a sample data collection using a sensory system. Figure 1 shows a detailed explanation of the working process of each local model. It represents the training of a machine learning model using a Nutrient dataset for identifying whether the soil is fertile or not.

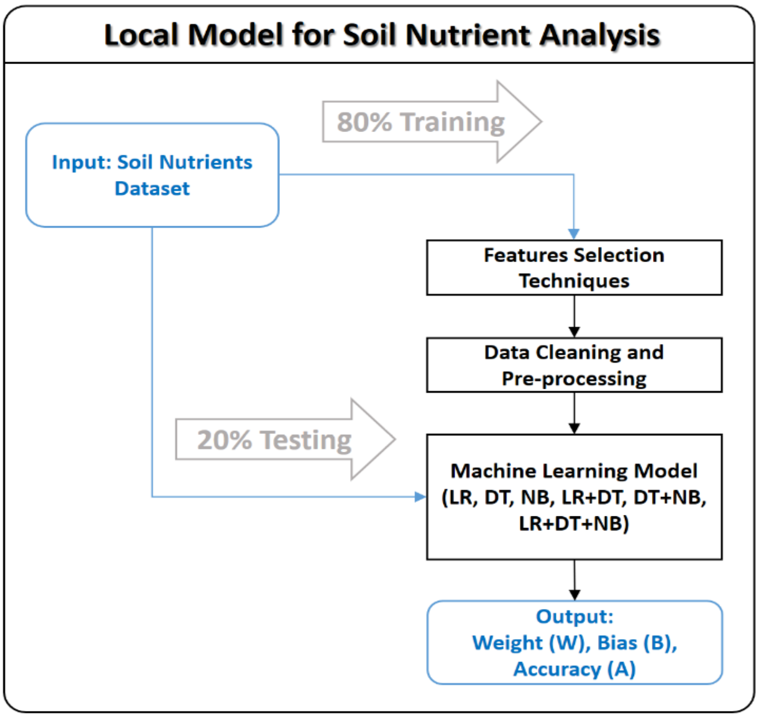


Figure 1. Proposed client model for analyzing soil nutrients dataset.

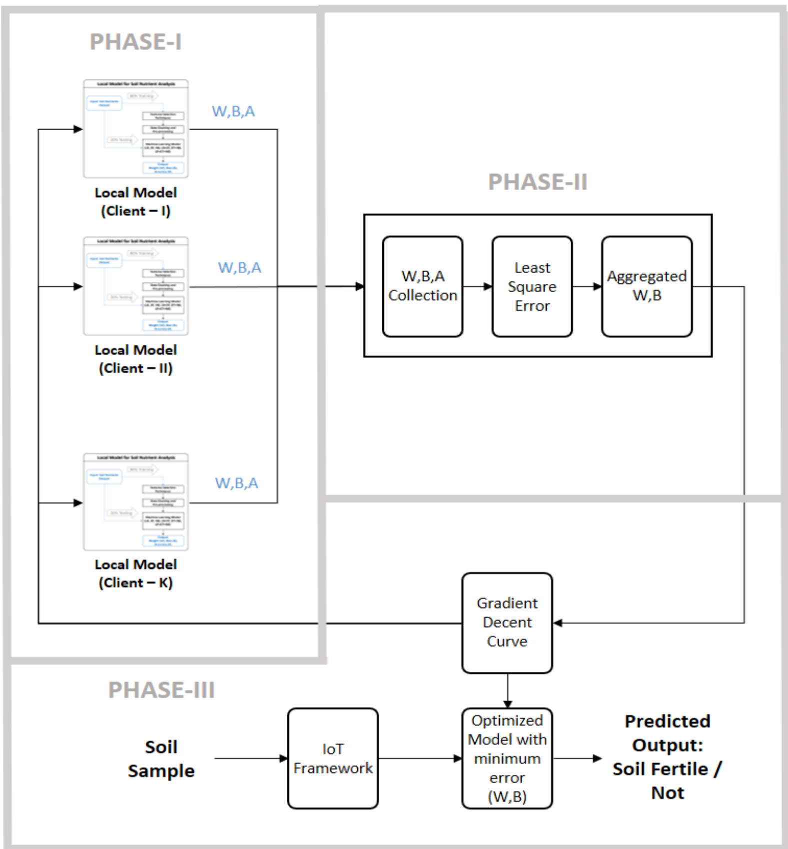


Figure 2. Proposed Centralized Federated Learning Model to combine, analyze, and aggregate the client results.

4.1. Phase-I

At the local client machine, a standard dataset for soil nutrients has been used which consists of 12 different features such as N, P, K, pH, EC, OC, S, Zn, Fe, Cu, Mn, B, and Output. The dataset consisting of 11,440 records which represents the binary classification status of soil is fertile or not in feature Output. The dataset is cleaned to duplicate and missing values, removed outliers, and noises, and normalized some features so that it will be suitable for binary classification. Then we shall apply feature selection techniques to reduce the overfitting and enhancement of accuracy. We shall find the Variance Threshold, Chi-Square, RFE Ranking, PCA Loadings, RF Importance, Mutual Information, Lasso Coefficients, RFECV Support, and Boruta Support. Further, we applied Logistic regression, Decision Tree, Naïve Bayes, and their ensemble approaches. Then we shall estimate the ranking of techniques based on accuracy, precision, recall, and F1-Score. Now based on ranking the best technique shall be selected to predict the soil fertility status with higher accuracy shown in Figure 1.

Figure 2 Phase I shows that there are multiple local models of individual clients which generates the local results of only the weight, bias, and accuracy to supply to the server-end model without supplying the actual results. So the results are kept securely at individual local machines without sharing with others.

4.2. Phase-II

It consists of a centralized model at a server that receives the weight, bias, and accuracy from each client and stores them by combining them. The least-square error method is applied to these collected details to identify the error data of each sample. Based on a minimum error and with an aggregation approach suitable weight, bias values were generated which is further verified using a gradient decent curve. Through this again it fed back to the clients' models for testing the dataset repeatedly.

4.3. Phase-III

It contains mainly soil nutrient collection from fields with the help of sensory devices. A set of sensors connected with an Arduino board collects nutrients such as Nitrogen, Phosphorus, Potassium, pH, Electrical Conductivity, Organic Carbon, Sulfur, Zinc, Iron, Copper, Manganese, and Boron. The Arduino is connected to an NPK sensor, a pH sensor, an EC sensor, an NIR sensor, an Ion-Selective Electrode sensor, and a Colorimetric RGB Sensor with a supply of 5 volts. We used these sensors and collected the soil nutrients in a systematic way which we used for testing the soil sample. There is a central model that holds the optimized weight and bias with the least error. The central model tests and analyses the soil nutrients collected and identifies its fertile state with a higher accuracy.

5. Results and Discussion

We applied initially feature selection techniques to verify their role in improving the performance of fertility analysis. The Research question and the solution during implementation are discussed in a table of ranking while analyzing different attributes.

5.1. Research Question #1

Is it possible to use the feature selection technique for soil fertility analysis? If so how do the different feature selection techniques impact the performance of the machine learning model?

Solution: yes, it is possible to use the various feature selection techniques for soil fertility analysis. Our experimental result reveals that RFECV (Recursive Feature Elimination with Cross-Validation) is the best feature because it is one of the iterative methods. This FS algorithm removes the least important features from our dataset. We have repeated this process to get the best of the best features from our dataset. Features like N, P, pH, and Cu, suggest these features are robust across multiple folds. It is the best because it does feature elimination as well as model validation.

Table 1 shows the outcome for 7 different techniques applied on 12 features where we can identify the ranking of them.

Table 1. An analysis table representing ranking for different features of the dataset.

	Variance Threshold	Chi-Square	RFE Ranking	PCA Loadings	RF Importance	Mutual Information	Lasso Coefficients	RFECV Support	Boruta Support
N	5982.230	11,447.547	9	0.107	0.550	0.494	0.005	TRUE	TRUE
P	482.034	1347.885	7	0.108	0.127	0.144	0.006	TRUE	TRUE
K	15,413.778	101.166	11	0.107	0.029	0.012	0.000	TRUE	FALSE
pH	0.216	0.223	2	0.143	0.048	0.077	0.000	TRUE	TRUE
EC	0.020	0.025	3	0.004	0.026	0.026	0.000	TRUE	FALSE
OC	0.710	1.420	10	0.151	0.024	0.000	0.000	TRUE	FALSE
S	19.551	10.896	4	0.166	0.026	0.000	0.001	TRUE	FALSE
Zn	3.584	14.134	5	0.167	0.029	0.005	0.000	TRUE	FALSE
Fe	9.661	7.822	6	0.177	0.038	0.028	0.003	TRUE	TRUE
Cu	0.217	3.891	1	0.113	0.039	0.049	0.000	TRUE	TRUE
Mn	18.459	5.506	8	0.179	0.034	0.000	0.000	TRUE	TRUE
B	0.325	1.350	12	0.120	0.030	0.032	0.073	FALSE	FALSE

Also, we tried to represent the important score in the chart in Figure 3 which describes the importance scores of different techniques. It discusses how each feature ranks according to the different feature selection techniques.

Also, the heat map generated shown in Figure 4 implements a multiple feature selection method for visualizing the ranking of different features of our dataset. It compares all the results of feature selection techniques and identifies the most important features for our predictive model. Here it shows that the soil nutrient properties are very important. Potassium is identified as a key factor for analysis, based on cross-validation the ranking of nitrogen is high. Overall by focusing on the soil properties the prediction model can be optimized.

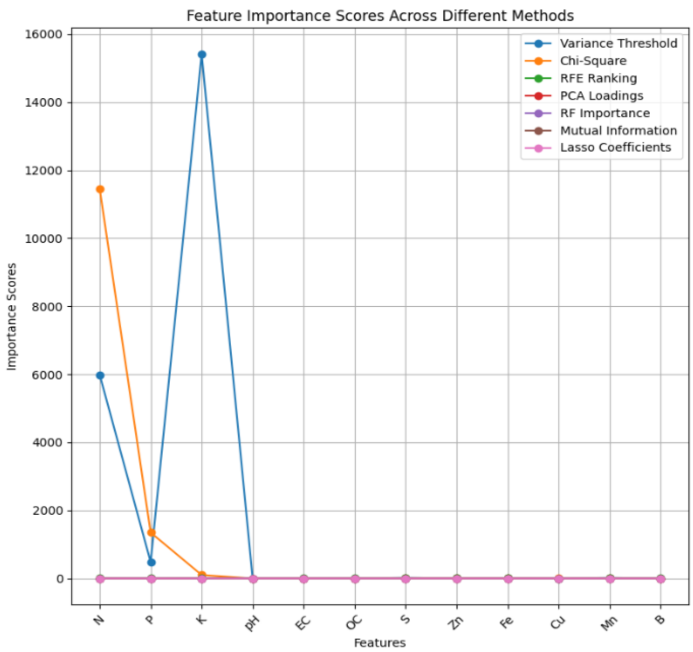


Figure 3. Feature importance chart for each technique.

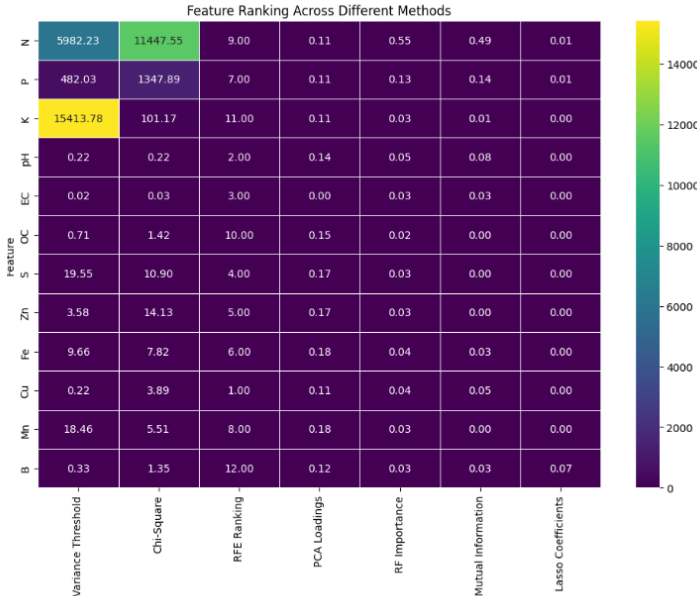


Figure 4. Feature ranking of different techniques.

In the context of machine learning, FS plays a vital role in model building and recognizing the important features of the dataset. In this paper, we use various techniques (such as Variance Threshold, Chi-Square, Recursive Feature Elimination (RFE), PCA Loadings, Random Forest (RF) Importance, Mutual Information, and Lasso Coefficients) to suggest the best model. We estimate the rank of all the features as well as how they performed on the model accuracy, precision, recall, and F1-Score. The chi-square as well as RF has high scores for some of the features as compared to other ones. The RFECV and Boruta are represented as 1 and 0 for true or false. With the help of this, we will get more consistent features for model development.

5.2. Research Question #2

Is it possible to make a hybridization model by considering the individual models for soil fertility? As well as can a hybrid model combining all these classifiers improve overall predictive accuracy and robustness in a soil fertility prediction system?

Solution: In this paper, we have taken the baseline classifier and the experimental result reveals that there is a significant difference in performance metrics when dealing with the variability of the dataset. We also consider the hybridization of models like Logistic Regression + Decision Tree, Decision Tree + Naive Bayes, and Logistic Regression + Decision Tree + Naive Bayes). The experimental result discusses that the combination of Logistic Regression + Decision Tree performs well in comparison to other combinations. The performance metrics' accuracy as well as precision increases. The ensemble model suggested one hybridization i.e., (LR + DT) to improve overall performance in soil fertility prediction.

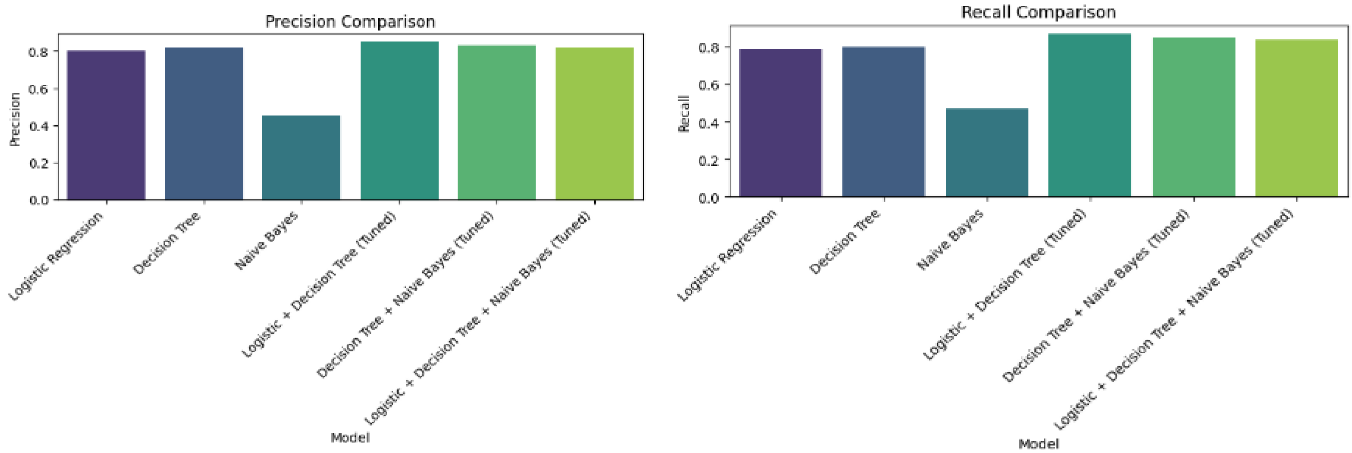
Table 2 shows the techniques of logistic regression, DT, and NB along with their ensemble approaches applied for data analysis. These methods are verified for four measuring parameters such as accuracy, precision, recall, and F1-Score so it identifies the most suitable technique.

Table 2. The performance analysis details of different algorithms and their ensemble approaches.

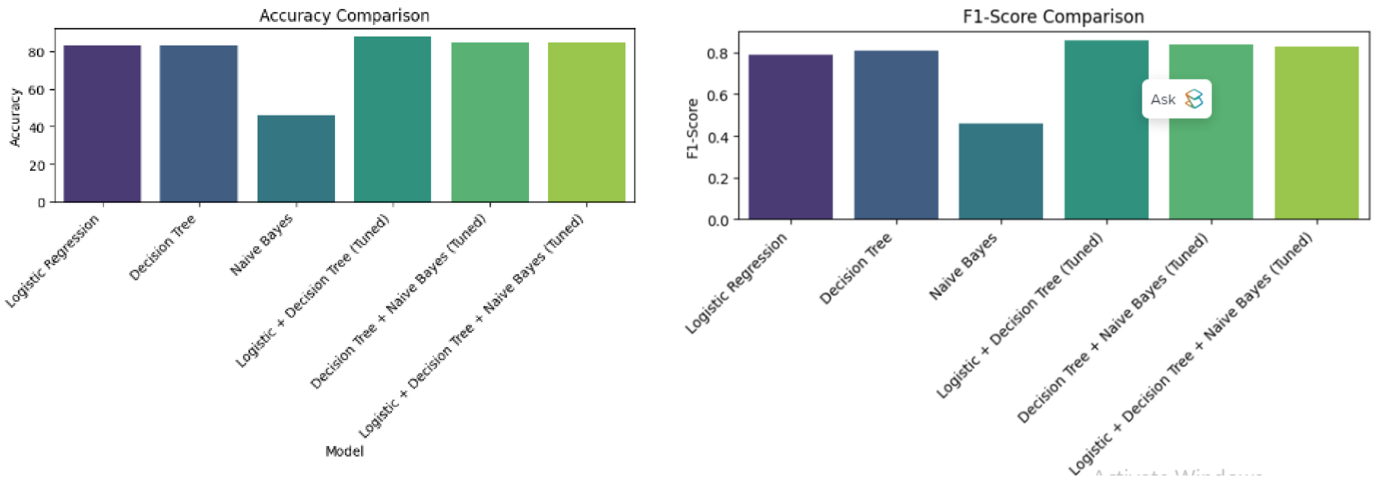
Techniques / Metrics	Logistic Regression	Decision Tree	Naive Bayes	LR + DT	DT + NB	LR + DT + NB
Accuracy	0.83	0.82	0.49	0.87	0.85	0.85
Precision	0.78	0.81	0.59	0.87	0.84	0.83

Recall	0.83	0.82	0.49	0.87	0.85	0.85
F1-Score	0.81	0.81	0.41	0.86	0.83	0.83

Figure 5 shows a comparative analysis using a Bar chart based on Precision, Recall Accuracy, and F1-Score. It is observed that the Naïve Bayes has low performance in each metric where whereas it shows the ensemble of logistic regression with the Decision Tree having higher performance.



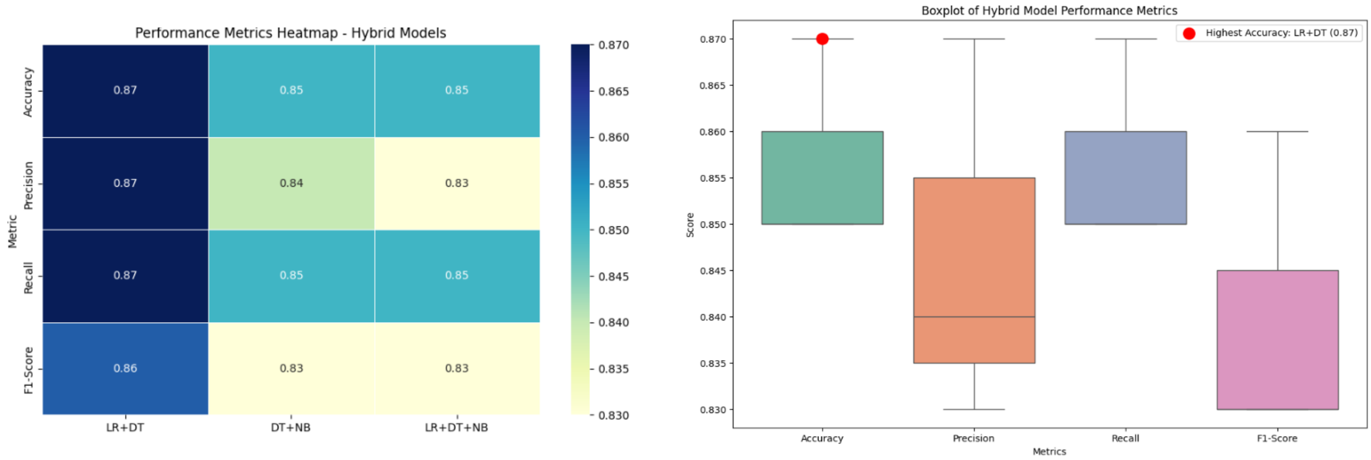
(**a**) Precision Comparison (**b**) Recall COmparison



(**c**) Accuracy Comparison (**d**) F1-Score Comparison

Figure 5. Comparison of Machine learning techniques based on Precision, Recall Accuracy and F1-Score.

Also, the performance comparison is done on the hybridization of Naïve Bayes, Decision Tree, and logistic regression using a Heat map and Box plot shown in Figure 6.

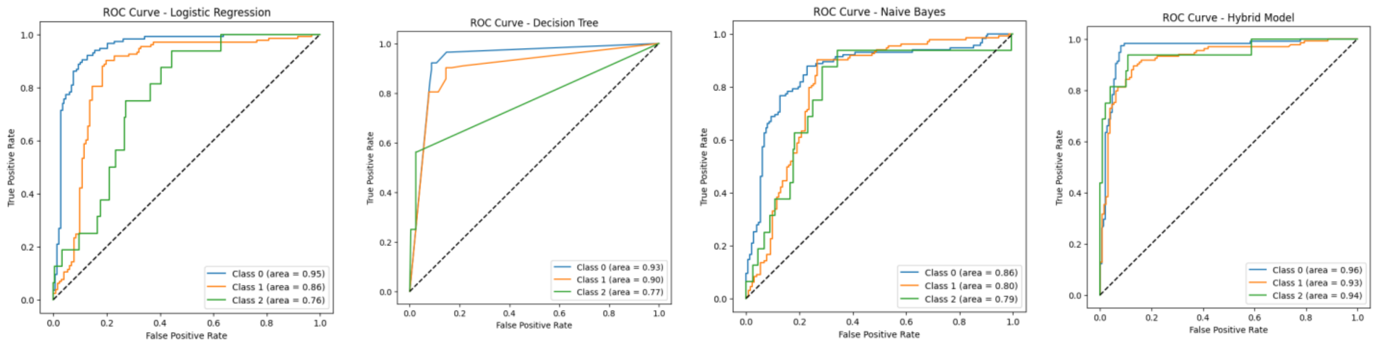


(**a**) Heatmap (**b**) Box plot

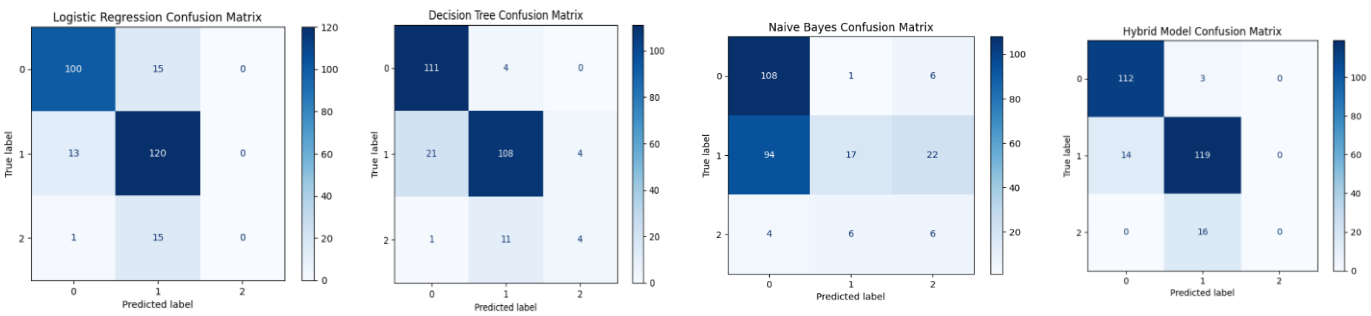
Figure 6. Heatmap and Box Plot of Hybrid models.

The boxplot provides the accuracy of having the best hybridization model. It determines which model is the best according to the accuracy of the performance metrics. The red color identifies the highest accuracy (LR + DT).

Figure 7 represents the ROC curve and confusion matrices of individual methods along with the hybrid approach.



(**a**) ROC Curve—Logistic regression (**b**) ROC Curve—Decision Tree (**c**) ROC Curve—Naïve Bayes (**d**) ROC Curve—Hybrid Model



(**e**) Logistic regression Confusion matrix (**f**) Decision Tree Confusion matrix (**g**) Naïve Bayes Confusion matrix (**h**) Hybrid Model Confusion matrix

Figure 7. ROC and confusion metrics of the individual method along with the hybrid approach.

6. Conclusions and Future Work

It is observed that a necessary approach is applied at every stage to improve the performance of the model. At the client-end model, for the selection of appropriate features 7 different selection techniques were applied, and based on calculated boruta support the feature selection was done. It allows us to have better generalization, reduces overfitting, removes highly correlated features that lead to have more reliable model, and improves accuracy.

Then the implementation of algorithms such as Naïve Bayes, Decision Tree, Logistic regression, and ensemble approaches allows us to analyze and improve the performance of the predictions. It is observed that the accuracy of the Decision tree is 82% and Logistic regression is 83% whereas by ensemble of LR + DT leads to have better accuracy of 87% in predicting soil fertility.

Again an optimized approach is incorporated at the Server-end centralized model with the least square error and gradient decent curve to make a recursive feedback to clients for improving the weight and bias values. The implementation of the recursive feedback approach is in progress which shall be the future work of contributions. Further, a sensory system was introduced at the server end to test a new sample nutrient value collected from fields.

Author Contributions: Conceptualization, M.K.S. and N.P.; methodology, A.R.; software, M.K.S.; validation, N.P., A.R. and N.P.; formal analysis, N.P.; investigation, N.P.; resources, M.K.S.; data curation, M.K.S.; writing—original draft preparation, A.R.; writing—review and editing, M.K.S.; visualization, M.K.S.; supervision, N.P.; project administration, N.P.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study is publicly available and can be accessed at the request of the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chandra, H.; Pawar, P.M.; Elakkiya, R.; Tamizharasan, P.S.; Muthalagu, R.; Panthakkan, A. *Explainable AI for Soil Fertility Prediction*; IEEE Access: Piscataway, NJ, USA, 2023.
2. Chen, G.; Cai, L.; Chen, H.; Cao, L.; Li, C. Analysis and Evaluation of Soil Fertility Status Based on Weighted K-Means Clustering Algorithm. In *Computer and Computing Technologies in Agriculture VII*; Springer: Berlin/Heidelberg, Germany, 2014; pp 89–97.
3. Kanade, P. Soil Analysis Using Machine Learning. *Br. J. Multidiscip. Adv. Stud.* **2023**, *4*, 1–11.
4. Madhumathi, R.; Arumuganathan, T.; Iyer, S.; Shruthi, R.; Shruthhi, K. Soil Nutrient Analysis Using Machine Learning Techniques. In Proceedings of the National E-Conference on Communication, Computation, Control and Automation (CCCA-2020), Coimbatore, India, 17 July 2020.
5. Folorunso, O.; Ojo, O.; Busari, M.; Adebayo, M.; Joshua, A.; Folorunso, D.; Olabanjo, O. Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review. *Big Data Cogn. Comput.* **2023**, *7*, 113.
6. Pant, J.; Pant, P.; Pant, R.P.; Bhatt, A.; Pant, D.; Juyal, A. Soil Quality Prediction for Determining Soil Fertility in Bhimtal Block of Uttarakhand (India) Using Machine Learning. *Int. J. Anal. Appl.* **2021**, *19*, 91–109.
7. BlesslinSheeba, T.; Anand, L.D.V.; Manohar, G.; Selvan, S.; Wilfred, C.B.; Muthukumar, K.; Padmavathy, S.; Ramesh Kumar, P.; Asfaw, B.T. Machine Learning Algorithm for Soil Analysis and Classification of Micronutrients in IoT-enabled Automated Farms. *J. Nanomater.* **2022**, *2022*, 5343965. https://doi.org/10.1155/2022/5343965.
8. Yadav, J.; Vivekanand Education Society's Institute of Technology, Mumbai, India; Chopra, S.; Vijayalakshmi; Vivekanand Education Society's Institute of Technology, Mumbai, India; Vivekanand Education Society's Institute of Technology, Mumbai, India. Soil Analysis and Crop Fertility Prediction Using Machine Learning. *Int. Res. J. Comput. Sci.* **2021**, *8*, 32–40. https://doi.org/10.26562/irjcs.2021.v0802.001.
9. Rajamanickam, J. Predictive Model Construction for Prediction of Soil Fertility Using Decision Tree Machine Learning Algorithm. *INFOCOMP J. Comput. Sci.* **2021**, *20*.
10. Suchithra, M.S.; Pai, M.L. Improving the Prediction Accuracy of Soil Nutrient Classification by Optimizing Extreme Learning Machine Parameters. *Inf. Process. Agric.* **2020**, *7*, 72–82. https://doi.org/10.1016/j.inpa.2019.05.003.