Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 6, Issue 12, December -2019

Soil Health Analysis for Crop Suggestions using Machine Learning

Pratiksha Yadav[1], Prof. Prashant Ahire[2]

[1] Computer Engineering, Dr. D. Y. Patil Institute of Technology
[2] Computer Engineering, Dr. D. Y. Patil Institute of Technology

Abstract — *Indian economy is depending on agriculture. Agriculture is the main source of income for most of the population. So farmers are always curious about yield prediction. Many factors are responsible like soil, weather, rain, fertilizers and pesticides to increase yield production. Agriculture being a soil-based industry, an increase in yield can only be attained by ensuring that the soil provides a balanced and an adequate supply of nutrients. Soil testing is pivotal in understanding the deficiencies in soil and avoiding nutrient imbalance. This survey and study focuses on the different soil types, crop types and soil test reports. Soils are complex mixtures of air, water, minerals, organic matter, and countless organisms that are the decaying remains of once-living things. We can say soil is an important ingredient of agriculture. There are several types of soils and each type of soil can have different kinds of features and different kinds of crops grow on different types of soils. We must know which type of crop is go better in our soil. We can apply machine learning techniques to classify soil and to predict the crop suitable.*

Keywords-*: Soil series; Chemical factors; KNN; ID3;K-means.*

I. INTRODUCTION

The Government of India has launched many schemes for the welfare of the farmers. But these schemes have some shortcomings. Enhancement of food technology is today's need. Huge no of people live in India. So ratio and proportion of food and humans has no toning, resulting in high rates of inflation. Agriculture is totally dependent on the soil quality but as time passes more agricultural production results in the loss of nutrients present in the soil. We have to identifying techniques that will slow down this elimination of nutrients and also will return the required nutrients with the soil, so that we keep getting high quality and good quantity crop productions. There are so many soil series available in India. Every soil series have different features and every soil is suitable for different crop. Sometimes or we can say every time it happens that farmer soil is best for some specific crop but as he don't know. We need to know the characteristics of soil types to understand which crops grow better in certain soil types. Machine learning techniques can be helpful in this case**.** The main purpose of the proposed work is to create a suitable model for classifying various kinds of soil series data along with suitable crops suggestion. Series are recognized by machine learning methods using various chemical features and possible crops for that soil series are suggested using geographical attributes.
In the system, the agriculture department officers are assigned with a area in particular district. It is the responsibility of these officers to collect the soil samples from the farmers of the given area. These soil samples are then tested in the laboratory and based on the results of the soil test report recommendations, like the type and amount of fertilizer suited to the crop, are generated by experts in the Department of Agriculture.

II. MOTIVATION AND OBJECTIVES

Farmers fall into debts because they have to face a scarce crop productivity, which increases the risk of their profit. This creates a vicious cycle and most of the time farmers suffer mental distress. Without knowing the type of soil and suitable crop, Farmer plants his farm and Loss due to wrong crop and Waste of money. Hence, we motivate to help solve farmer's issues.
The key motivation for developing this project is as we say every part of world is developing but we can see that there is no such big achievement or development in soil or crop related issues. So we can give preference to this soil field and if we suggest suitable crop to farmers then it is beneficial for them.
1.The main goal of this project is to classify soil series and predict suitable crop for that soil.
2.To improve crop productivity by recommendation of soil fertilizers.
3.Continuously, test of soil and then give suggestions about crops.

III. LITERATURE SURVEY

Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M.Mohidul Islam [1] conducted a study to predicting soil series and providing suitable crop yield suggestion for that specific soil. The model has been tested by applying different kinds of machine learning algorithm. Bagged tree and K-NN shows good accuracy but among all the classifiers, SVM has given the highest accuracy in soil classification. The proposed model is justified by a properly made dataset and machine learning algorithms.

Pramudyana Agus Harlianto, Noor Akhmad Setiawan, Teguh Bharata Adji [2] conducted a study of machine learning algorithms such as neural network, decision tree, naïve bayes, and SVM could be used to automate soil type classification with satisfactory accuracy (> 70%). using standard algorithm in RapidMiner, SVM is the best performance for classifying soil type. Attribute selection did not improve SVM's accuracy, as well as class reduction did not improve significantly.
Jay Gholap [5] conducted a study of Soil fertility is considered to be one of the critical attributes for deciding cropping pattern in particular area. J48 algorithm's accuracy for predicting soil fertility was highest. For increasing accuracy of J48 they use some other meta-techniques like attribute selection and boosting.
Vrushali Bhuyar [7] conducted a study of classification of soil fertility rate using J48, Naïve Bayes, and Random forest algorithm. J48 algorithm gives mostly good result than other algorithms. J48 algorithm in decision tree helps the farmer and decision makers to identify the the fertility rate of soil and on the nutrients found in the soil sample different fertilizers can be recommended.
K.K.Deshmukh [6] conducted a study of evaluate the soil fertility status from Sangamner area, Ahmednagar district, Maharashtra. Classification of soil in different categories based on different quality of soil by evaluation of the soil attributes. soil samples were analyzed for various soil fertility parameters by standard procedures. Improper agriculture knowelge loss the productivity. So, overcome the adverse effect, complimentary use of fertilizers, organic manures in suitable combination of chemical fertilizers were suggested.

IV. ALGORITHMIC SURVEY

A. ID3:

Decision tree ID3 is used **Iterative** to generate **Dichotomiser** a decision**3(ID3)** tree. It Classifies data using the given attributes. Tree consists of decision nodes and leafs. Nodes can have two or more branches which represents the value for the attribute tested. Leafs nodes produces a homogeneous result.
Algorithm:
Step 1.Create a root node for the tree
Step 2.If all examples are positive, Return the single node tree Root with label assign +.
Step 3.If all examples are negative, Return the single node tree Root with label assign -.
Step 4.If number of predicting attributes is empty, then return the single node Root, with label = most common value of the target attribute in the examples.
Step 5. Else
X = The Attribute that best classifies examples.
Decision Tree attribute for Root = X.
For each possible value, vi, of X,
Step 6.Add a new tree branch below Root, corresponding to the test X = vi.
Step 7.Let Examples (vi), be the subset of examples that have the value vi for X

B. k-nearest neighbors:

The k-nearest neighbors (KNN) algorithm is a supervised machine learning algorithm. That can solve both classification and regression problems. K-nearest neighbours (KNN) algorithm uses 'feature similarity' to predict the values of new data points which further means that the new data point will be assigned a value based on previous closed values it matches the points in the training set.
Algorithm:
Step 1.Take training dataset.
Step 2. Then, choose the value of cluster K. It means the nearest data points K is any integer.
Step 3.For each point in the test data do the following steps Calculate the distance between test data and each row of training data with the help of any of the method which are Euclidean distance, Manhattan distance.
But Euclidean distance method is simple so it is most commonly used.
Now, based on the distance value, sort that values in ascending order and give rank to each. then, it will choose the top K rows from the sorted array. at last, it will assign a class to the test point based on most frequent class of these rows.

C. K-means clustering:

k-means is one of the unsupervised learning algorithms that solve the clustering problem. The procedure follows a simple to classify a given data set through a certain number of clusters. The main idea is to assume k centers, one for each cluster. So, the place them as much as possible far away from each other. Then next step is to take each point belonging to a given dataset and allocate it to the nearest center. Then re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words

centers do not move any more. The K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.
Algorithm:
Step 1.First we initialize k points, called means, randomly.
Step 2.We categorize each item to its closest mean and we update the mean's coordinates, which are the averages of the items categorized in that mean so far.
Step3.We repeat the process for a given number of iterations and at the end, we have our clusters.

V. PROPOSED SYSTEM

The main purpose of the proposed work is to create a suitable model for classifying various kinds of soil series data along with suitable crops suggestion .Soil is an important ingredient of agriculture. There are several kinds of soil. Each type of soil can have different kinds of features and crops grow on different types of soils. We need to know the characteristics of various soil types to understand which crops grow better in certain soil types. Machine learning techniques can be helpful in this case. Here we can use clustering technique to group data, and then classified the data by the order of soil and places with algorithm. Then apply apriority Mining process to generate an association rule for finding suitable crops for the specific soil. Soil series and land type combine represents the soil class in the database. The machine learning methods are used to find the soil class. Different methods are used: K-means clustering, K-NN and Naive bayse.
In that system many Farmers can registration at a time. When Farmer register that time Agriculture officer gets Email notifications. Then Agriculture officer takes results of soil sample test from laboratory. Then that result store in soil database. Do some soil analysis on that soil database and suggest which crops are best for that soil.
In that process we consider Soil Attributes are a major source of nutrients for plant growth. Nutrients supplied by the soil are called mineral nutrients. The non-mineral nutrients such as carbon (C), hydrogen (H), oxygen (O) come from air and water during photosynthesis. Soil mineral nutrients are divided into two groups the macro and micronutrients. The macronutrients are further divided into two groups the primary and the intermediate nutrients. The primary nutrients are required by plants in relatively large proportions. These are the most famous; the nitrogen (N), sulphur (S), phosphorus (P) and potassium (K) commonly referred to as NPK. The intermediate nutrients are required by plants in medium quantities, these are magnesium (Mg), calcium (Ca).
Also consider moisture and weather conditions. On the base of these factors our system recommend best crop for soil and improve crop productivity by recommendation of soil fertilizers. Continuously, suggest test of soil and then give suggestions about crop.

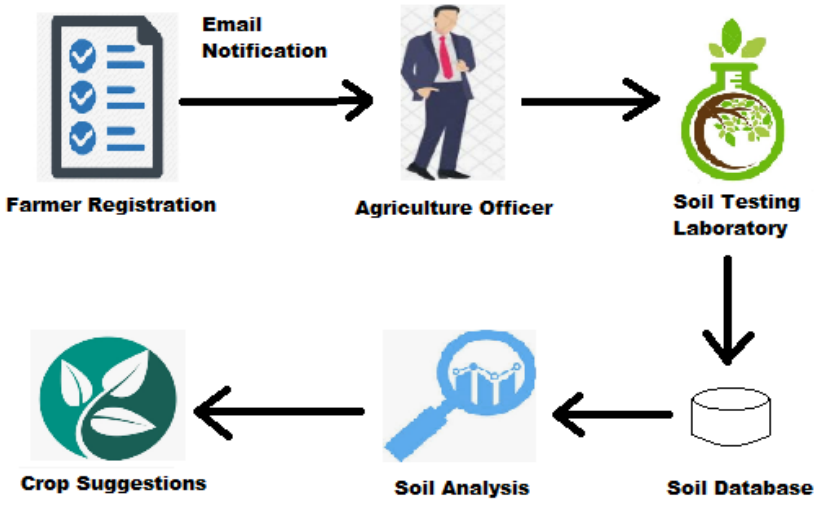

Figure 1. Architecture Diagram

REFERENCES

[1] Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M.Mohidul Islam "Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series", International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018.
[2] Pramudyana Agus Harlianto, Noor Akhmad Setiawan, Teguh Bharata Adji "Comparison of Machine Learning Algorithms for Soil Types Classifications", International Conference of Science And Technology - Computer (ICST), 3, 2017.

[3] Ramesh, D., & Vardhan, B. V. "Data mining techniques and applications to agricultural yield data", International journal of advanced research in computer and communication engineering, 2(9), 3477-3480 2013.
[4] Bhargavi P. & Jyothi S., "Soil Classification Using Data Mining Techniques: A Comparative Study", International Journal of Engineering Trends and Technology, 2011.
[5] Jay Gholap, "Performance Tuning Of J48 Algorithm For Prediction Of Soil fertility", Asian Journal of Computer Science and Information Technology,Vol 2,No. 8 2012
[6] K.K.Deshmukh, "Evaluation Of Soil Fertility Status From Sangamner Area, Ahmednagar District, Maharashtra, India", Rasayan Journal of Chemistry, 2012.
[7] Vrushali Bhuyar, "Comparative Analysis Of Classification Techniques On Soil Data To Predict Fertility Rate for Aurangabad District", International Journal of Emerging Trends & Technology in Computer Science, Apr 2014.
[8] Department of Agriculture & Cooperation Ministry of Agriculture Government of India, "Methods Manual-Soil Testing in India", 2011.
[9] B.V.RamaKrishna, Dr B.Satyanarayana, "Agriculture Soil Test Report Data Mining for Cultivation Advisory", International Journal of Computer Application Volume 6– No.2, March-April 2016.
[10] Elma Hot, Vesna Popović-Bugarin, "Soil Data Clustering by Using K-means and fuzzy K-means Algorithm", Telfor Journal, Vol. 8, No. 1, 2016.
[11] Rachana P. Koli, V. D. Jadhav, "Agriculture Decision Support System As Android Application", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.

