Eco. Env. & Cons. 30 (August Suppl. Issue) : 2024; pp. (S458-S463)

ISSN 0971–765X

DOI No.: http://doi.org/10.53550/EEC.2024.v30i05s.071

Precision Agriculture: Forecasting Plant Nutrient Requirements with Machine Learning

J. Porkodi[1], B. Karunai Selvi[2] and A. Nagavaratharajan[3]

[1]Department of Computer Applications (UG), V.V.Vanniaperumal College for Women, Virudhunagar, T.N., India
[2]Department of Botany, V.V.Vanniaperumal College for Women, Virudhunagar, T.N., India
[3]Department of Commerce with Computer Applications, V.H.N. Sentikumara Nadar College, Virudhunagar, T.N., India

(Received 20 January, 2024; Accepted 28 March, 2024)

ABSTRACT

Fertilizers are indispensable for modern agriculture, providing vital nutrients crucial for crop growth, yield, and nutritional quality. They optimize soil fertility, alleviate nutrient deficiencies, and promote sustainable farming practices. This study explores the prediction of crop fertilizer needs using machine learning algorithms. The dataset consist of the following features such as Farmer Name, Father_or_Husband Name, Survey Number, Block Village, PH, EC, Organic Carbon (%), Nitrogen (Kg/ha), Phosphorous (Kg/ha), Potassium (kg/ha), Sulphur (ppm), Zinc (ppm), Boron (ppm), Iron (ppm), Manganese (ppm), Copper (ppm) and Lime Status. These are the independent variables and the fertilizer is the dependent variable to be predicted. By assessing farmer details, soil attributes, and nutrient levels, Logistic Regression emerges as the most precise model to predict which major nutrient is required for the crop. Logistic Regression Algorithm, boasting an impressive 97% accuracy rate of prediction. Through meticulous analysis of classification reports, Logistic Regression proves superior among the algorithms considered, surpassing XG Boost, Decision Tree, and Support Vector Machine, which achieved accuracies of 92.7%, 90.5%, and 94.2% respectively. Hence machine learning technique proves to be an effective method for prediction of soil nutrient and as well as its analysis. This finding holds immense significance for agricultural practices. Precise fertilizer prediction optimizes resource allocation, boosts crop yields, and minimizes environmental harm. Harnessing machine learning empowers farmers to make informed decisions, tailoring fertilizer application to individual crop and soil requirements.

Key words : *Alleviate nutrient deficiencies, Nutrient prediction, Logistic Regression, XG Boost Classifier, Decision Tree Classifier, Support Vector Machine*

Introduction

Agriculture, encompassing the cultivation of plants and raising of livestock, stands as a cornerstone of global economies, playing a pivotal role in sustaining human societies. It serves as a fundamental pillar for ensuring food security, which entails providing safe and nutritious sustenance to all individuals at all times. Agriculture not only furnishes the nourishment necessary for human well-being but also generates employment, income, and foreign exchange for numerous nations. Fertilization emerges as a critical aspect of agricultural practices, facilitating the provision of essential nutrients for optimal

([1]Head and Assistant Prof. [2]Head and Associate Prof., [3]Associate Prof.)

plant growth and productivity. However, conventional fertilization approaches often prove inefficient and wasteful, frequently resulting in the over application of fertilizers beyond actual requirements. This not only poses risks of environmental contamination but also leads to financial setbacks for farmers.

Machine learning represents a swiftly advancing technology poised to transform fertilization strategies. By leveraging vast datasets encompassing soil analyses, meteorological information, and crop productivity data, machine learning algorithms can discern intricate patterns and correlations that traditional methods might overlook. When integrated with nutrient forecasting, machine learning enables the creation of exceptionally accurate fertilization guidelines. Such advancements can aid farmers in minimizing fertilizer usage, enhancing crop yields, and safeguarding environmental integrity.Top of Form

Experimental Methods or Methodology

In the area of machine learning (ML), algorithms typically engage in the iterative process of learning from experience or examples, commonly referred to as training data. The objective is to optimize a specific task, and when the learning outcomes align with predefined mathematical and statistical performance metrics, these ML algorithms can be employed to make accurate predictions for new data, often referred to as test data. To facilitate this learning process, the raw data and information collected undergo a thorough pre-processing phase, wherein efforts are dedicated to tasks such as data cleaning, integration, and conversion to ensure its suitability before being fed into the trained model.

In the context of predicting fertilizer requirements, the collected raw data and information related to soil composition, crop characteristics, and environmental factors would similarly undergo a meticulous pre-processing phase. This phase involves tasks such as data cleaning, integration of diverse datasets, and conversion into a format suitable for analysis. The trained ML model, specifically tailored for fertilizer prediction, is then fed with these organized training samples. Five commonly employed ML algorithms are applied in this study to forecast optimal fertilizer requirements for enhancing the growth of crops.

This article addresses the importance of correct amount of fertilizer for the better growth of plants and to improve the agriculture.

Logistic Regression

Logistic regression is a statistical method commonly employed for binary classification tasks, where the outcome variable has two possible categories. The logistic regression model transforms the linear combination of input features using the logistic function, also known as the sigmoid function. The logistic function is represented as:

$$P(Y=1)=1/(1+e-(b0+b1X1+b2X2+...+bnXn))$$

Here, $P(Y=1)$ denotes the probability of the dependent variable being 1, e is the base of the natural logarithm, and $b0$, $b1$,..., bn are the coefficients associated with the input features $X1$, $X2$,..., Xn. The logistic function maps the linear combination to a probability value between 0 and 1, allowing the model to predict the likelihood of an instance belonging to a particular class. During the training process, the logistic regression model aims to adjust the coefficients to minimize the difference between the predicted probabilities and the actual class labels. This adjustment is typically done through techniques like Maximum Likelihood Estimation (MLE), which seeks to maximize the likelihood of observing the given outcomes given the input features. The resulting decision boundary, determined by the learned coefficients, divides the input space into regions corresponding to the two classes, enabling the model to make predictions on new data based on this learned relationship. Logistic regression finds wide application in fields such as healthcare, finance, and social sciences for tasks involving binary classification.

Decision Tree Classifier

In the domain of machine learning, Decision Trees

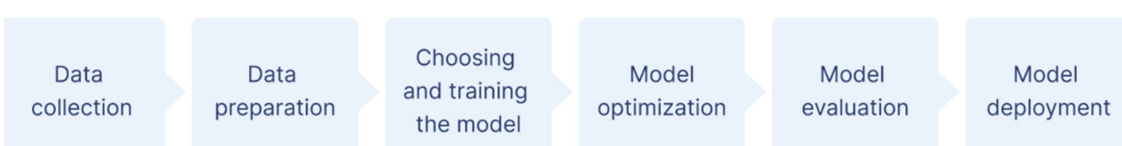

Fig. 1. Machine Learning process flow

(DT) function as predictive models, establishing a mapping relationship between object attributes and their corresponding values. An instance of DT, known as gradient boosting DT, is a supervised machine learning model renowned for its interpretability and swift classification. Pioneering DT algorithms include ID3 and C4.5, developed by Quinlan in 1986 and 1994, respectively, alongside Classification and Regression Trees (CART) proposed by Breiman *et al.*, 1984. Unlike models with specific equations encapsulating datasets, DT construction comprises three key steps: the creation of nodes (root, internal, and leaf nodes) based on features, minimization of sample splits, and addressing impurity. The quality of the DT's fit to training samples improves with decreasing impurity, and two commonly employed impurity indices are cross-entropy and Gini. These indices are computed as follows: " Formulas for cross-entropy and Gini here".

$$Entropy = -\sum_{i}(p_i)\log_2(p_i)$$

$$Gini = 1 - \sum_{i=0}^{n-1}(p_i)^2$$

Here pi denotes the proportion of samples in class i to the total number of samples.

XG Boost Classifier

"XGBoost" stands for "Extreme Gradient Boosting," and it is a powerful machine learning algorithm. XG Boost is an advanced implementation of the gradient boosting algorithm that offers high performance, efficiency, and flexibility. It is commonly used for classification tasks.

XG Boost is based on the gradient boosting framework, which involves combining weak learners (typically decision trees) into a strong predictive model. XG Boost includes regularization techniques to prevent overfitting, such as L1 (Lasso) and L2 (Ridge) regularization on the leaf weights.

Support Vector Machine

Support Vector Machines (SVM) belong to the class of generalized linear classifiers utilized for binary classification through supervised learning. The decision boundary in SVM is determined by the maximum margin hyperplane concerning the learning samples. SVM is versatile, performing tasks like regression and classification. It employs the kernel trick to initially map input data into a high-dimensional feature space. Subsequently, an insensitive loss is applied for linear regression in the feature space, minimizing the empirical risk and reducing model complexity. This approach facilitates SVM's capacity to handle nonlinear classification problems using methods originally designed for linear classifiers, presenting a notable advantage. The function of the linear classifier is expressed as:

$$f(x) = W^T X + b$$

Here W is the weight vector, b is the bias. The function of the nonlinear boundaries generated by the introduction of the kernel is:

$$f(x) = \sum_{i=1}^{n} k(X, Xi) + b$$

Results and Discussion

Dataset

The dataset comprises various features including Farmer Name, Father_OR_Husband Name, SurveyNumber, BlockVillage, pH, EC (Electrical Conductivity), Organic Carbon percentage, Nitrogen content (kg/ha), Phosphorus content (kg/ha), Potassium content (kg/ha), Sulphur concentration (parts per million - ppm), Zinc concentration (ppm), Boron concentration (ppm), Iron concentration (ppm), Manganese concentration (ppm), Copper concentration (ppm), Lime Status, and Fertilizer. These features serve as independent variables, while the type of fertilizer required is the dependent variable to be predicted.

The goal is to classify the fertilizers into three distinct categories based on the aforementioned features. This classification task involves utilizing machine learning algorithms to analyze the dataset and discern patterns among the features that correlate with different types of fertilizers. By training the model on the dataset, it learns to recognize relationships between the independent variables (such as soil characteristics, nutrient levels, and lime status) and the dependent variable (type of fertilizer).

Through this process, the machine learning model can predict the appropriate category of fertilizer needed for a given set of input features. This predictive capability offers valuable insights to farmers, enabling them to make informed decisions about fertilization strategies tailored to specific soil conditions and crop requirements. Ultimately, the aim is to optimize fertilizer usage, maximize crop yields, and promote sustainable agricultural practices.

Fig. 1. Depicts the count of 3 fertilizers in the dataset which are Urea, Super Phosphate and Potash. Thus this histogram represents the type of fertilizer in the x-axis and its count in y-axis.

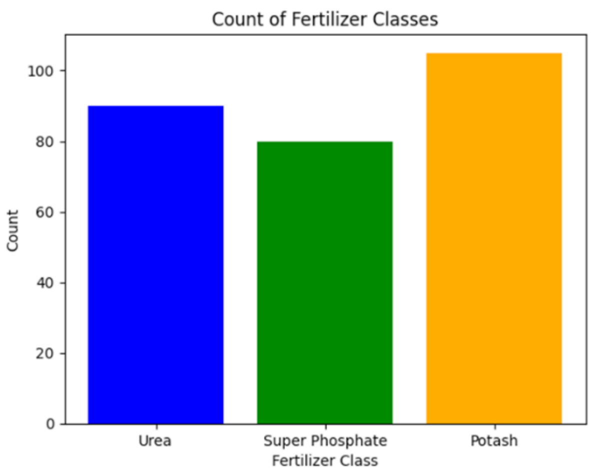

Fig. 1. Histogram for Count of Fertilizer Class

Fig. 2. This graph depicts the distribution of pH level in the fertilizers. The x-axis represents the pH and y-axis represents the count.

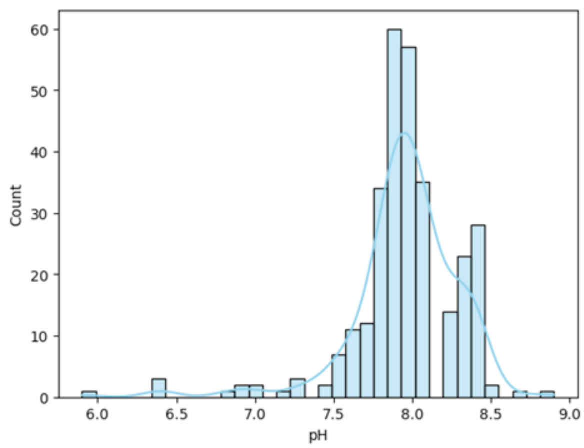

Fig. 2. Distribution of pH

Fig. 3. This is the correlation Heat Map of the various attributes like pH, Organic Carbon, Nitrogen, phosphorous, potassium, Sulphur, Zinc, Boron, Iron, Managenese and Copper content.

Performance Metrics

Evaluation metrics such as the root mean squared

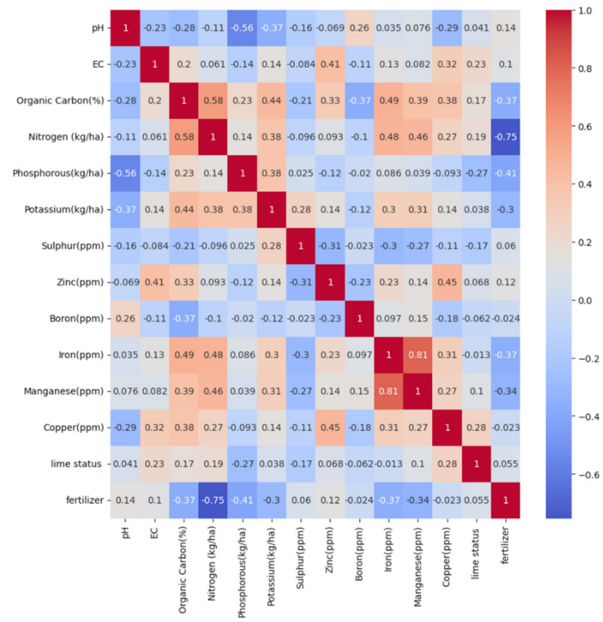

Fig. 3. Correlation of Heat Map

error (RMSE), mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R2) were utilized to gauge and compare the accuracy of distinct models. To validate the predictive performance of the top-performing model, assessments were extended to datasets drawn from external sources beyond the original data pool.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(yi - yi')^2}{N}}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|yi - yi'|$$

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

$$R^2 = \sqrt{\frac{\sum_{i=1}^{N}(yi - \bar{y})^2 - \sum_{i=1}^{N}(yi - yi')^2}{\sum_{i=1}^{N}(yi - \bar{y})^2}}$$

Where N is the sample size, yi is the measured value, yi2is the predictedvalue, y is the mean value of yi.

Fig. 4. This bar graph represents the accuracy of various models like Logistic Regression, XG Boost

Classifier, Decision Tree Classifier and Support Vector Machine .The x-axis represents the various machine learning models and y-axis represents the percentage of accuracy.

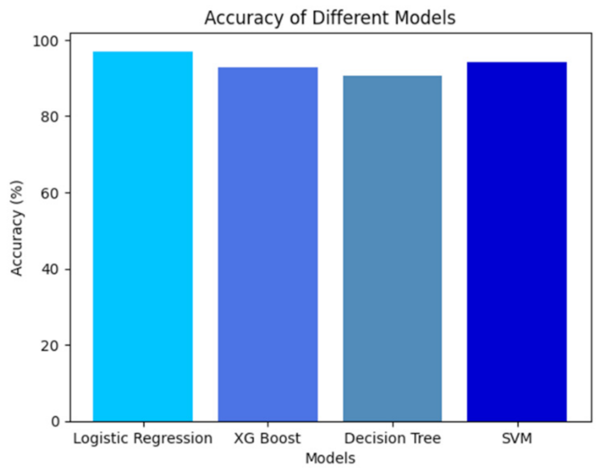

Fig. 4. Accuracy of various models

Table 1 and Table 2 represents the performance metrics of the training dataset and the testing dataset.The Logistic Regression gives the accuracy of 97%, XGBoost gives the accuracy of 92.7% , Decision Tree gives 90.5% and finally the Support Vector Machine gives the accuracy as 94.2%. The classification report is analysed and the performance metrics for each machine learning algorithm is compared and finally for this dataset Logistic algorithm performs well compared to other machine learning algorithms.

Conclusion

Previous research and findings underscore the pivotal role of fertilizer quantity in facilitating optimal plant growth. The integration of machine learning-driven nutrient prediction into fertilization practices holds immense potential to transform agriculture and overcome challenges associated with traditional methods. The fusion of advanced data analytics with agricultural techniques offers a pathway to optimize nutrient management, elevate crop yields, and uphold environmental sustainability.

Machine learning algorithms have showcased their capacity to analyze extensive datasets sourced from diverse outlets such as soil samples, weather data, crop varieties, and historical yield records. By discerning meaningful patterns and relationships, these algorithms can furnish precise forecasts regarding crop nutrient requirements. This predictive prowess empowers farmers with actionable intelligence, enabling them to apply fertilizers with greater precision and efficiency. The advantages of such a paradigm shift in fertilization strategies are multifaceted.

Primarily, it can lead to substantial reductions in resource wastage. Conventional practices often result in over-fertilization, triggering issues like nutrient runoff, soil depletion, and groundwater pollution. By customizing nutrient application to match crop demands precisely, machine learning-driven approaches can alleviate these adverse environmental consequences. Moreover, optimizing nutrient

Table 2. Represents the performance metrics of the testing dataset.

Model	Test Dataset			
	RMSE	MAE	MSE	R^2
SVM	1627.8	1311.7	2.6	0.36
XGBoost	1344.8	1219.8	1.8	0.71
DT	1003.4	849.1	1.0	0.78
Logistic Regression	814.7	715.1	0.66	0.87

Table 1. Represents the performance metrics of the training dataset.

Model	Training Dataset			
	RMSE	MAE	MSE	R^2
SVM	710.8	536.1	0.51	0.73
XGBoost	569.7	463.8	0.32	0.83
DT	709.4	573.2	0.5	0.73
Logistic Regression	461.6	386.1	0.21	0.89

delivery holds the promise of directly boosting crop productivity.

Fine-tuning nutrient levels based on predictive models can bolster crop health, expedite growth, and amplify yields. This assumes particular significance amid the expanding global population, which places heightened pressure on agricultural outputs. Furthermore, the economic advantages are undeniable. Precision fertilization stands to yield cost savings for farmers by curbing unnecessary fertilizer expenses. Additionally, the improved yields stemming from optimized nutrient management can enhance overall profitability.

References

Bhanumathi, S., Vineeth, M. and Rohit, N. 2019. Crop Yield Prediction and Efficient use of Fertilizers. *2019 International Conference on Communication and Signal Processing* (ICCSP), 25 April 2019.

Devan, K.P.K.B., Swetha, P. Uma Sruthi and Varshini, S. 2023. Crop Yield Prediction and Fertilizer Recommendation System Using Hybrid Machine Learning Algorithms. *2023 IEEE 12th International Conference on Communication Systems and Network Technologies* (CSNT), 31 May 2023.

Elizabeth Rani, G., Ede Venkatesh and Karnam Balaji, 2022. An automated prediction of crop and fertilizer disease using Convolutional Neural Networks (CNN). *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering* (ICACITE),18 July 2022.

Hongjian Zhang, Chunbao Xu, 2021. Fertilizer Strength Prediction Model Based on Shape Characteristics. *IEEE Access*. 9.

Jeevaganesh, R., Harish, D. and Priya, B. 2022. A Machine Learning-based Approach for Crop Yield Prediction and Fertilizer Recommendation. *2022 IEEE 6th International Conference on Trends in Electronics and Informatics* (ICOEI), 28-30 April 2022.

Krutika Hampannavar, Vijay Bhajantri and Shashikumar G. Totad, 2019. Prediction of Crop Fertilizer Consumption. *2018 Fourth International Conference on Computing Communication Control and Automation* (ICCUBEA), 25 April 2019.

Laasya Kanuru, Amit Kumar Tyagi and Aswathy S U. 2021. Prediction of Pesticides and Fertilizers using Machine Learning and Internet of Things. IEEE 2021 *International Conference on Computer Communication and Informatics* (ICCCI -2021), Jan. 27-29, 2021

Michela Battisti, Barbara Moretti and Dario Sacco, 2021. Soil Olsen P response to different phosphorus fertilization strategies in long-term experiments in NW Italy. *Soil Use and Management.* 38(1): 549-563.

Monika, K., Balakrishnan Ramprakash, Sankayya Muthuramalingam and Mirdula, K. 2022. Crop Fertilizer Prediction using Regression analysis and Machine Learning algorithms", 2022

Rama Devi, O., Naga Lakshmi, P. and Naga Babu, S. 2023. Fertilizer Forecasting using Machine Learning. *2023 International Conference on Inventive Computation Technologies* (ICICT), 01 June 2023.

Sai Yasehwanth Chaganti and Prajwal Ainapur, 2019. Prediction Based Smart Farming. *2019 2nd International Conference of Computer and Informatics Engineering (IC2IE).*