JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Machine Learning Based Nutrition Recommendations System for Soil Fertilization

[1]J.S.Sowmiya, [2]Saranya.S [3]Sriponbala.K, [4]Sarangam Sai Indu, [5]Swathi.A

[1]Assistant Professor, Department of Computer Science Engineering, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering college, Chennai, India

[2,3,4,5]Student, Department of Computer Science Engineering, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering college, Chennai, India

Abstract: Expert guidance is often necessary for farmers to achieve higher yields and minimize fertilizer loss since they typically have limited control over fertilizer use. The impact of rainfall on nutrient loss is also significant, with moderate rainfall at the right time being beneficial for nutrient penetration into the soil and dissolution of dry fertilizers. However, excessive rainfall can result in runoff and increase the risk of losing vital nutrients such as manganese (Mn), phosphorus (P), boron (B), nitrogen (N), and potassium (K) from the soil. To address this issue, we have proposed a nutrient recommendation framework that utilizes a revised version of the random forest algorithm and time-series data to analyze rainfall patterns and crop fertility. The method predicts the optimal nutrient quantity required for different crops and provides customized recommendations to farmers based on local conditions. By adopting this approach, farmers can enhance soil fertility, improve crop growth, and reduce the risk of leaching and runoff.

IndexTerms **– Random Forest, Decision making, Leeching, Nutrition Recommendation, Fertilizer**

I. INTRODUCTION

Agriculture plays a significant role in the national economic growth, contributing 17-18% to India's GDP and ranking second globally in farm outputs. For plants to grow optimally, they require compost and fertilizers to replenish nutrients depleted from the topsoil. Inadequate compost can lead to a decline in crop yield. However, fertilization must be done with precision, taking into account rainfall patterns and the specific nutrient requirements of each crop. Machine learning technology can effectively address this issue by analyzing available data on crop fertility and rainfall, providing farmers with valuable information. The proposed model uses a random forest regression algorithm with k-fold cross-validation technique (a machine-learning algorithm) and prompts users for two inputs, namely, crop and location. After applying the algorithm, the model predicts the required amount of nutrients and the optimal time for fertilizer application. The website is developed using Flask Python (a web framework), ensuring accessibility on all platforms and sharing among users.

II. PROPOSED MODEL

Using random forest, a predictive model was developed in this research to estimate the nutrients essential for crops. The model was represented by random forest regression along with k-fold cross-validation technique, and after achieving satisfactory accuracy for the prediction, the final model was obtained. To evaluate the algorithm, seven features were utilized.

The proposed system aims to mitigate farmers' losses by offering valuable insights on fertilizer usage, reducing water pollution caused by leaching, and bridging the gap between farmers and modern technology. Its ultimate goal is to help farmers enhance their crop yield while reducing the quantity of inputs required. The system takes the form of a website, ensuring platform-independent functionality for users to access it from any device. The user interface has been designed to prioritize functionality over complexity, making it user-friendly even for inexperienced individuals. To estimate the appropriate levels of nitrogen, potassium, and phosphorus for a given crop, As shown in fig.1 the website utilizes machine learning algorithms after obtaining information such as state, city via drop-down menus and crop type. The system also utilizes a Weather API to collect data on environmental conditions like temperature, humidity, and rainfall, and it will alert users to potential heavy rainfall that may necessitate precautionary measures. The proposed algorithm is otherwise followed, and the system is known for its high degree of accuracy in nutrient decision-making.

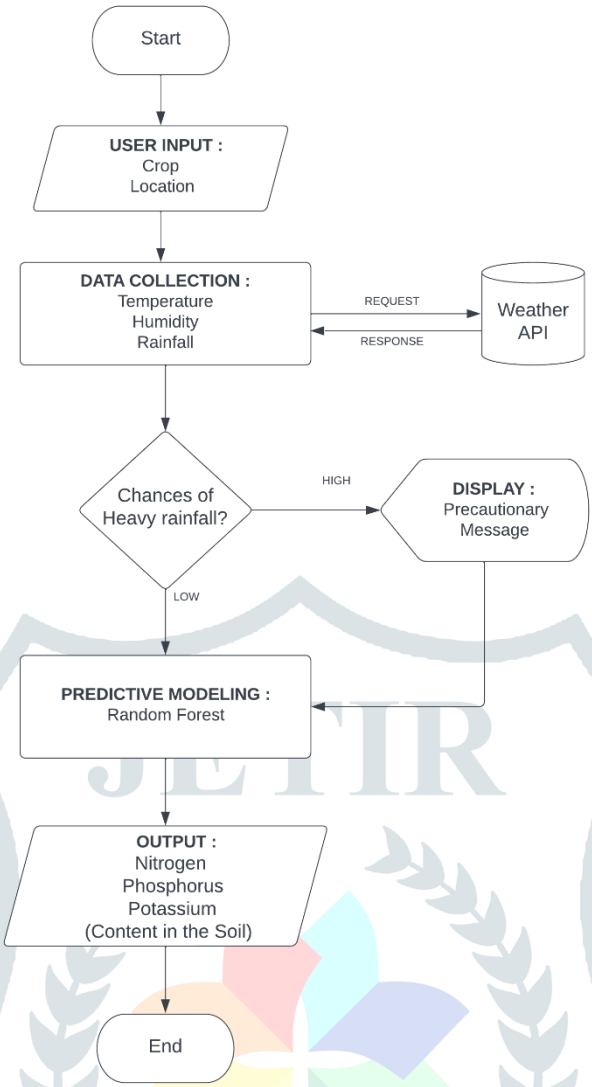

Figure 1: Project Flow

2.1 Random Forest Algorithm

Our project involves using a random forest (RF) algorithm, which is made up of multiple decision trees trained on different subsets of data and can be adjusted through hyper-parameters. We plan to predict the values of Potassium(K), Nitrogen(N), and Phosphorous(P) based on input of crop and location. The first step in our approach is to partition the dataset into two subsets, namely training and test sets. The training set will contain 80% of the original data, while the test set will hold the remaining 20% for assessing the models' performance. After that, we will construct three independent random forests, each consisting of 50 decision trees, to forecast the values of K, N, and P. The average of the classes produced by each random forest will serve as the overall prediction for each nutrient. We have tested for different n_estimator values, but the upmost accuracy achieved for *N_Label* is 0.87 for two decimal digit precision. As shown in fig.2 so we selected n_estimator as 50 for each label.

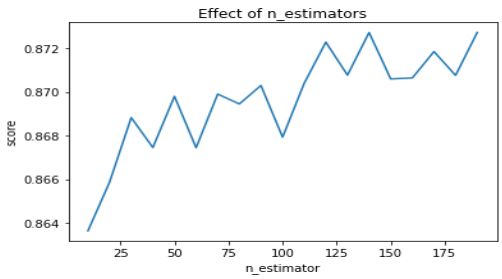

Figure 2: n_estimators

The initial step involves dividing the dataset, which includes 2200 observations, into two separate datasets - a training dataset and a test dataset. Specifically, the training dataset is composed of 80% or 1760 of the observations, while the remaining 20% or 240 observations make up the test dataset.

Next, in step two, random forest regression is applied to each of the three variables - Nitrogen, Phosphorus, and Potassium - with 50 decision trees (n estimators) for each.

During step three, the training dataset is used to train the K Label, N Label, and P Label by using the dependent variable as the target. Specifically, the K Label is trained with K as the dependent variable, the N Label is trained with N as the dependent variable, and the P Label is trained with P as the dependent variable.

Finally, In last step, a set of 50 decision trees is generated as output by each of the K Label, N Label, and P Label based on the training dataset.

In general, adding more trees will increase performance and predictability while decreasing calculation speed. The end solution for regression issues is the mean of all the trees. The samples in the tree target cell are the initial level of means in a random forest method regression model, followed by all trees. In contrast to linear regression, it estimates values beyond the observed range using prior observations.

The accuracy in the decision tree depends on the number of right predictions made divided by the total number of predictions, since it uses huge value attributes at each node, and it produces less accurate results when we apply an algorithm to handle the regression problem in a random forest. Decision trees are greedy and may be deterministic, meaning they produce different answers if we add or remove any additional rows. So, compared to decision trees, random forest forecasts outcomes with higher accuracy.

One of the most significant challenges in machine learning is the problem of overfitting. Overfitting is a common bottleneck that arises when we apply algorithms in machine learning. Essentially, overfitting occurs when machine learning models are unable to perform well on new or unknown datasets, and this is often reflected in poor performance on testing or validation datasets. Overfitting is often detected when the error on these datasets is significantly larger than the error on the training dataset. Overfitting occurs when models gain knowledge from non-constant data in the training data, which can negatively impact the performance of the model on new data. This is because the model has learned specific patterns in the training data that may not be generalizable to new data.

One approach to reducing the risk of overfitting is to use an ensemble of decision trees, such as in the case of a random forest. Because random forests employ multiple decision trees, the danger of overfitting is generally lower than that of a single decision tree. By averaging the results of many trees, the random forest can capture a broader range of patterns and avoid overfitting to specific patterns in the data. Overall, overfitting is a significant challenge in machine learning that can negatively impact the performance of models on new data. Using techniques like ensembling can help reduce the risk of overfitting and improve the generalizability of machine learning models.

When it comes to classifying data sets, decision tree classifiers can be useful because they contain a large number of splits. These splits can make it easier to overfit the data, which means that the classifier is too closely tailored to the training set and may not perform well on new data. However, decision trees can also be validated to ensure that they are not overfitting.

In our case, we needed to predict the required nutrients (Nitrogen, Phosphorus, Potassium) for a given crop. While decision tree classifiers could have been used, we decided to use a random forest model instead. Random forests are similar to decision trees, but they use multiple decision trees and their predictions are averaged to produce a more accurate result. Random forests can perform well in terms of computation if the n_estimator value is carefully adjusted. The n_estimator value determines the number of decision trees in the random forest, and in our case, we used a value of 50. To determine the best n_estimator value for our model, we tested its accuracy with different values and found that 50 was the optimal value.

Overall, while decision tree classifiers can be useful for data sets with many splits, random forests are a better choice for predicting the required nutrients for a given crop. By adjusting the n_estimator value carefully, random forests can provide accurate results with reasonable computation times.

2.2 Data Preparation

The dataset comprises of eight features, but not all of them are relevant to the proposed model. As a result, a dimension reduction technique known as feature selection was utilized to choose the seven most pertinent features for evaluation.

2.3 Input Features

The algorithm takes several characteristics as input, including the type of crops being grown (up to 22 different varieties, such as grapes, apple, orange, rice, jute, and so on.), the temperature in Celsius, the relative humidity in percentage, and the amount of rainfall in millimeters.

2.4 Output Features

The output features are Label K (proportion of Potassium present in the soil), Label N (proportion of Nitrogen present in the soil) and Label P (proportion of Phosphorous present in the soil).

2.5 System Architecture

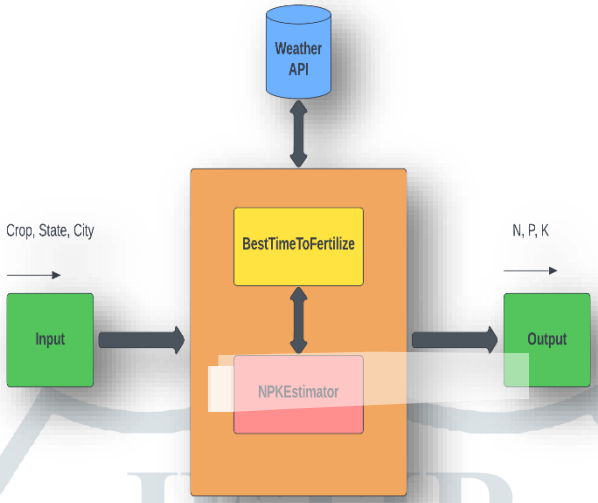

Fig 3: Block Diagram

The description of each component in the block diagram has the major functionalities with respect to the application:

- **Input:** Allows the user to provide input data such as the crop, state, and city by selecting options from the drop-down menu.
- **Weather API:** Fetches weather details such as temperature, humidity, and rainfall from an external Weather API.
- **BestTimeToFertilize:** Analyzes the weather data it retrieves and identifies the optimal timing for fertilizer application, while also issuing a notification in the event of anticipated heavy rainfall.
- **NPKEstimator:** Determine the optimal balance of Nitrogen, Phosphorus, and Potassium (NPK) in soil based on various factors related to the specific crop and soil.
- **Output:** Displays the estimated Potassium Nitrogen, and Phosphorus content on the website for the user to view.

III. OUTPUT AND RESULTS

The webpage is constructed using HTML, CSS, and JavaScript. Upon landing on the first page, the user enters details such as their state, crop type, and city. Once the information is submitted, the algorithm processes the input data and generates a report. The report includes the 7-day weather forecast and the recommended ratio of phosphorus, nitrogen, and potassium for the selected crop.

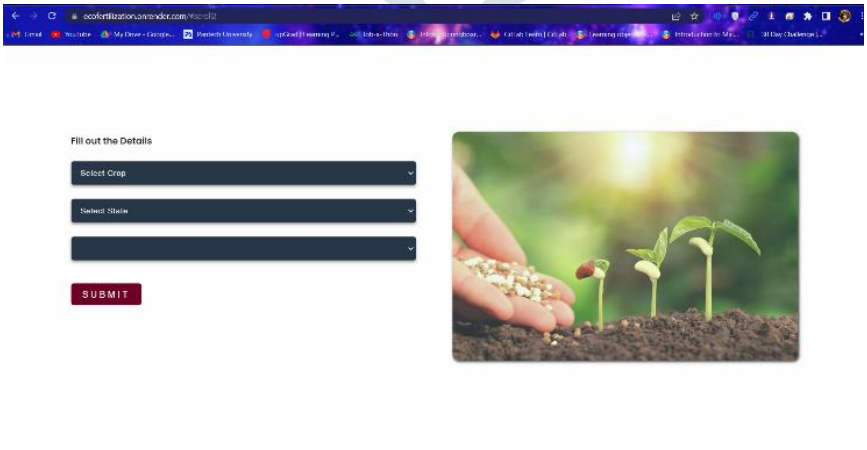

Figure 4: Input Page

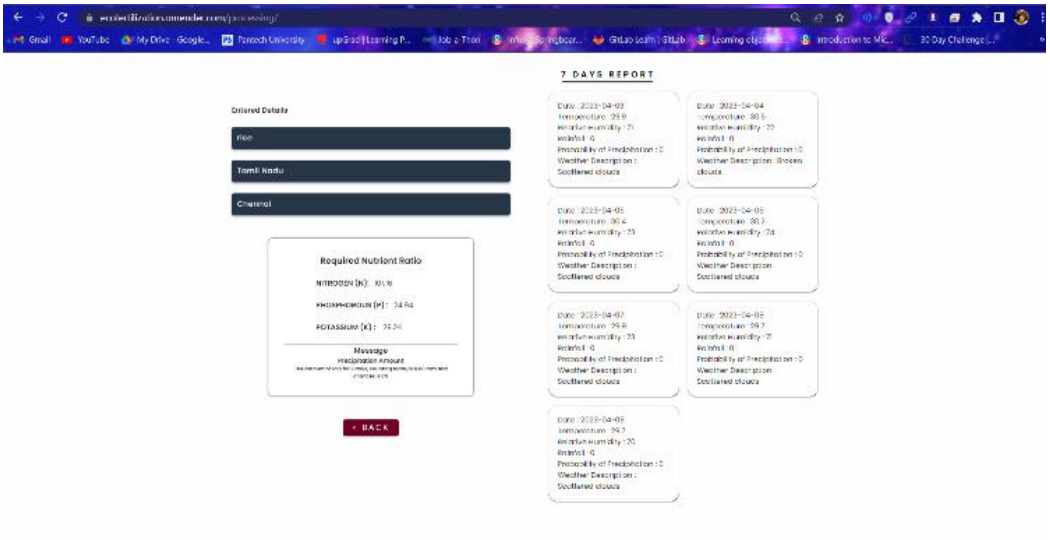

Figure 5: Output Page

IV. CONCLUSION AND FUTURESCOPE

The system proposed is capable of achieving a high accuracy rate of 92%, which is considered quite good for any predictive model. It offers valuable information regarding the optimal use and amount of nutrients required for the crops to ensure satisfactory growth and yield based on the prevailing weather conditions. Additionally, it provides weather alerts and messages to the user, with any relevant warnings being displayed in the application's output in the event of poor weather conditions. As technology continues to advance, there is scope for further improvement in accuracy.

The proposed system serves as a valuable tool to assist farmers by providing them with crucial information on the appropriate use and quantity of nutrients required for their crops. However, there is potential for further improvement in the system by incorporating a user interface in the local language. This would make it easier for users who are unfamiliar with the English language to operate the system seamlessly. Additionally, speech recognition systems can be added to enable illiterate users to use the system effectively. These enhancements would make the system more accessible and user-friendly, thereby increasing its adoption rate and usefulness to farmers.

REFERENCES

[1] M.S. Suchithra, Maya L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters", Elsevier,2022.

[2] Senthil Kumar Swami Durai, Mary Divya Shamili, "Smart farming using Machine Learning and Deep Learning techniques", Elsevier,2022.

[3] Rubby Aworkaa, Lontsi Saadio Cedric, Wilfried Yves Hamilton Adoni,Jérémie Thouakesseh Zoueuc,Franck Kalala Mutomboa, Charles Lebon Mberi Kimpoloa,Tarik Nahhal, Moez Kricheng,"Agricultural decision system based on advanced machine learning modelsfor yield prediction: Case of East African countries",Elsevier,2022.

[4] Saheed Garnaik, Prasanna Kumar Samant, Mitali Mandal, Hari Sankar Nayak, "Untangling the effect of soil quality on rice productivity under a 16-year long-term fertilizer experiment using conditional random forest “, Elsevier,2022.

[5] QIAO Fang-bin, HUANG Ji-kun,"Farmers’ risk preference and fertilizer use”,Elsevier,2022.

[6] Akhter, R., Sofi, S.A., "Precision Agriculture using IoT Data Analytics and Machine Learning", Journal of King Saud University - Computer and Information Sciences ,2021.

[7] Benny Antony,"Prediction of the production of crops with respect to rainfall”,Elsevier,2021.

[8] Zujiao Shi, Donghua Liu, Miao Liu , Muhammad Bilal Hafeez , Pengfei Wen ,Xiaoli Wang , Rui Wang , Xudong Zhang , Jun Li ,"Optimized fertilizer recommendation method for nitrate residue control in a wheat–maize double cropping system in dryland farming",Elsevier,2021.

[9] Pallavi Kamath, Pallavi Patil, Shrilatha S, Sushma, Sowmya S,"Crop Yield Forecasting using Data Mining",Global Transitions Proceedings,2021

[10] Usman Ahmed , Jerry Chun-Wei Lin , Gautam Srivastava , Youcef Djenouri,"A nutrient recommendation system for soil fertilization based on evolutionary computation",Elsevier,2021.

[11] Mahmood A. Hussein, Diogenes L. Antille , Shreevatsa Kodur, Guangnan Chen,Jeff N. Tullberg,"Controlled traffic farming effects on productivity of grain sorghum, rainfall and fertiliser nitrogen use efficiency",Elsevier,2021.

[12] Thomas van Klompenburga, Ayalew Kassahuna, Cagatay Catal,"Crop yield prediction using machine learning: A systematic literature review",Elsevier,2020.

[13] János Kátai, Ágnes Oláh Zsuposné, Magdolna Tállai, Tarek Alshaal,"Would fertilization history render the soil microbial communities and their activities more resistant to rainfall fluctuations?",Elsevier,2020.

[14] Janmejay Pant, R.P. Pant , Manoj Kumar Singh , Devesh Pratap Singh , Himanshu Pant,"Analysis of agricultural crop yield prediction using statistical techniques of machine learning",Elsevier,2020.

[15] Tony Yang , Kadambot H.M. Siddique, Kui Liu,"Cropping systems in agriculture and their impact on soil health-A review",Elsevier,2020.

[16] Potnuru Sai Nishant,Pinapa Sai Venkat,Bollu Lakshmi Avinash,B. Jabber, "Crop Yield Prediction Based on Indian Agriculture using Machine Learning.", INCET, 2020.

[17] Laura J.T. Hess,Eve-Lyn S. Hinckley,G. Philip Robertson,Pamela A. Matson,"Rainfall intensification increases nitrate leaching from tilled but not no-till cropping systems in the U.S. Midwest",Elsevier,2020.

[18] Yulong Yin, Hao Ying,Huifang Zheng ,Qingsong Zhang ,Yanfang Xue,Zhenling Cui,"Estimation of NPK requirements for rice production in diverse Chinese environments under optimal fertilization rates",Elsevier,2019.

[19] G. Prabakarana, D. Vaithiyanathan, Madhavi Ganesan,"Fuzzy decision support system for improving the crop productivity and efficient use of fertilizers",Elsevier,2018.

[20] Kaggle, " https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset " (accessed on 16th November 2021)