TuijinJishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 2 (2023)

A Study on Variable Selections and Prediction for Crop Recommender System with Soil Nutrients Using Stochastic Model and Machine Learning Approaches

S. Dhanavel[1], A. Murugan[2]

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalainagar – 608 002, Tamil Nadu, India.

[2]Assistant Professor, Department of Computer Science, Periyar Arts College, Cuddalore, (Deputed from Annamalai University, Annamalainagar) Tamil Nadu, India.

Abstract

To develop a crop recommendation system using soil nutrient data, you'll need a dataset containing details on soil nutrients and the crops that thrive in particular soil conditions. While I can't supply a specific dataset, I can offer guidance on the types of data you should seek or gather for building such a system. Machine learning, a subset of artificial intelligence (AI) and computer science, centers on harnessing data and algorithms to replicate human learning processes, steadily enhancing its precision over time. This paper considers crop recommender dataset with soil nutrients-related dataset like N, P, K, ph, EC, S, Cu, Fe, Mn, Zn, B, label. The machine learning approaches are used to analyze and predict the dataset using Logistic, Multilayer Perceptron, Simple Logistic, Hoeffding Tree, random forest, random tree, and REP tree. Numerical illustrations are provided to prove the proposed results with test statistics or accuracy parameters.

Keywords: Machine learning, crop recommender dataset with soil nutrients, decision tree, correlation coefficient, and test statistics.

1. Introduction and Literature Review

A successful crop recommendation system necessitates ongoing fine-tuning and adjustment to fit specific local circumstances. Its effectiveness is intrinsically linked to the excellence and appropriateness of the training data and the resilience of the employed machine learning models.

Data mining finds application in various domains, such as customer relationship management, fraud detection, market basket analysis, recommendation systems, medical diagnosis, and scientific research, among numerous others. Its utilization empowers organizations to make data-informed decisions, recognize trends, and unearth valuable insights from extensive datasets. Machine learning finds extensive use in diverse domains, such as natural language processing, image and speech recognition, healthcare, finance, autonomous vehicles, and more. Its versatile applications are expanding, presenting opportunities to automate processes, extract insights from data, and enhance decision-making within intricate, data-driven contexts.

A system for predicting crop yield based on historical data. We accomplish this by employing machine learning algorithms such as Support Vector Machine and Random Forest on agricultural data. Additionally, we offer recommendations for suitable fertilizers tailored to specific crop types. The primary focus of this study is the creation of a predictive model that can be applied for future crop yield forecasts. It also provides a concise analysis of crop yield prediction through machine learning techniques [1].

Machine learning is harnessed to predict the yields of four widely cultivated crops throughout India. Once the crop yield is accurately predicted for specific locations, it enables the precise application of fertilizers based on the anticipated crop and soil requirements. We utilize machine learning methods to develop a trained model that identifies patterns in data for crop prediction. The study concentrates on predicting the yields of four of the most commonly cultivated crops in India: Maize, Potatoes, Rice (Paddy), and Wheat [2].

An application of machine learning for the classification of soil into hydrologic groups. By utilizing attributes such as percentages of sand, silt, clay, and the value of saturated hydraulic conductivity, machine learning models are trained to classify soil into four hydrologic groups. The results of this classification, achieved through algorithms such as k-Nearest Neighbors, Support Vector Machine with Gaussian Kernel, Decision Trees, Classification Bagged Ensembles, and TreeBagger (Random Forest), are compared with estimation based on soil texture. The performance of these models is assessed using various metrics. Notably, k-Nearest Neighbors, Decision Trees, and TreeBagger performed better than Support Vector Machine with Gaussian Kernel and Classification Bagged Ensemble. Among the four hydrologic groups, it was observed that group B had the highest false positive rate [3].

The hypothesis that a machine learning approach enhances the accuracy of soil properties prediction. The study presents multiple research findings and a comparison of six commonly used techniques: Random Forest, Decision Tree, Naïve Bayes, Support Vector Machine, Least-Square Support Vector Machine, and Artificial Neural Network. It demonstrates that the most accurate predictions are not always achieved with the most common and complex methods. The choice of nutrient characterization category is also explored, indicating better prediction with a multi-component strategy. Additionally, the study investigates the influence of category levels and compares soil from a local farm with soil from different locations in Slovenia. Finally, the impact of principal component analysis on machine learning performance is validated using various numbers of principal components [4].

A model to assess soil fertility, the viability of sowing crop seeds, and predicting crop yields based on various soil features. Using machine learning algorithms such as Support Vector Machine (SVM), Random Forest, Naive Bayes, Linear Regression, Multilayer Perceptron (MLP), and Artificial Neural Networks (ANN), the study focuses on soil classification and crop yield prediction. Test results demonstrate that the proposed ANN method, which follows a deep learning architecture, achieves higher accuracy than existing methods [5].

Data mining is a valuable tool for uncovering previously unknown information within large existing databases. In this study, a weather dataset is used to predict whether conditions are conducive to playing golf. Seven classification algorithms, including J48, Random Tree (RT), Decision Stump (DS), Logistic Model Tree (LMT), Hoeffding Tree (HT), Reduce Error Pruning (REP), and Random Forest (RF), are employed to measure accuracy. Among these algorithms, the Random Tree algorithm stands out, achieving an accuracy of 85.714% [6].

Author suggest investigates parameters in the literature used to define soil characteristics and how they can be used as inputs for machine learning algorithms to predict soil fertility. The results indicate that prediction techniques can be efficiently applied to optimized soil parameters for more accurate soil fertility predictions with minimal human intervention [7].

The research's objectives involve conducting a comparative assessment of nutrient management strategies for major cereals, considering productivity, profitability, and nutrient use efficiencies. Various methods, including the Nutrient Expert (NE) Decision Support System, the APSIM cropping system simulation model, and machine learning (ML) approaches, are used for data analysis. The study aims to estimate potential yields and yield gaps and understand the causes of yield variability across on-farm trials in Nepal. Machine learning approaches, specifically Linear Mixed Effect models (LME) and Random Forest models (RF), are used to analyze data from the trials and make predictions [8].

A framework for predicting the absolute Crop Growth Rate (CGR) in hydroponic tomato crops using machine learning techniques. Input variables such as Electric conductivity (EC) limit, Nutrient solution (NS), ion concentration uptake, and dry weight matter of the fruits contribute to the CGR. The study explores the dynamics

of nutrient ion uptake and its impact on absolute growth, aiding in determining essential variables affecting CGR [9].

The paper proposes the use of stochastic modeling and data mining approaches to assess groundwater levels, rainfall, population, food grains, and enterprises data. It introduces a novel data assimilation analysis to predict groundwater levels effectively. The experimental results demonstrate the effectiveness of this approach [10] and [11].

The research uses chronic disease data to conduct assessments and training for five classification algorithms. The paper provides an analysis of the accuracy and performance of these algorithms, highlighting the M5P decision tree approach as the best-performing algorithm among the five tested [12].

2. Backgrounds and Methodologies

A data mining decision tree is a widely used machine learning technique for classification and regression tasks. It visually depicts a sequence of decisions and their possible outcomes in a tree-like structure. Each internal node represents a decision based on a specific feature, and each branch corresponds to the potential result of that decision. The tree's leaf nodes represent the final decision or the predicted outcome [13].

2.1 Logistic Regression

Logistic Regression is a statistical method used for binary classification, which means it's used to predict the probability of an observation belonging to one of two classes (usually labeled as 0 and 1). It's a type of regression analysis that's particularly suited for categorical outcome variables. The formula for logistic regression involves the logistic function (also known as the sigmoid function) to transform the linear combination of input features into a value between 0 and 1, representing the predicted probability of the positive class. The formula is as follows:

$$P\left(Y=\frac{1}{X}\right) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}}$$

P(Y=1/X) is the probability that the dependent variable Y is the binary outcome equal to 1 given the input features $X_1 + X_2 + \cdots + X_n$. e is the base of the natural logarithm. $\beta_0 + \beta_1 + \cdots + \beta_n$ are the coefficients that need to be estimated from the training data. $X_1 + X_2 + \cdots + X_n$. are the input features. Logistic regression is often implemented using optimization algorithms to find the best-fitting coefficients that minimize the prediction error.

2.2 Multilayer Perception

A Multilayer Perceptron (MLP) is an artificial neural network consisting of multiple layers of interconnected nodes or neurons. It's a fundamental architecture in deep learning and is used for various tasks, including classification, regression, and more complex tasks like image recognition and natural language processing. The architecture of an MLP typically includes three types of layers:

i. **Input Layer:** This layer consists of neurons receiving input data. Each neuron corresponds to a feature in the input data, and the values of these neurons pass through the network.
ii. **Hidden Layers:** These layers come after the input layer and precede the output layer. They are called "hidden" because their activations are not directly observed in the final output.
iii. **Output Layer:** This layer produces the network's final output. The number of neurons in the output layer depends on the problem type.

2.3 Hoeffding Tree

A Hoeffding Tree, also known as VFDT (Very Fast Decision Tree) or Incremental Decision Tree, is a machine learning algorithm designed for online, incremental learning on streaming data. It's beneficial when you have large volumes of data that are continuously arriving and you want to update your model in real-time without retraining the entire dataset. Here's a simplified overview of how the Hoeffding Tree algorithm works:

Step 1. **Initialization**
Step 2. **Data Arrival**

Step 3. **Splitting Nodes**
Step 4. **Leaf Node Prediction**
Step 5. **Adaptation**

2.4 Random Forest

Random Forest is a popular machine learning ensemble method for classification and regression tasks. It is an extension of decision trees and is known for its high accuracy, robustness, and ability to handle complex datasets. Random Forest is widely used in various domains, including data science, machine learning, and pattern recognition. The main idea behind Random Forest is to create an ensemble (a collection) of decision trees and combine their predictions to make more accurate and stable predictions. The following steps describe what Random Forest works like Bootstrap Aggregating (Bagging), Decision Tree Construction and Voting for Classification, Averaging for Regression. The steps involved in building a Random Forest are as follows:

Step 1. Data Bootstrapping
Step 2. Random Feature Subset Selection
Step 3. Decision Tree Construction
Step 4. Ensemble of Decision Trees
Step 5. Out-of-Bag (OOB) Evaluation
Step 6. Hyperparameter Tuning (optional)

2.5 Random Tree

In machine learning, a Random Tree is a specific type of decision tree variant that introduces randomness during construction. Random Trees are similar to traditional decision trees but differ in how they select the splitting features and thresholds at each node. Random Trees are commonly used as building blocks in ensemble methods like Random Forests. The critical characteristics of Random Trees are as follows Random Feature Subset, Random Threshold Selection, No Pruning and Ensemble Methods. Steps involved in Random Tree.

Step 1. Data Bootstrapping:
Step 2. Random Subset Selection for Features:
Step 3. Decision Tree Construction:
Step 4. Voting (Classification) or Averaging (Regression):

2.6 REP Tree

REP (Repeated Incremental Pruning to Produce Error Reduction) Tree is a machine learning algorithm for classification and regression tasks. A decision tree-based algorithm constructs a decision tree using incremental pruning and error-reduction techniques. The key steps in building a REP Tree are recursive binary splitting, pruning, and repeated pruning and error reduction. Below are the steps involved in building a REP Tree.

Step 1. Recursive Binary Splitting
Step 2. Pruning
Step 3. Repeated Pruning and Error Reduction
Step 4. Model Evaluation

2.7 Kappa statistic

The Kappa statistic, also called Cohen's Kappa or simply Kappa, is a statistical metric utilized to assess the level of agreement between two or more raters or classifiers when assigning categorical ratings or labels to items. It goes beyond considering agreement by chance alone. The Kappa statistic is represented on a scale from -1 to 1. A Kappa value of -1 signifies perfect disagreement between the raters or classifiers. A Kappa value of 0 indicates agreement that is no better than chance. A Kappa value of 1 implies perfect agreement between the raters or classifiers. The calculation of Kappa employs the formula:

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} \quad \ldots(1)$$

$$P_o = \frac{\text{Number of items with agreement}}{\text{Total number of items}}$$

$$P_e = \sum \frac{\text{Total count in row} \times \text{Total count in column}}{\text{Total number of items}}$$

Where, P_o denotes the observed agreement, i.e., the proportion of items on which raters or classifiers agree. P_e represents the expected agreement, i.e., the agreement expected by chance.

2.8 Mean Absolute Error

Mean Absolute Error (MAE) is a metric used to measure the average absolute difference between predicted and actual (true) values in a regression problem. It is commonly used to assess the accuracy of a regression model's predictions [14]. The formula to calculate Mean Absolute Error (MAE) is as follows:

MAE = Σ |(Actual Value - Predicted Value)| / n ... (2)

Where:

Σ represents the summation symbol, which sums up the values for all data points, | | denotes the absolute value, ensuring the differences are positive. In this formula, Actual Value: Refers to the true value of the target variable (ground truth) for a specific data point. Predicted Value: Refers to the value predicted by the regression model for the same data point. n: Represents the total number of data points in the dataset.

2.9 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is a commonly used metric to assess the accuracy of a regression model's predictions. It measures the average magnitude of the errors between the predicted and actual (true) values, considering both the direction and magnitude of the errors. The formula to calculate Root Mean Squared Error (RMSE) is as follows [15]:

RMSE = √(Σ (Actual Value - Predicted Value)² / n) ... (3)

Where:

❖ Σ represents the summation symbol, which sums up the values for all data points. (Actual Value - Predicted Value) ² denotes the squared difference between each data point's actual and predicted values. n is the total number of data points in the dataset.

2.10 Relative Absolute Error (RAE)

Relative Absolute Error (RAE), also known as Mean Absolute Percentage Error (MAPE), is a metric used to evaluate the accuracy of predictions in regression tasks. It measures the average percentage difference between the absolute and actual (valid) values, providing a relative measure of the prediction errors [16]. The formula to calculate Relative Absolute Error (RAE) is as follows:

RAE = (Σ |Actual Value - Predicted Value| / Σ |Actual Value|) * (100 / n) ... (4)

Where:

Σ represents the summation symbol, which sums up the values for all data points. | | denotes the absolute value, ensuring the differences are positive. n is the total number of data points in the dataset.

2.11 Root Relative Squared Error (RRSE)

"Root Relative Squared Error" is not a standard or widely recognized metric in statistics or machine learning. It appears to be a combination of the terms "Root Mean Squared Error (RMSE)" and "Relative Absolute Error

(RAE)." It's possible that the time was created or used in a specific context or literature, but it is not a commonly used or established metric. For clarity, let's briefly define the two individual metrics mentioned. The formula to calculate RMSE is:

$$\mathbf{RMSE = \sqrt{(\Sigma\ (Actual\ Value - Predicted\ Value)^2 / n)}}$$

Relative Absolute Error (RAE): Also known as Mean Absolute Percentage Error (MAPE), RAE measures the average percentage difference between the absolute errors and the actual (true) values, providing a relative measure of the prediction errors. The formula to calculate RAE is:

$$\mathbf{RAE = (\Sigma\ |Actual\ Value - Predicted\ Value| / \Sigma\ |Actual\ Value|) * (100 / n)}$$

As there is no established metric called "Root Relative Squared Error," it's crucial to use standard evaluation metrics such as RMSE, RAE (MAPE), or others that are well-known and have clear interpretations in the context of your specific problem.

Numerical Illustrations

The corresponding dataset was collected from the open-source Kaggle data repository. The crop recommender and soil nutrients dataset includes 12 parameters with data categories like N, P, K, ph, EC, S, Cu, Fe, Mn, Zn, B, and label [17]. A detailed description of the parameters is mentioned in the following Table 1.

Table 1. Crop recommender dataset with soil nutrients (sample dataset)

N	P	K	ph	EC	S	Cu	Fe	Mn	Zn	B	label
143	69	217	5.90	0.58	0.23	10.20	116.35	59.96	54.85	21.29	pomegranate
170	36	216	5.90	0.15	0.28	15.69	114.20	56.87	31.28	28.62	pomegranate
158	66	219	6.80	0.34	0.20	15.29	65.87	51.81	57.12	27.59	pomegranate
133	45	207	6.40	0.94	0.21	8.48	103.10	43.81	68.50	47.29	Pomegranate
98	85	191	6.00	1.45	0.28	14.47	179.63	85.79	47.76	65.83	mango
95	87	141	4.80	0.88	0.27	18.63	71.11	64.77	20.86	58.64	mango
107	79	127	5.70	0.75	0.25	9.46	80.98	89.12	29.14	60.91	mango
123	74	134	6.00	0.83	0.31	19.90	140.22	99.24	23.29	70.52	mango
174	79	300	7.70	1.65	0.02	12.23	163.19	65.96	18.22	7.49	ragi
147	96	346	6.30	0.85	0.01	15.11	206.88	60.28	20.08	4.32	ragi
164	96	242	6.20	0.78	0.02	18.33	76.05	55.59	18.42	6.88	ragi
179	86	301	7.70	0.84	0.02	13.76	248.54	59.01	26.88	8.39	ragi
103	35	69	5.60	1.47	0.12	19.00	38.50	184.24	37.92	11.01	potato
72	32	196	6.00	1.92	0.09	29.00	40.40	116.88	23.33	13.25	potato
177	42	171	5.70	2.23	0.11	17.00	39.70	254.08	41.96	16.10	potato
83	31	173	6.00	2.32	0.10	14.00	41.60	147.19	42.60	17.61	potato
165	57	53	5.10	1.61	0.11	16.00	41.50	175.07	48.55	19.36	potato

Table 2: Machine Learning Models with Correctly and Incorrectly Classified Instances

Function and Trees	Correctly Classified Instances	Incorrectly Classified Instances
Logistic	591.0000	29.0000
Multilayer Perceptron	590.0000	30.0000
Simple Logistic	600.0000	20.0000
Hoeffding Tree	601.0000	19.0000
Random Forest	600.0000	20.0000
Random Tree	576.0000	44.0000
REP Tree	595.0000	25.0000

Table 3: Machine Learning Models with Correctly and Incorrectly Classified Instances (%)

Function and Trees	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)
Logistic	95.3226	4.6774
Multilayer Perceptron	95.1613	4.8387
Simple Logistic	96.7742	3.2258
Hoeffding Tree	96.9355	3.0645
Random Forest	96.7742	3.2258
Random Tree	92.9032	7.0968
REP Tree	95.9677	4.0323

Table 4: Machine Learning Models with Kappa statistic

Function and Trees	Kappa statistic
Logistic	0.9439
Multilayer Perceptron	0.9419
Simple Logistic	0.9613
Hoeffding Tree	0.9632
Random Forest	0.9613
Random Tree	0.9148
REP Tree	0.9516

Table 5: Machine Learning Models with MAE and RMSE

Function and Trees	MAE	RMSE
Logistic	0.0261	0.1188
Multilayer Perceptron	0.0226	0.1174
Simple Logistic	0.0365	0.1151
Hoeffding Tree	0.0106	0.0993
Random Forest	0.0255	0.1055
Random Tree	0.0237	0.1538
REP Tree	0.0228	0.1145

Table 6: Machine Learning Models with RAE and RRSError (%)

Function and Trees	RAE (%)	RRSE (%)
Logistic	9.3928	31.8846
Multilayer Perceptron	8.1319	31.5085
Simple Logistic	13.1465	30.8742
Hoeffding Tree	3.8121	26.6572
Random Forest	9.1955	28.3209
Random Tree	8.5161	41.2692
REP Tree	8.1920	30.7324

Table 7: Machine Learning Models with Time Taken to Build Model (Seconds)

Function and Trees	Time taken (seconds)
Logistic	0.7600
Multilayer Perceptron	1.7100
Simple Logistic	0.8200
Hoeffding Tree	0.1000
Random Forest	0.4700
Random Tree	0.0200
REP Tree	0.0300

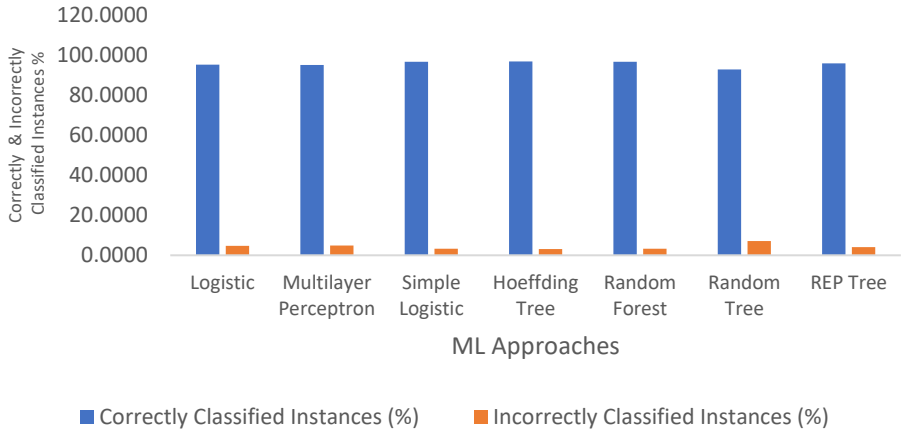

Figure 1. Machine Learning Models with Correctly Classified Instances (%) and Incorrectly Classified Instances (%)

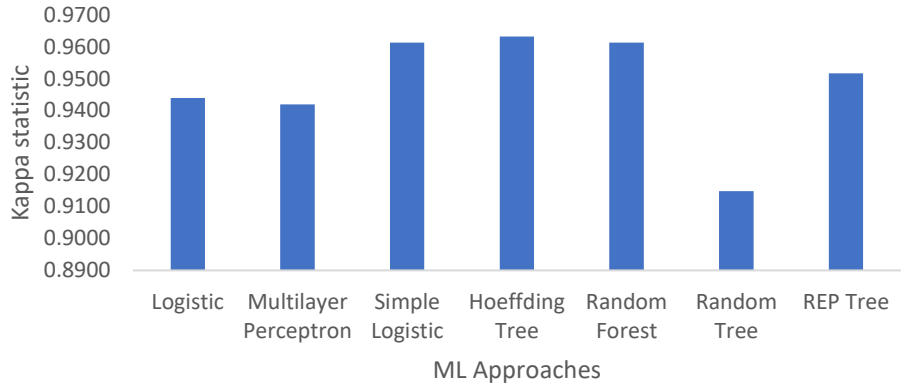

Figure 2. Machine Learning Models with Kappa statistic

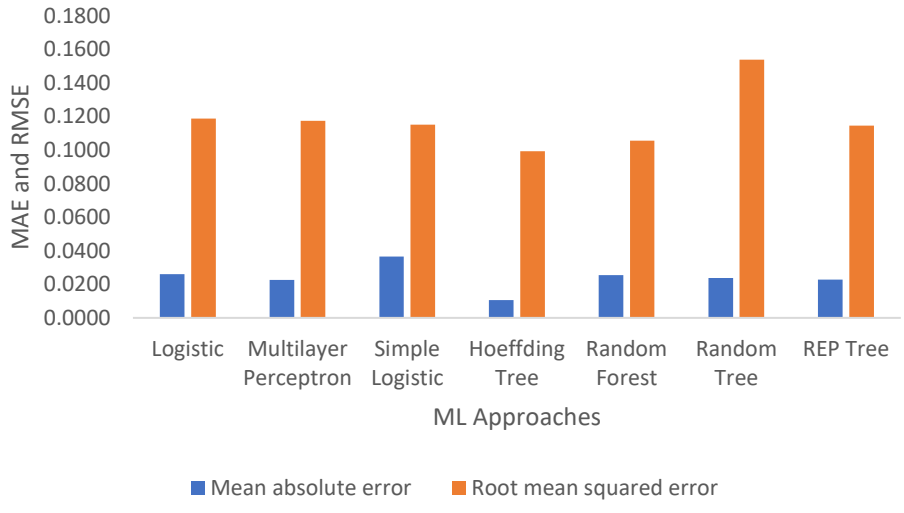

Figure 3. Machine Learning Models with MAE and RMSE

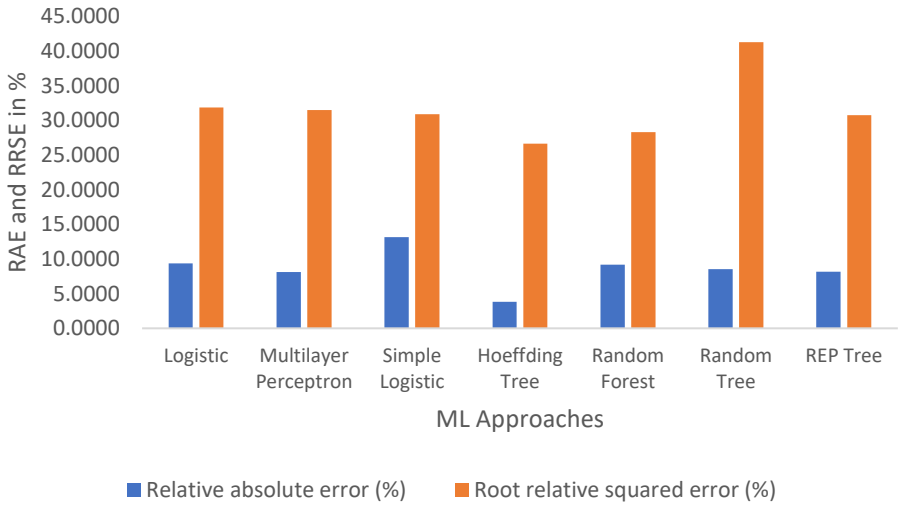

Figure 4. Machine Learning Models with RAE (%) and RRSE (%)

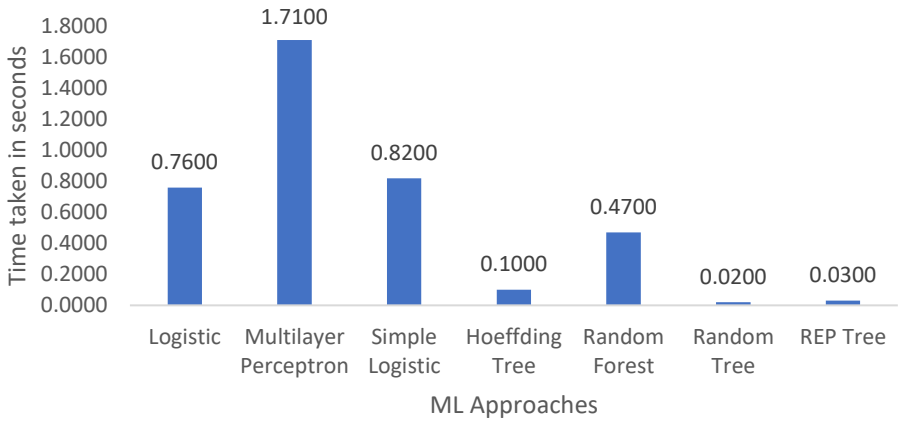

Figure 5. Machine Learning Models and its Time Taken to Build the Model (Seconds)

3. Result and Discussion

Table 1 explains 12 parameters encompassing various data categories, including Soil micronutrients, macronutrients, and information on agricultural products. The analysis of this dataset reveals the utilization of seven distinct machine learning approaches - Logistic, Multilayer Perceptron, Simple Logistic, Hoeffding Tree, random forest, random tree, and REP tree. These approaches are employed to uncover hidden patterns and determine the most influential parameter for future predictions. Comprehensive results and numerical representations are presented across Table 1 to Table 7 and Figure 1 to Figure 5.

Table 2 delves into the quality assessment of the data, distinguishing correctly and incorrectly classified instances. Similarly, Table 3 highlights the percentage of correctly and incorrectly classified instances. The outcomes of these assessments are depicted in Figure 1.

The basis for these evaluations lies in Equation 1 and references to Table 4 and Figure 2. Equation 1 is employed to calculate kappa statistics, a measure of inter-rater agreement or reliability, often used to assess the consistency of ratings or classifications among multiple observers. Notably, among the results, it is observed that the random forest method yields the lowest kappa value when using the seven machine learning approaches. The other

approaches demonstrate kappa values close to 0.95, indicating superior reliability. Visual representations are included in Figure 3 to support these findings.

Moving forward, the Mean Absolute Error (MAE), defined by Equation 2, is utilized to assess model errors, leveraging seven different machine learning algorithms. All seven approaches exhibit exceptional error performance, with MAE values approaching 0. The Root Mean Square Error (RMSE), as described in Equation 3, measures the disparity between predicted and actual values. Like MAE, all machine learning approaches deliver a commendable error performance, with RMSE values close to 0. Corresponding numerical data is displayed in Table 5 and Figure 3.

To evaluate the accuracy, the Relative Absolute Error (RAE), as per Equation 4, is employed to compare predicted and actual values in percentage terms. Seven ML classification algorithms are considered in this context. Notably, logistic regression returns the highest error rate, while the remaining six ML approaches exhibit minimal error. This trend is reflected in Relative Root Mean Square Error (RRSE) as well, with analogous numerical representations provided in Table 6 and Figure 4.

Time efficiency is a critical factor in machine learning approaches. The data presented in Table 7 and Figure 5 indicate that Multilayer Perceptron require the maximum time for problem-solving. Conversely, Random tree, REP tree, Hoeffding tree, and Random Forests are the most time-efficient for model development. Furthermore, logistic regression demonstrates a minimal time requirement for model generation. Similar trends are observed in the visualizations.

4. Conclusion and further research

Considering the constraints of the model, it is important to acknowledge potential biases in the dataset, as well as factors specific to machine learning algorithms that may contribute to variations in performance. Additionally, computational constraints might have influenced the model development. To enhance this research, it would be beneficial to explore additional data sources to validate and augment the findings. Investigating more advanced algorithms and fine-tuning hyperparameters could further improve model performance. Additionally, addressing potential dataset biases and working to reduce computational constraints would be pivotal for refining the model. This research holds significant potential for the Department of Agriculture and other stakeholders aiming to optimize the agriculture sector through micro and macronutrient level insights.

Reference

1. Bondre, D.A. and Mahagaonkar, S., 2019. Prediction of crop yield and fertilizer recommendation using machine learning algorithms. International Journal of Engineering Applied Sciences and Technology, 4(5), pp.371-376.
2. Pant, J., Pant, R.P., Singh, M.K., Singh, D.P. and Pant, H., 2021. Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings, 46, pp.10922-10926.
3. Abraham, S., Huynh, C. and Vu, H., 2019. Classification of soils into hydrologic groups using machine learning. Data, 5(1), p.2.
4. Trontelj ml, J. and Chambers, O., 2021. Machine Learning Strategy for Soil Nutrients Prediction Using Spectroscopic Method. Sensors, 21(12), p.4208.
5. Yadav, J., Chopra, S. and Vijayalakshmi, M., 2021. Soil analysis and crop fertility prediction using machine learning. Machine Learning, 8(03).
6. Rajesh, P. and Karthikeyan, M., 2017. A comparative study of data mining algorithms for decision tree approaches using the Weka tool. Advances in Natural and Applied Sciences, 11(9), pp.230-243.
7. Rose, S., Nickolas, S. and Sangeetha, S., 2018, August. Machine Learning and Statistical Approaches used in Estimating Parameters that Affect the Soil Fertility Status: A Survey. In 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT) (pp. 381-385). IEEE.
8. Timsina, J., Dutta, S., Devkota, K.P., Chakraborty, S., Neupane, R.K., Bishta, S., Amgain, L.P., Singh, V.K., Islam, S. and Majumdar, K., 2021. Improved nutrient management in cereals using Nutrient Expert and

machine learning tools: productivity, profitability and nutrient use efficiency. Agricultural Systems, 192, p.103181.

9. Verma, M.S. and Gawade, S.D., 2021, March. A machine learning approach for prediction system and analysis of nutrients uptake for better crop growth in the Hydroponics system. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 150-156). IEEE.
10. Rajesh, P., Karthikeyan, M. and Arulpavai, R., 2019, December. Data mining approaches to predict the factors that affect the groundwater level using a stochastic model. In AIP Conference Proceedings (Vol. 2177, No. 1). AIP Publishing.
11. Rajesh, P. and Karthikeyan, M., 2019. Data mining approaches to predict the factors that affect agriculture growth using stochastic models. International Journal of Computer Sciences and Engineering, 7(4), pp.18-23.
12. Rajesh, P., Karthikeyan, M., Santhosh Kumar, B. and Mohamed Parvees, M.Y., 2019. Comparative study of decision tree approaches in data mining using chronic disease indicators (CDI) data. Journal of Computational and Theoretical Nanoscience, 16(4), pp.1472-1477.
13. Kohavi, R., & Sahami, M. 1996. Error-based pruning of decision trees. In International Conference on Machine Learning, pp. 278-286.
14. Akusok, A. 2020. What is Mean Absolute Error (MAE)? Retrieved from https://machinelearningmastery.com/mean-absolute-error-mae-for-machine-learning/
15. S. M. Hosseini, S. M. Hosseini, and M. R. Mehrabian, 2019. Root mean square error (RMSE): A comprehensive review, International Journal of Applied Mathematics and Statistics, vol. 59, no. 1, pp. 42–49,.
16. Chi, W. (2020). Relative Absolute Error (RAE) – Definition and Examples. Medium. https://medium.com/@wchi/relative-absolute-error-rae-definition-and-examples-e37a24c1b566
17. https://www.kaggle.com/datasets/manikantasanjayv/crop-recommender-dataset-with-soil-nutrients