**PREDICTION OF SOIL NUTRIENTS USING VARIOUS REGRESSION METHODS**

***1Madhumathi R, Department of Computer Science and Engineering,***

***Sri Ramakrishna Engineering College, Coimbatore, India***

***Email:*** [***madhumathi.r@srec.ac.in***](mailto:madhumathi.r@srec.ac.in)

***2Arunmuganthan T,  
ICAR-Sugarcane Breeding Institute, Coimbatore, India***

***Email:***[***arumuganthan@gmail.con***](mailto:arumuganthan@gmail.con)

***3Dhanishtaa R, Department of Computer Science and Engineering,***

***Sri Ramakrishna Engineering College, Coimbatore, India***

***Email:*** [***dhanishtaa.2101038@srec.ac.in***](mailto:dhanishtaa.2101038@srec.ac.in)

***4Elakkiya E, Department of Computer Science and Engineering,***

***Sri Ramakrishna Engineering College, Coimbatore, India***

***Email:*** [***elakkiya.2101051@srec.ac.in***](mailto:elakkiya.2101051@srec.ac.in)

***5Gunashri.R, Department of Computer Science and Engineering,***

***Sri Ramakrishna Engineering College, Coimbatore, India***

***Email:*** [***gunashri.2101059@srec.ac.in***](mailto:gunashri.2101059@srec.ac.in)

**ABSTRACT**

**Soil nutrient prediction can be used as a key input to increase crop yield. The nutrient present in the soil plays the major roles in the healthy growth of the plants. There are many nutrients present in the soil such as ca, mg, s, …, etc. In this paper we are majorly focusing on the three macro nutrients present in the soil such as Nitrogen(N), Phosphorus(P), and Potassium(K) which are needed for healthy plant growth. Farmers do not have sufficient knowledge about the nutrient present in the soil. So, they apply large quantity of fertilizers in their field which leads to the depletion of nutrient content in the soil. Laboratory soil test is time consuming. Hence the goal of our project is to predict the soil nutrient using Machine Learning technique.** **The current study is to predict the fertility indices for soil nutrients such as Nitrogen, Phosphorous, Potassium using the regression method which are available. Specifically, there are 76 regressors which belonging to 20 families, including neural networks, deep learning, support vector machine regression, random forests, lasso and ridge regression, Bayesian models and so on.**

**Keywords- agriculture, machine learning Regression, Nutrients, Dataset, NPK.**

1. **INTRODUCTION**

India stands second to have largest population in the world. Due to the increase in population the basic needs such as food, cloths, water, and more increases day by day. Agriculture is considered, as the primary source of food production. Around 70 percent of rural areas in India depends primarily on agriculture for their live hood [1].

Due to urbanization and industrialization the number of cultivable lands is decreasing. So, there is a need to increase the agricultural yield in a continual way without mistreating the environment. This is achieved by increasing the soil fertility by supplying essential nutrients in a sufficient amount at a right time to yield the healthy growth of the crops.

Mostly, the already existed approaches use machine learning (ML) to estimate the soil nutrient, but very little work has been done to predict region-specific soil nutrient based on soil parameters. Parameters such as soil type, soil nutrients (nitrogen, phosphorus, and potassium), micronutrients (iron, boron, and manganese), moisture, pH impact on the crop cultivation. Since the parameters differ for every zone [2].

Today smart farming is so evolved and influenced in the agriculture. It can also give the productive crops for the people. The data captured and exaggerated the role of machine learning algorithms in paddy rice production, by implementing the applications of machine learning strategists’ paper also sum up all the machine learning algorithms and gives outline to the overall management for the crop production. Smart farming helps farmers to get better understanding about the important factors such as water, topography, aspect, vegetation, and soil types.

1. **RELATED WORK**

The soil analysis can be achieved by various machine learning algorithms like extremely randomized regression trees which has a high accuracy. Some researchers evaluated the soil fertility for few villages by using the concept of pedo transfer function (PTF) [3].

Predict the value of the SOM (Soil Organic Matter) and Ph for the assessment of soil fertility. The vis-NIR Spectra to inquire the wavelength of the SOM and Ph of the soil [4]. The work is to combine the ML with the computer vision to analyse the crops images and checking the growth of the crop and quality of the crop [5].

The SWRC (soil Water Retention Curve) which is based on the SWRC curve. It is used to determine the conductivity, shear strength, and volume differences of unsaturated soils [6].

The authors used the —Boosted regression tree (BRT), crop yield prediction, Gaussian process regression (GPR)technique with the NVDI (normalized difference vegetation index) dataset to categorize the crop quality in the maize field [7].

By using DL models like SVM models the authors can develop the fertility level in the index positions for the soil organic carbons and soil nutrients level. The MRFO algorithm were used to boost the predictive performance of DL models and data splitting can be done into trained dataset and test dataset the final performance can be measured by the fusion models [8].

The feature extraction is the fed for all ML algorithms which are used to predict the suitable soil crops. the amin contribution is work with the modified recursive feature elimination (MRFE) technique to select key features for feature extraction and applying ML algorithm for prediction [9].

For the faster classification and prediction results using extreme learning algorithm. They used the neural network concepts to get the better results and prediction over the statistical method to analyse the overall performance of the soil they used the different DL models like (Gated Recurrent Units), DBN (Deep Believe Network). MRFO (Manta Ray Foraging Optimization) is the model which has a ability to predict the soil nutrient and pH level with high precision [10]. the authors examine the NVDI concept to find the feasibility of the sensed data and they developed the set of algorithms to formulate the changes in the model.[11]

The authors Ronny Hansch, Tomas Jadhuber have used the ML approaches for train and testing the data for the better accuracy. the used the RF(Random Forest ) algorithm to identify the soil permittivity of the given soil .the classify the ranges of soil permittivity for the different soils and predict the high volume flow in the soil .[12]Soil moisture is an important indicator which is used in analysing soil fertility. Recently, NASA used the microwave sensor to analyse the soil moisture with high precision. The global navigator satellite system reflectometry is a new technology which uses the GNSS signal which has the reflected property the surface produces the reflected signal that can be relevant to the permittivity the soil moisture can be detected by the strength of the reflected signal.[13]The aim of the work is to use the different dataset and various pre-processing techniques used for the spectroscopic method(VNIR-SWIR) to analyse the soil organic matter (SOC) and soil organic carbon level in the soil.[14]the VIS-NIR method is used to predict the potassium level in the soil. But most of the regression models provide the worst performance in the case of prediction. the researchers used the boosting algorithms to improve the performance of the algorithms [15].

1. **PROPOSED MODEL**

In this project, several machine learning algorithms like Naive Bayes, Random Forest, Support vector machine. The dataset has been collected. Among dataset they are trained dataset and testing dataset. The dataset contains macro nutrients like nitrogen, phosphorous, potassium and micronutrients like zinc, iron, copper etc. The code is implemented in Collab using python language. The aim is to predict the soil fertility using machine learning algorithm.

Algorithm

1.Start

2.Collect the dataset.

3.Read N, P, K values.

4.Import the Library.

5.Fitting the algorithm to training set.

6.X-> Independent variable

7.Y->Dependent Variable

8.Divide the dataset into training and testing data.

9.Divide

X\_train, X\_test, Y\_train, Y\_train.

10.Predict the N, P, K result.

11.Visualize the model.

12.Stop.

**A) Random Forest Algorithm**

Random forest is one of the supervised learning Algorithm. It groups various decision tree and join to get a more accurate and constant prediction. It is based on the basics of ensemble learning. The Algorithm help to reduce the problem of over fitting. Based on averaging it will predict the result.

1. Data pre-processing step.
2. Import the library.
3. Import the dataset.
4. Import the algorithm to the training set.
5. Extract the Independent and Dependent variable.
6. Divide the dataset into training and test set.
7. Predict the test result.
8. Visualise the graph.

Among the regression algorithm, Random Forest gives more accuracy than other algorithm.

**B) Decision Tree Algorithm**

The algorithm has both vertical and horizontal lines that are dividing the dataset according to the variable. In attribute selection measures (ASM) information gain and Gini index techniques were used.

1. Data pre-processing step.
2. Import the library.
3. Import the dataset.
4. Import the algorithm to the training set.
5. Extract the Independent and Dependent variable.
6. Divide the dataset into training and test set.
7. Predict the test result.
8. Visualise the graph.

**C) Support Vector Machine**

The aim of the SVM algorithm is to make the best line that can divide n-dimensional space into classes. The support vector used to decide the position of the optimal hyperplane. It is based on statistical learning theory.

1. Data pre-processing step.
2. Import the library.
3. Import the dataset.
4. Import the algorithm to the training set
5. Extract the Independent and Dependent variable.
6. Divide the dataset into training and test set.
7. Predict the test result.

8. Visualise the graph.

**D)Naïve Bayes Algorithm**

Naïve Bayes is one of the supervised algorithms that is based on the Bayesian algorithm to perform classification. It assumes that the occurrence of a particular attribute is independent of the occurrence of other attributes. It is an approach used for Classification due to its performance. The algorithm will predict based on the probability of the value.

1. Data pre-processing step.
2. Import the library.
3. Import the dataset.
4. Import the algorithm to the training set.
5. Extract the Independent and Dependent variable.
6. Divide the dataset into training and test set.
7. Predict the test result.

ml algorithm

1. Visualise the graph.

**E) Logistic or Ridge Regression Algorithm**

naive bayes

Logistic Regression Algorithm is one of the supervised machine learning. It helps to predict the categorical dependent variable using the dataset of independent variable. It is based on the concept of predictive modelling. The dependent variable should be categorical in nature. The independent variable should not have multi collinearity.

output prediction

1. Data pre-processing step.
2. Import the library.
3. Import the dataset.
4. Import the algorithm to the training set.
5. Extract the Independent and Dependent variable.
6. Divide the dataset into training and test set.
7. Predict the test result.
8. Visualise the graph.

The dependent variable should be categorical in nature. The independent variable should not have multi collinearity.

Dataset collection

Analyse and pre-processing the data

                                             
  
                                 

Train data

Soil fertility prediction

Naïve Bayes

Vector machine

Random forest

Test and evaluate the data

1. **RESULTS**

By using different machine learning algorithms, the various metrics accuracy has been predicted.

Decision tree:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | F1-score | support |
|  | 0.88 | 0.74 | 0.80 | 38 |
|  | 0.44 | 0.67 | 0.53 | 12 |
| accuracy |  |  | 0.72 | 50 |
| Macro avg | 0.66 | 0.70 | 0.67 | 50 |
| Weighted avg | 0.77 | 0.72 | 0.74 | 50 |

Naïve Bayes:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | F1-score | support |
|  |  | 1.00 | 1.00 | 5 |
|  | 0.80 | 1.00 | 0.89 | 4 |
| accuracy | 1.00 | 0.95 | 0.98 | 21 |
| Macro avg |  |  |  |  |
| Weighted avg |  |  |  |  |

Random forest:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | F1-score | support |
|  | 0.88 | 0.76 | 0.82 | 38 |
|  | 0.47 | 0.67 | 0.55 | 12 |
| accuracy |  |  | 0.74 | 50 |
| Macro avg | 0.67 | 0.71 | 0.68 | 50 |
| Weighted avg | 0.78 | 0.74 | 0.75 | 50 |

Support vector machine:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | F1-score | support |
|  | 0.97 | 0.79 | 0.87 | 38 |
|  | 0.58 | 0.92 | 0.71 | 12 |
| accuracy |  |  | 0.82 | 50 |
| Macro avg | 0.77 | 0.85 | 0.79 | 50 |
| Weighted avg | 0.87 | 0.82 | 0.83 | 50 |

1. **CONCLUSION**

Agriculture is a backbone of the Indian economy. Which is depends upon various parameters like Ph, Rainfall etc…. By considering results over all the soil dataset, random forest classifier algorithm is a most efficient algorithm with better accuracy. The soil fertility can be analysed by this algorithm and various macro nutrients like Nitrogen, Phosphorous, Potassium and some other micronutrients were predicted by using different parameters.

**REFERENCES**

[1] Li Gong, et.al “An IoT-based intelligent irrigation system with data fusion and a self-powered wide-area network”, Journal of Industrial Information Integration, Elsevier, vol. 29, pp. 1-9, 2022.

[2] G. Mariammal, A. Suruliandi, S. P. Raja, and E. Poongothai “Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique with Various Classifiers”, IEEE transactions on computational social systems, VOL. 8, IEEE, 2021.

[3] M.S. Sirsat, E. Cernadas, M. Fernández-Delgado, S. Barro “Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods”, Computers and Electronics in Agriculture, 154, pp. 120–133, 2018.

[4] Meihua Yang. Dongyun Xu, Songchao Chen, Hongyi Li and Zhou Shi “Evaluation of Machine Learning Approaches to Predict Soil Organic Matter and pH Using vis-NIR Spectra”, Sensors, MPDP, pp. 2-14, 2019.

[5] Abhinav Sharma, Arpit Jain, Prateek gupta, and Vinay Chowdary “machine Learning Applications for Precision Agriculture: A Comprehensive Review”, IEEE Access, volume 9, 2020.

[6] Kevin Achieng “Modelling of soil moisture retention curve using machine learning techniques Artificial and deep neural networks vs support vector regression models”, Computers and Geoscience, Research gate, 133, 2019.

[7] Hossein Aghighi, Mohsen Azadbakht, Davoud Ashourloo, Hamid Salehi Shahrabi, and Soheil Radiom “Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI” IEEE journal of selected topics in applied earth observations and remote sensing, vol. 11, pp. 12, IEEE, 2018.

[8] José Escorcia-Gutierrez, Margarita Gamarra Roosvel Soto-Diaz, Meglys Pérez, Natasha Madera and Romany F. Mansour” Intelligent Agricultural Modelling of Soil Nutrients and pH Classification Using Ensemble Deep Learning Techniques”, Agriculture, MPDI, 2022.

# [9] G. Mariammal; A. Suruliandi; S. P. Raja; E. Poongothai, “Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique with Various Classifiers”, IEEE transactions on computational social systems, vol. 8, issue 5, pp.1132, 2021.

[10] José Escorcia-Gutierrez, Margarita Gamarra, Roosvel Soto-Diaz, Meglys Pérez, Natasha Madera and Romany F. Mansour “Intelligent Agricultural Modelling of Soil Nutrients and PH Classification Using Ensemble Deep Learning Techniques”MPDI,2022**.**

[11] Michael L. Mann, James M. Warner & Arun S. Malik “Predicting high-magnitude, low-frequency crop losses using machine learning: an application to cereal crops in Ethiopia”, climate change, pp.211–227,2019.

[12] [Ronny Hänsch](https://ieeexplore.ieee.org/author/37400918500), [Thomas Jagdhuber](https://ieeexplore.ieee.org/author/37400470000), [Benjamin Fersch](https://ieeexplore.ieee.org/author/38506891800) “Soil-Permittivity Estimation Under Grassland Using Machine-Learning and Polarimetric Decomposition Techniques” [IEEE Transactions on Geoscience and Remote Sensing](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=36) , Volume 59, [Issue 4](https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=9386007&punumber=36), pp 2877 – 2887 April 2021,

# [13] [Yan Jia](https://ieeexplore.ieee.org/author/37085709558), [Shuanggen Jin](https://ieeexplore.ieee.org/author/37896250800), [Haolin Chen](https://ieeexplore.ieee.org/author/37089395773); [Qingyun Yan](https://ieeexplore.ieee.org/author/37085812179" \t "_blank), [Patrizia Savi](https://ieeexplore.ieee.org/author/37296452700), [Yan Jin](https://ieeexplore.ieee.org/author/37086213918), [Yuan Yuan](https://ieeexplore.ieee.org/author/37088405424) “Temporal-Spatial Soil Moisture Estimation from CYGNSS Using Machine Learning Regression With a Preclassification Approach” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE, VOL. 14, pp.4870-4880 2021.

**[14]** [Theodora Angelopoulou](https://sciprofiles.com/profile/592116), [Athanasios Balafoutis](https://sciprofiles.com/profile/614706),[George Zalidis](https://sciprofiles.com/profile/898360),[Dionysis Bochtis](https://sciprofiles.com/profile/171786),” From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation—A Review”,  [Sustainability](https://www.mdpi.com/journal/sustainability),MPDI,[Volume 12](https://www.mdpi.com/2071-1050/12) ,[Issue 2](https://www.mdpi.com/2071-1050/12/2),pp.2-24,2020

# [15] Jin, X.; Li, S.; Zhang, W.; Zhu, J.; Sun, J. “Prediction of Soil-Available Potassium Content with Visible Near-Infrared Ray Spectroscopy of Different Pretreatment Transformations by the Boosting Algorithms.” *Appl. Sci.*,MPDI, VOL 10,issue 4,2020