International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.014

Precision Farming Tool featuring Crop Recommendation and Intelligent Fertilization

RiaSehgal
Department of Computing Technologies, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute of Science and Technology,
Kattankulathur, 603203
rs6904@srmist.edu.in

ShayanaZaman
Department of Computing Technologies, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute of Science and Technology,
Kattankulathur, 603203
sz9919@srmist.edu.in

P Selvaraj
Department of Computing Technologies, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute of Science and Technology,
Kattankulathur, 603203jj
selvarap@gmail.com

***Abstract*—Agriculture is a major economic driver in India but productivity related issues pinpoints to the significance of choosing the right crop under given conditions. Agriculture is in turn intrinsically dependent on the fertilizer industry - one of the eight core industrial sectors in India. However, tackling issues overuse and runoff leakages has been a persistent problem because of a "one-size-fits-all" approach and can be corrected through the principles of Precision Farming. One of the ways of addressing geographical variances in agricultural needs could be by informing cultivators of site-specific required NPK (Nitrogen - Phosphorus - Potassium) ratio for various crops. Underlying this idea, this work pitches a two - dimensional recommendatory ML-enabled tool that assesses soil quality parameters to provide features of (i) Crop Recommendation using ensemble technique built by six classification base learners for greater accuracy encompassing greater number of crops, (ii) Intelligent and balanced fertilization by predicting weather-affected nutrient (specific to macro-nutrients NPK ratio) requirements of a region by calling live weather API.**

Keywords*— precision farming, nutrient management, intelligent fertilization, Ensemble Learning, NPK ratio

I. Introduction

The Green Revolution was the harbinger of mass incorporation of fertilizers in Indian agriculture and made the country self-sufficient but also brought forth newer complexes. The sole focus on increasing productivity through High Yielding Varieties (HYV) of crops has still kept optimal and sustainable farming a farsighted goal because of unawareness related to agricultural sciences, fragmented land holdings, dependence on weather etc. Also, the concentration of its success in Punjab, Haryana and the subsequent rice-wheat boom highlights that the framework did not entirely address geographical or climatic variations across the nation. Precision farming introduces smart practices that address these differences.

Precision Farming is a concept that uses tech to ensure crops and soil receive what they need for optimal growth and productivity. It involves site-specific agricultural practices. Experts at the International Society of Precision Agriculture (ISPA) have highlighted the [1] 5Rs that can help attain the goals of Precision Farming are - right place, right time for the right input, in the right quantity in the right manner.

The Indian Council on Agricultural Research (ICAR) and the National Commission of Farmers 2006 concurred that the nation-wide [2] standard ratio of 4:2:1 for primary fertilizers NPK doesn't hold for all the regions of the country. Several studies show that farmers plant an inappropriate crop based on conventional or non-scientific approaches and try to ensure high yield through mindless use of fertilizers.

The aim of this project is to advocate the principles of smart farming by providing accurate crop recommendations and integrating nutrient management (of macronutrients - N, P, K) at a regional scale using weather conditions instead of standardizing it. The presented work is intended to present a prototype for introducing precision nutrient management in a more dynamic and dependable form.

The objectives of the project are : (i) To customize crop-fertilizer management (ii) To advocate for integrated nutrient management in farming (iii) To build a competent advisory service to eliminate unawareness and negligence in agricultural practices regarding fertilizer use inline with recommendations of ICAR and National Commission of Farmers (2006) (iv) Use of machine learning techniques to enhance accuracy and legitimacy.

II. Literature Survey

[3] This work seconds the need for agricultural practices integrated with nutrient management. The optimal nutrient input ratio has been greatly impacted by the large interstate and interregional variability in fertilizer use.

[4] A precision agriculture project in Karnataka used GIS mapping to accommodate spatial variations of soil properties to ascertain yield variability of three crops most commonly grown in three north-eastern regions. This pinpoints the fact that deployment of information technology at micro-level can much better cater to agricultural goals. Climatic conditions however were not considered.

[5] Precision farming practices address geographic and temporal variations in nutrient availability that facilitates better decision making based on the vital gap between plant demand and soil supply. Under this premise, the use of sensors for capturing soil properties, aerial crop images and site maps is not established for nutrient management. Their

DOI: rp-9788770040723.014

potential remains to be validated under various agro-climatic conditions.

[6] Focus should be on solving productivity related issues by thoughtful crop selection instead of trying to increase the yield of a lesser apt crop for the region. Eradication of problem at the source through better crop and fertilizer management would lead to diversification of crops and reduced fertilizer consumption. fter comparing with other technological frameworks, this work suggests ensembling technique to be best suited for crop recommendation. Three independent base learners have been used to build the recommendatory model, where the advantage in using independent frameworks lies in time efficiency in comparison to dependent frameworks.

[7] SVM, Random Forest have been used for soil classification and crop yield prediction whereas [8] Decision tree, K- nearest neighbor, Linear Regression, Naive Bayes also showed impressive results. The choice of base learners for our adopted ensemble learning framework was guided by the conclusions of these works.

[9] Taking NPK ratio for a single crop (i.e. rice) helps make the following conclusions: climatic factors such as temperature variations and total rainfall, change nutrient concentrations - rainfall in the right quantity can augment nutrient sufficiency but extreme rainfall would lead to runoffs.

[10] The prediction of crops as a function of rainfall in four states of India is determined using a comprehensive evaluation of all regression techniques. This is done predicting crop yield for a given amount of rainfall.

While random forest has been widely used independently for crop recommendation or crop yield prediction, [11] shows the abilities of conditional random forest models to evaluate data to create useful information in fertilization studies and that such techniques can be used in other long-term techniques.

III. Proposed System

The proposed system emphasizes on the primary fact that climatic conditions and soil properties affect crop production and agricultural practices like fertilization should be attuned to these variations. Two operational modules are taken into consideration

A predictive model based on Random Forest Regression with K-fold cross-validation technique is used for determining nutrient requirement for a crop. The user is required to enter input features of crop (to be harvested) and location. The state and city entries direct the model to the live weather API which accumulates values such as regional rainfall, humidity and temperature to finally predict values of N, P , K requirements.

The second module employs the ensemble technique for crop recommendation across multiple Indian states for 26 crops through independent base classification learners. The N, P, K ratio for a particular location is used to classify crops. The choice of classification techniques is being guided by similar works done on crop recommendation. Entering soil inputs to get best suited crops and finding optimal soil requirements for maximum, sustainable output - the module can work both ways. The system is presented under the assumption that farmers or cultivators have access to nutritional information of the soil like nutrient presence and pH values through their respective soil health cards provided by the government. The use of sensors otherwise for quantifying nutritional values of soil in an IoT based environment, has seen limited on-site implementation and hence lab-based tests become a more dependable source of input data.

IV. Methodology

1. Data Preprocessing of raw input

The most crucial step for an accurate model requires pre-processing that can convert raw, unorganized data into a form that can be understood by computers and used for machine learning models. Mathematical concepts of mean, mode, median are employed along with scaling values to a particular range, data cleaning, data encoding and determining correlations among variables.

2. Deploying exploratory data analysis

Exploratory Data Analysis is an essential step of doing initial assessment on data in order to uncover patterns, identify irregularities, test hypotheses, and validate assumptions using summary statistics and graphs. This can take the form of univariate, bivariate or multivariate analyses. PDF, CDF are some instances of univariate analysis and box plot and heat map are forms of multivariate analysis.

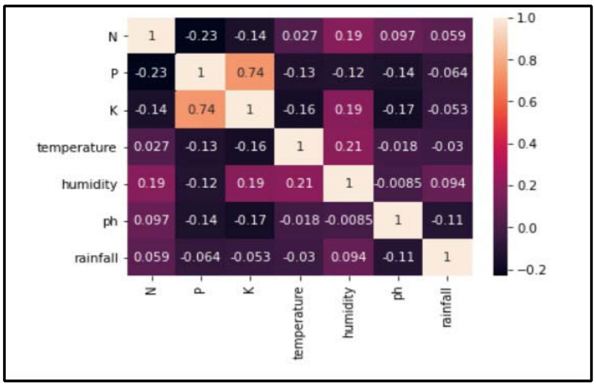

Fig.1. Heatmap-correlation metrics of the dataset for crop recommendation

3. Data Splitting

Training and testing datasets are formed on a 80:20 ratio. The 20% data is used to test the model on the unknown data and perform cross-validation to determine the hyperparameter - [13] these are parameters that govern the learning process acquired by a learning algorithm. These are 'top-level' parameters that drive the learning process and other model parameters that come from it.

4. Training multiple models on the same dataset

The preprocessed training set is fed onto multiple independent learning models whose results will be accumulated for ensemble technique that will finally give a class label from every classifier. Our choice of classifiers was followed by as assessment of related works, post which the most frequently used classifiers were chosen

a) Decision Tree - as a supervised learning tool based on a tree-like structure, it can be used for both regression

DOI: rp-9788770040723.014

and classification. [14] A decision tree has a structure similar to a flowchart, with a tree representing attributes and class labels.

b) Naive Bayesian Classifier - the advantage being that it is simple and straightforward and works under the premise that all features are independent of each other.

c) Logistic Regression - classification of a datapoint in an n-dimensional space. It is based on statistical recognition whereas SVM is based on geometrical properties.

d) Support Vector Machine (SVM) - where the main task involves determining the best hyperplane. Each data having n features is plotted onto a n-dimensional space and the hyperplane must segregate any two given classes (in a linear kernel) with the highest margin. [15] SVMs require lesser variables for a similar or higher MCR value

e) XGBoost - in comparison to decision trees, it is a more advanced tree-based approach sed for gradient boosting implementations

f) Random Forest - a miniature of ensembling technique in which multiple decision trees are assessed for maximum depth until nodes split with minimum variability and bias

5. Testing data

A query point is presented to each classifier. A class label is the output of every classification model.

V. Implementation

The dataset has been taken from kaggle and referred from an earlier work [12] because of the presumed ease with data preprocessing. The sample size for this work comprises 2200 crop samples of 22 unique crops. This collection of data had the scope of encompassing multiple features. For the fertilization module, feature selection was used as a dimensionality reduction technique to simplify the model.

Various public datasets about India are augmented and combined in this project, such as weather, soil, and so on. Contrary to complicated features affecting yield of crops, this data is relatively simple and has only a few but useful features. It includes nitrogen, phosphorus, potassium, and pH values. In addition, it contains the humidity, temperature, and rainfall needed for each crop.We have used Pandas library through which we preprocessed our data efficiently. We converted the raw data to mergeable data so that it can be appropriately passed through our model. Libraries used in python are - pandas, numpy, matplotlib, pyplot, seaborn, sklearn etc. The corresponding features are nitrogen, potassium, phosphorus, temperature, humidity using which 22 unique features are predicted.

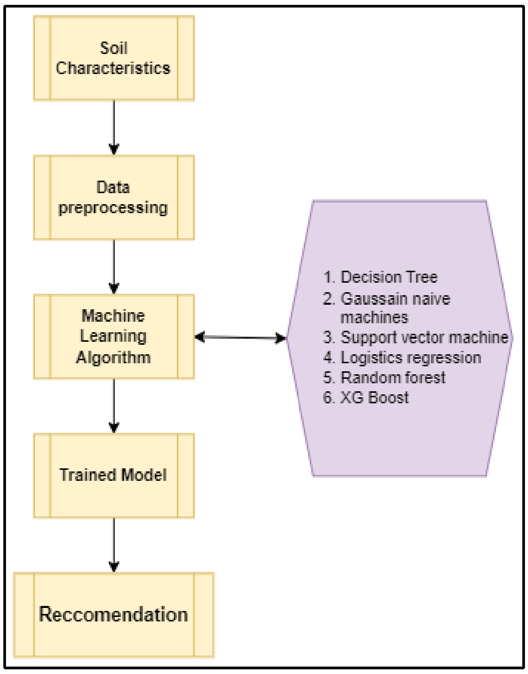

Fig. 2. Flow Chart of steps for Crop Recommendation using Ensembling Technique

The distinct class labels obtained from each classifier are applied to the majority voting technique to produce an ensembled class label as the final prediction.

Ensemblimg Technique for Crop Recommendation

Ensemble methods is a machine learning technique that integrates numerous base models to create a single best predictive model.The ensemble framework should be used primarily because it offers a classifier that performs better than each of the classifiers learning independently.

[16] Ensembling technique is of two forms - dependent and independent. In the former technique, the result of one classifier is further used serially in the learning process of the consecutive classifier. In an independent model, a class label is generated by each classifier without any input from preceding learners. Here, all classifiers work in parallel. Parallel execution makes it a more time - efficient alternative taking lesser execution time. It consists of the following components: training data (S) , base inducers (learners denoted by I), diversity generator and a combiner. Where the model M, can be represented as: M = I(S).

Random Forest Algorithm for Fertilizer Recommendation

Random Forest algorithm is also a classification algorithm that uses ensembling technique that combines the results of similar models to create a strong model built by multiple decision trees. The training is done via what is called as bagging (form of ensemble method) where the combination of multiple tree-like classifiers, trained on varying subsets of data, increases the overall accuracy. Three different random forests each having 50 decision trees (n estimators = 50), are created for each of the nutrients - N, P, K. The end result is the output of the mean of classes.

DOI: rp-9788770040723.014

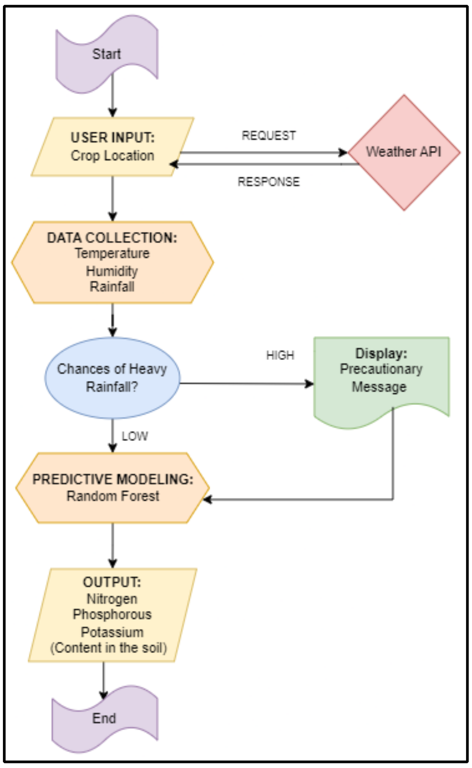

Fig. 3. Flow Chart of decision flow for Fertilizer Recommendation using Random Forest

We see here that the random forest model uses live weather API information to attain humidity levels or probabilistic values of precipitation using which required NPK values are displayed for a particular crop.

Majority Voting Technique

Majority voting is a concept using which the final result is obtained in independent ensemble learning. The labeling of an unknown data is done based on the highest number of votes obtained by a classifier - known as plurality votes. Majority voting is the most commonly used combiner in ensembling technique

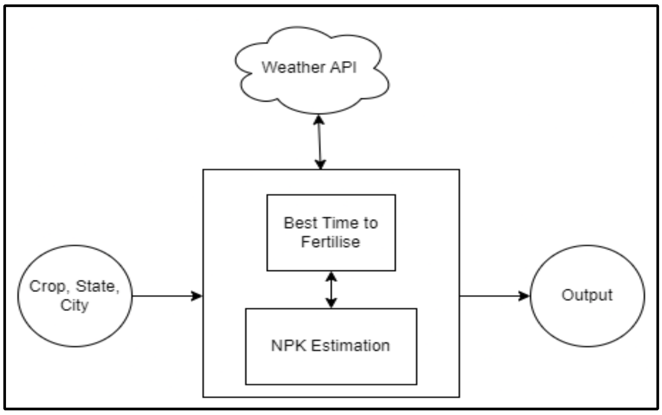

Fig 4. Block Diagram of Proposed NPK Fertilization

VI. Results and Discussion

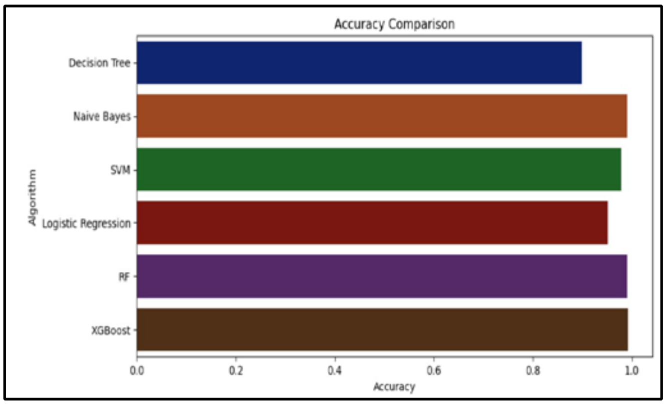

Fig 5. Accuracy comparison of different classifiers used in ensemble learning

Using Majority Voting Technique (MVT) different classifiers have been compared and the one with the highest votes is used to label the unknown data. The accuracy for crop recommendation has been shown to be 99% while required N, P,K ratio is being generated with an accuracy of about 87%

Hence, this project demonstrates how nutrient management can be implemented, through machine learning tools of ensemble learning and random forest with k-cross validation for better decision making in line with recommendation of commissions, that focused on a localized model for agricultural practices rather than a standardized one. A regional model will more comprehensively address the needs

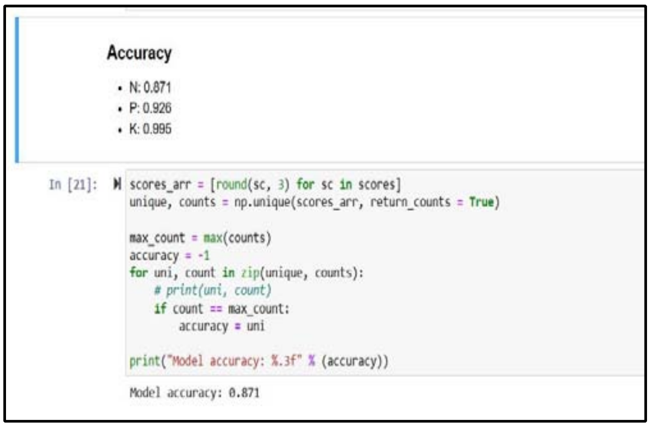

Fig 6: Model Accuracy of Proposed NPK Fertilization

VII. Conclusion and Future Works

A prototype Smart Farming tool has been presented with noteworthy accuracy to aid crop-fertilizer management by considering site-specific conditions of soil properties and rainfall. Fertilizer usage has been optimized through intelligent decision making.

Using crop sensors for soil assessment does allow nondestructive analysis of samples, however their validity has not yet been determined for precision nutrient management. This increases the scope of such a tool that can be deployed with more local information at micro-level for actual implementation as can be said from similar precision agriculture projects in states like Karnataka and

DOI: rp-9788770040723.014

Kerala, where good reviews were obtained. The main concern however with introduction of technology in conventional agricultural practices remains awareness and capacity-building through training etc.

REFERENCES

[1] U.K. Shanwad, M.B. Patil, N.L, Rajesh, H. Veeresh, G.M. Hiremath, A. Amaregouda, P.S. Kamannavar, V.C. Patil, and R.Khosla "Precision Nutrient Management in Cotton: A Case Study From Karnataka"

[2] NAAS, Crop Response and Nutrient Ratio". Policy Paper No. 42, National Academy of Agricultural Sciences, New Delhi, 2009.

[3] A.K. Shukla, S.K. Behera, S.K. Chaudhari, and G. Singh, "Fertilizer Use in Indian Agriculture and its Impact on Human Health and Environment," Indian Journal of Fertilizers, 2022

[4] M.B. Patil, U.K. Shanwad, H. Veeresh, P.R. Mastan Reddy, R. BG, S. NL, L.B. Hugar, B.T. Pujari, and B.V.Patil, "Precision agriculture initiative for Karnataka: A new direction for strengthening the farming community," Sch. J. of Agric. Sc., vol. 3, no. 10, pp. 445-52, 2013.

[5] B.S. Brar, and A.Kaur, "Precision Nutrient Management: A Review," Indian Journal of Fertilizers. 2016 Nov1.

[6] N.H. Kulkarni, G.N. Srinivasan, B.M. Sagar, and N.K. Cauvery,"Improving crop productivity through a crop recommendation system using ensemble technique," In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), IEEE, pp. 114-119, 2018 Dec 20.

[7] Rajesh, M., &Sitharthan, R. (2022). Image fusion and enhancement based on energy of the pixel using Deep Convolutional Neural Network. Multimedia Tools and Applications, 81(1), 873-885.

[8] D.A. Bondre, and S.Mahagaonkar, "Prediction of crop yield and fertilizer recommendation using machine learning algorithms," InternationalJournal of Engineering Applied Sciences and Technology, 2019 Sep

[9] Yulong Yin, Hao Ying, Huifang Zhen, Q ingsong Zhang, Y anfangXue, and I. Zhenling, "Estimation of NPK requirements for rice production in diverse Chinese environments under optimal fertilization rate," Agricultural and Forest Meteorology

[10] Benny Antony, "Prediction of the production of crops with respect to rainfall.," Environmental Research, vol. 202, pp. 1-5, 2021.

[11] S. Garnaik, P.K. Samant, M. Mandal, T.R. Mohanty, S.K. Dwibedi, R.K. Patra, K.K. Mohapatra, R.H. Wanjari, D. Sethi, D.R. Sena, T.B. Sapkota, "Untangling the effect of soil quality on rice productivity under a 16-years long-term fertilizer experiment using conditional random forest," Computers and Electronics in Agriculture,vol. 1, no. 197, p. 106965, June2022.

[12] R.K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A.Devare, "Crop recommendation system to maximize crop yield using machine learning techniques," International Research Journal of Engineering and Technology,vol. 4, no. 12, pp. 950-3, Dec 2017.

[13] Pazhani. A, A. J., Gunasekaran, P., Shanmuganathan, V., Lim, S., Madasamy, K., Manoharan, R., &Verma, A. (2022).Peer–Peer Communication Using Novel Slice Handover Algorithm for 5G Wireless Networks.Journal of Sensor and Actuator Networks, 11(4), 82.

[14] H. Zhang and R. Zhou, "The analysis and optimization of decision tree based on ID3 algorithm," 2017

[15] D.A. Salazar, J.I. Vélez, and J.C..Salazar, "Comparison between SVM and logistic regression: Which one is better to discriminate?"RevistaColombiana de Estadística,vol. 35,no. SPE2, pp. 223-37, Jun2012.

[16] O. Sagi, and L. Rokach, "Ensemble learning: A survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. vol. 8, no. 4, p. e1249, July 2018.