A bootstrap aggregation approach for adequate crop fertilizer and nutrition recommendation

Varshitha D. N., Savita Choudhary
Department of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, India

ABSTRACT
Agriculture is the largest workforce of India and biggest contributor to the Indian economy. Improving agricultural practices with the help of modern computer science technologies have great scope. Helping the farmers to know about their soil fertility, crops which can be grown and fertilizers or nutrients required for their land will be valuable inputs for them. Too much or too little fertilizers may harm the soil, so right amount of fertilization is also important. In this paper we have discussed about the bootstrap aggregation regression method, which is an ensemble machine learning technique to recommend the optimum level of nutrients and fertilizers. Hence customized nutrients recommendation reports could be generated to suggest the fertilizers and nutrients with their adequate quantities. This will be really beneficial for farmers to maintain the soil health and helpful for better crop growth and yield. We consider the features and levels of soil parameters such as nitrogen, phosphorus, potassium (NPK), pH level, organic carbon, electric conductivity, humidity, rainfall and other micro nutrients for predicting the right amount of fertilizers and nutrients. We have also checked other regression methods to compare the results based on the previous work done in the same field.

Keywords:
Bootstrap aggregation
Fertilizers
Machine learning
Nutrients
Regression
Soil parameters

Corresponding Author:
Varshitha D. N.
Department of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology
Bengaluru, Karnataka, India
Email: varsha.swara@gmail.com

1. INTRODUCTION
There is always a scope for improvement in the area of agriculture to reduce the burden on farmers and guide them in right direction. The low productivity of agriculture always is a challenge, to meet the need to supply the food for people and raw materials to industries. Low soil fertility, poor soil health, lack of proper knowledge about fertilizers and pesticides, minimum knowledge about the soil properties and actions needed to improve the soil are few reasons of low productivity of crops [1], [2].

Soil in its original form contains several macro and micro nutrients like nitrogen, phosphorus, potassium (NPK), sulphur, copper, iron, and manganese. Also some features like pH, humidity, temperature, rainfall, monsoon cycles and type of the soil are all necessary for any individual crop to grow in the soil [3]–[6]. Even a single nutrient deficiency may lead to abnormal growth of crops or plants. Plants need all the components of soil in its optimum level to grow healthy. More crop yield is possible only when soil is healthy enough to grow the crops.

Fertilizers act like boosters to improve the fertility of the soil and facilitate the crop growth and they fill the necessary portion of the soil which is missing, so the requirements of the soil can be satisfied by using fertilizers in soil. Always the right amount of fertilization is considered important. Too much or too little nutrition is dangerous for crop growth and soil health. So adequate quantity of each substance should be
informed to farmers. Some farmers may not be able to use fertilizer as it is costlier or willing to organically cultivate the crops, in that case right amount of organic manure which can be decomposed naturally can be also guided to farmers.

There are lot of challenges farmers are facing every day to get adopted to modern technique. ‘Cost’ and ‘ease of use’ are the two main fears Indian farmers are facing. Educating the farmers regarding soil health and crops in easier and efficient ways, would always support them well in every phase of the agriculture. Fertilizer recommendation will improve the soil health and encourage the farmers to enhance the yield and income [7], [8]. Artificial intelligence and machine learning techniques are believed to guide the farmers to perform the right actions, also the machine learning techniques provide time and cost efficient solutions for the same [9]–[12].

In this paper we have discussed about the implementation of soil fertilizer recommendation module. Fertilizer recommender is a part of our ‘soil fertility and crop friendliness detection and monitoring using artificial intelligence (AI) system, discussed in methodology section of this paper. We use bootstrap aggregation method which is a machine learning technique used for predicting the right amount fertilizers and nutrients. Like other recommender systems we are not just classifying the fertilizer type needed for a specific crop rather we will fetch the customised fertilization report for each soil sample given. The results obtained using this approach to recommend the fertilizers, nutrients with their quantities are most accurate and work satisfactorily compared to other approaches. We can also see the results and comparison of our work with other methods for fertilizer recommendation in the ‘results’ section of this paper.

2. RELATED WORK

The K-nearest neighbors (KNN) is a supervised learning regression algorithm which can be used for crop and fertilizer recommendation [13]–[20]. KNN algorithm predicts any value using ‘feature similarity’. Whenever a new point arrives, it will be assigned to the group which has most similar data point. First, the distance between new point and other training points are calculated, based on the distance, the nearest ‘k’ data points are selected. Average of the data points will be considered for final prediction. Euclidian distance will be used to find the distance between new point and every training point.

Support vector machines (SVM) are supervised learning models which can be used for both classification and regression. SVM models can also be used to predict crops and fertilizers suitable for the soil [16], [19]–[21]. Hyperplane is a line in SVM models to classify the data points. The data points nearer to the hyperplane will be considered as support vectors. The hyperplane will be used to categorise the data points into classes. Support vector regressions will try to find the best fit line which has maximum points.

Deep neural networks (DNN) are well known to solve the agricultural problems, deep neural networks can be used to recommend crops and fertilizers relevant for suggested crops [22]–[24]. Deep neural models use statistical and predictive methods to analyse and learn to produce the output. Artificial neural network (ANN) is a kind of deep learning which imitates human brain and neurons. System uses many layers of nodes to derive high-level functions from input information. Input layer, hidden layers and output layer are present in deep neural network.

Ensemble models can be used for predicting crops and fertilizers [25]–[27] ensemble technique works on few or several base models where each base model gives its own result and final prediction will be the average or mean of all the base models result. Sometimes different type of base models are used in ensemble system. Ensemble model is meant to improve the performance of any system and they are robust [28].

3. ARCHITECTURE & METHODOLOGY

3.1. Dataset

The paddy growing soil data of Mandya district of Karnataka-India has been collected from soil health cards given by government of India through website, 3094 samples of soil testing data having values of soil parameters such as pH, electric conductivity (EC), organic carbon, nitrogen (N), phosphorous (P), potassium (K), sulphur, zinc, boron, iron, manganese, copper along with soil fertilizers and nutrients like diammonium phosphate (DAP), urea, potassium muriate, ammonium phosphate, farm yard manure, zinc sulphate, borax, ferrous sulphate, copper sulphate, manganese sulphate have been collected. The dataset is named as ‘fertilizer dataset’ and stored in comma separated values (CSV) format. Part of fertilizer dataset can be seen in the Figure 1, in which we can see the values or quantities of soil parameters, micro nutrients and macro nutrients in kilograms per hectare of paddy growing soil.
3.2. Methodology

The fertilizer recommendation is the part of our research work ‘Soil fertility and crop friendliness detection and monitoring system using AI’ [29], [30]. The block diagram for the same is shown in Figure 2. We use artificial intelligence and machine learning techniques along with ‘internet of things’ features to predict the fertility rate of the soil, yield and recommend the right crops which can be grown in that soil. After suggesting the right crops, the right fertilizer or nutrients with optimum quantities needed for the suggested crops are also recommended. All the prediction results can be stored in cloud and recommendation report are sent to farmers in an easily readable and understandable format.

Nutrition recommender is the important module of our work, where the fertilizers and nutrients required and their quantities in kilograms per hectare for the suggested crops are recommended. The crop recommender module suggests the crops relevant to the soil parameters in the proposed system. Nutrition recommender system gives the optimum amount of nutrition and fertilization necessary for the suggested crops in detailed manner.

Fertilizers are not just meant for filling the necessary portion of soil, they directly affect the health of soil. Excessive use of fertilizers will degrade the soil properties. When optimum levels are given along with the combination of all the nutrients required, it will be really beneficial for farmers to grow the healthy crops as well as to maintain soil health. Like other fertilizer recommenders existing or proposed in other works, our nutrition recommender module will not just give a fertilizer name suggestion to farmers but is expected to give the customized report to each soil sample given, report of nutrition recommender will be the combination of fertilizers and their quantities needed for the particular soil and micro nutrients and their quantities along with organic manure quantity.
The nutrition recommender module is implemented using bootstrap aggregation method where several sub models or base models work to give the result. Each sub model is a ‘decision tree’. Each decision tree gets a portion or some samples of training data of nutrition dataset to get trained, all the sub models in the system get the reshuffled part of the samples and thus each sub model (decision tree) gets trained with different samples of nutrition dataset. One sub model's samples may or may not be same as that of another sub model’s samples. This process is called as bootstrap method.

Whenever a new value of fertilizer has to be predicted or test data has to be checked, all the decision trees give the output which will be aggregated. Hence average value of all the predictions become the final prediction value and this procedure can be referred as aggregation. The overall process of training several sub models and getting average result is called as bootstrap aggregation. So, recommender system forms an ensemble model with trees of random forest. This is illustrated in the Figure 3. The fertilizer recommender is not just meant to give a category of fertilizer. Since we need combination of fertilizers and their quantities, so we need to obtain the predictions of continuous dependent variables from numerous independent soil parameters therefore multioutput regressions are performed. For example, if the model has to predict the optimum level of ‘DAP’ fertilizer for a given soil sample, each decision tree regression predicts the ‘DAP’ value and average of all the values will be taken and given as one final ‘DAP’ quantity to the farmer. Some regression models get trained well with dataset and fail to perform well on test samples or new predictions, this issue is called as ‘overfitting problem’ in machine learning. Since each sub model gets trained on random samples in bootstrap aggregation, the overfitting issue can be overcome.

Ensemble technique has several advantages, so is the bootstrap aggregation method. Performance becomes the important factor for any system. Since the final prediction is the average of multiple predictions in ensemble technique, high performance can be expected. The variance can be reduced and the model overcomes the problem of overfitting. Random forest is robust to outliers and believed to work on highly nonlinear datasets.

![Bootstrap aggregation method used in fertilizer recommender module](image)

The necessary libraries and packages are loaded into out colab file. The nutrition recommender module is implemented in colab platform, the nutrition dataset is read into our colab file. Data is normalised to put the data in the common scale so that system accuracy will be more. Later parameters are set to input and target. Data is split as 80% training and 20% testing data then regression models are run and data is being fit to the model. Predictions are performed, evaluation metrics are derived and analysed along with plotting the same on graphs. R² Score value and mean squared error is analysed to see the efficiency of the model.

4. RESULTS

When soil sample is given or parameters are known to farmers, the system suggests the best crops which can be grown and also the combination of fertilizers and nutrients along with their necessary levels for
A bootstrap aggregation approach for adequate crop fertilizer and nutrition (Varshitha D. N.)

each suggested crop. As discussed earlier paddy growing soil dataset of Mandya is used for training the model. To analyse the regression models, there are few evaluation metrics used. Mean squared error, root mean squared error, mean absolute error and R squared are some important evaluation metrics. R squared score or $R^2$ also referred as score in the below result graphs is the best evaluation metric, $R^2$ is a statistical method used to evaluate the performance of the machine learning regression models. Score is nothing but coefficients of determination. For perfect prediction the $R^2$ score value will be approximately equal to 1. If the $R^2$ score is ‘0’ the model will not perform well at all. We have got the $R^2$ scores for predicting the quantities of DAP, potassium muriate, urea, farm yard manure as ‘0.99’ also for predicting the zinc sulphate, borox, ferrous and copper sulphate scores as ‘0.97’, hence the model can be considered the best one. $R^2$ is calculated using the formula given (1).

$$R^2 = 1 - \left(\frac{\text{Sum of Squared regression}}{\text{Total sum of squares}}\right)$$

(1)

The test data prediction results are plotted on graphs for fertilizers and nutrients combination of paddy growing soil as shown in Figure 4 and Figure 5. Also, $R^2$ score, mean squared error and root mean squared error results are captured. The combination of DAP, urea, potassium muriate along with micro nutrients like zinc sulphate, ferrous sulphate, copper sulphate and manganese sulphate and their quantities to be used in kgs/hectare to farmers are recommended. Red lines indicate that the predicted values and blue dots indicate original values in scaled form of data. 20% of testing data makes around 600 rows or samples of data for testing which is shown in the graph. We can see every red line meets the blue dots which implies that data predicted is almost accurate to that of original values.

Figure 4. The test data multi-regression results of paddy crop fertilizers

K-nearest neighbour and support vector machine are popular, simpler and effective machine learning models used in classification and regressions, deep neural networks are flexible models in building solutions for regression or classification problems. We have also implemented regression methods like KNN, deep neural network (DNN) and support vector regression (SVR) using same dataset for few fertilizers to see the performance. We can see that out bootstrap aggregation methods have highest $R^2$ score (score) among all the methods. The graphs of test data predictions of KNN, DNN, and SVR are also plotted and shown in the Figure 6. A part of predicted test data values can be seen in Figure 7 and Figure 8 where the grey shaded portion indicates the original
values and left hand side has predicted values of fertilizers and nutrients for paddy crops. The $R^2$ score (score) comparison of all the implemented models along with our bootstrap aggregation model is shown in Figure 9.

Figure 5. The test data muti regression results of paddy crop nutrients

Figure 6. Test data results of urea prediction of KNN, SVR, and DNN

Figure 7. A part of predictions vs real values of fertilizers-testing data
5. CONCLUSION

By using our fertilizer recommender, farmers can get custom report of all the fertilizers in combination with their quantities which can be very beneficial to them for better agricultural outcomes. If optimum level of fertilization is done in agricultural fields, it will result in healthy crop growth, yield can be enhanced and also the income of the farmers will be increased. Fertilization in adequate quantity will be helpful in maintaining soil health. The fertilizer recommender is implemented using ensemble technique known as bootstrap aggregate method which is nothing but a random forest of several decision trees. Using multiple output regressions, we are able to obtain the combinations of nutrients. We have also checked with other models and obtained the results. We obtained a R² score of 0.1 for support vector regression. For deep neural networks and KNN models, we got R² values of 0.57 and 0.72 respectively which correspond to average performance results and bootstrap aggregation method with 0.99 as R² score demonstrates the best performance. We can conclude from graphs and evaluation metrics that our proposed fertilizer recommender provides more accurate results than other methods. For fertilizers and nutrition recommendations, this can be considered as the best approach. Dataset can be expanded in the future to include more fertilizer combinations, as well as options for farmers to select fertilizer combinations can be thought of. Model can be trained with more number of crops, if the research is continued.

REFERENCES


**BIOGRAPHIES OF AUTHORS**

**Varshitha D. N.** is a research scholar at Sir M. Visvesvaraya Institute of Technology, Bengaluru, and Karnataka, India. She has 6 years of teaching and research experience. She holds a Master degree in Computer science and Engineering from VTU PG centre, Mysuru and Bachelor degree in Information Science and Engineering from NIEIT, Mysuru, Karnataka, India. She has published and presented papers in many international journals and conferences. Her research interests include data mining, machine learning, deep learning and DBMS. She can be contacted at email: varsha.swara@gmail.com.

**Dr. Savita Choudhary** is an associate professor working at Sir M. Visvesvaraya Institute of Technology, Bengaluru, Karnataka. She has been awarded Ph.D. for her research work in the area of soft computing from Bhagawanth university, Rajasthan. She has done Masters in Software Engineering from Banasthali university, Rajasthan, and Bachelor of Engineering from university of Rajasthan, India. She has 12+ years of teaching and research experience, she has guided several UG students and currently guiding four Ph.D. students. She has published and presented papers in many international journals and conferences. Her research interests include soft computing, machine learning, and IoT. She can be contacted at email: choudhary7.mvit@gmail.com.