International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-06, Issue-01, Apr 2020

Soil Quality Analysis and Crop Fertility Prediction

[1]Asst.Prof. Allan Lopes, [2]Mr. Amandeep Singh Bans, [3]Ms. Upasana Bhatt
[1]Asst.Professor, [2,3]UG Student, [1,2,3]Department of Information Technology, Universal College of Engineering, Kaman, Maharashtra, India.
[1]*allan.lopes@universal.edu.in*, [2]*amandeepbans@gmail.com*, [3]*bhattupasana1998@gmail.com*

Abstract Data Mining is a technique used to retrieve information for the analysis and discovery of hidden trends in large data sets. Data Mining extends to numerous areas such as education, banking, marketing, retail, communications and agriculture. Agriculture is the backbone of country's economy. It is the important source of livelihood. Agriculture depends primarily on the weather, geology, soil and biology. Agricultural Mining is a technology that can contribute information for the growth of agriculture. The current study presents the various techniques of data mining, and their role in soil fertility, nutrient analysis. Decision tree is a well-known Data Mining classification approach. C4.5 and Classification and Regression Trees (ID3) are two widely used decision tree algorithms for classification. The C4.5, ID3 and the proposed classifier have been trained using the soil sample data set by taking into account the optimal soil parameters pH (hydrogen power), EC (electrical conductivity) and ESP (exchangeable sodium percentage). The model is evaluated using a collection of soil samples test results. Classification of soil is the division of soil into classes or groups each having similar characteristics and likely similar behavior. Soil classification is easy to allow the farmer to know the type of soil and to plough the crops based on the soil type.

Keywords — C4.5, Data Mining, ID3, Prediction, Quality Analysis

I. INTRODUCTION

Overall, the need for food is intensifying on a daily basis, so agricultural scientists, farmers, government, and researchers are trying to improve production using diverse agricultural practices. As an impact, the information gathered from the agrarian data is increasing every day. Because the volume of information has been enlarged, this information needs a spontaneous method to be mined and analyzed once necessary. Data Mining techniques [5] will be used to predict agricultural process trends in the longer term. Data Mining techniques are of two types, one being descriptive, taking into account the existing data and the other being predictive, depending on the probability for future analysis.

Data Mining process involves [1]

a) Data collection, cleaning and loading into data warehouse system

b) Stores the data in multidimensional format

c) Provides access to information for analysts and strategy-makers

d) Analyzation of data using different applications

e) Presents the data using different patterns

Agricultural researchers and farmers deploy sensors in their far-off agricultural fields to obtain temperature, humidity, soil moisture and so on [7]. Automatic collection of those data helps their analytical work greatly. They are currently relying on network providers, such as cell phone networks, to achieve such automatic collection from their remote sites, which is not feasible for most farmers because of operating costs. The review and meta-analysis of yield data comparing organic and conventional agriculture showed that currently organic yields of individual crops are on average 80% of conventional yields. Analysis of 362 datasets also showed a high variation in organic farm yield gap (standard deviation of 21 per cent). Some of this variation would appear pervasive. Soybeans, some other pulses, rice and maize, for example, score more than 80% and wheat, barley and potato less than 80%. Most regions have fairly close relative yields to the overall average. Comparative study of current data mining techniques, such as cluster analysis and statistical methods, was conducted in a research conducted by Leisa J. Armstrong to establish the most effective technique. To conduct this research, they used a large data set extracted from the soils database of the Department of Agriculture and Food of Western Australia (AGRIC). The experiments analyzed a small number of features contained within the data sets to determine their effectiveness relative to standard statistical techniques.

 DOI : 10.35291/2454-9150.2020.0280

II. OBJECTIVE

Data mining methodologies are hugely populated in agriculture. Data mining encompasses the systematic assessment of huge information sets and data processing is an exciting and fashionable space for analysis in agricultural soil datasets [4]. The productive capability of a soil depends on soil fertility [4]. Today, data processing is employed in a very large areas and plenty of ready-to-wear data processing system product and domain specific data processing application software's are obtainable, however data processing of knowledge set of agricultural soil may be a comparatively young field of The objectives of this analysis were to describe and classify the soil in order to provide more accurate information on the nature of the soil [4]. The additional purpose of this analysis was to classify the estimated crops that could be grown in the region. This research was designed to describe and classify the soil in order to provide more detailed morphological details on the subsurface [3]. In addition, the pH meter is used to calculate the pH value and by using this pH value and algorithm [7], the N, P, K and Ec values are specified for further convenience [5].

III. LITERATURE SURVEY

1. Soil Toxicity Prediction and Recommendation System Using Data Mining in Precision Agriculture

The proposed method predicts and advises the farmer about the amount of toxicity present in the soil. Most farmers rely on rainfall, which is a factor in poor growth and reduces crop yields. The suggested program therefore recommends crops, soil fertility, toxicity and water supply to the farmer. The accuracy of the sensor recommendation scheme is very critical, as is the classification algorithm.

2. Crop Prediction Using Predictive Analytics

This research is to create a model for soil fertility testing. It also shows the crop to be planted, depending on the value obtained from the sensor. This also offers regionally sound information on the crop in the form of a graph. We have a farmer chat where the farmers can share and get an idea from the expert by registering for this application. It also indicates the nutrient that needs to be applied to the soil in order to improve crop production. It allows the farmer to evaluate the fertility of his farm and plant a better crop to increase productivity and income. This also provides details on the fertilizer to be applied to the soil and also provides information on the nearby fertilizer store.

3. Crop Recommendation System for Precision Agriculture

Illustrates The common problem among Indian farmers is that they do not choose the right crop on the basis of their soil requirements. Despite of this, they face a substantial decline in profitability. The problem of farmers has been solved by precision farming. Precision agriculture is a modern agricultural technique which uses soil characteristics, soil types, crop yield data collection data and recommends the right crop for farmers depending on their site-specific parameters. This eliminates the wrong crop option and increases productivity. In this paper, this problem is resolved by proposing a recommendation method via a majority voting ensemble model using Random tree to recommend a crop for site-specific parameters with high accuracy and efficiency.

4. Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data

The present research describes the thesis on soil classification analysis using the Support Vector Machine (SVM). Hyperion Hyperspectral [9] satellite data with 10 nm fine spectral resolution of the Phulambri area of the Aurangabad district of Maharashtra (India) was used for soil classification. Gaussian Radial Base Function (RBF) kernel SVM was used to extract five different soil types and obtained an overall accuracy of 71.18 per cent and a Kappa Value of 0.57 with appropriate training samples. It has been found that the soil of the area can be divided into five groups. The overall area (51 per cent) was brown sandy soil, while the minimum area (0.02 per cent) was gray clay soil. The result is of great significance for soil analysis of a very complex area without reducing the dimensionality of satellite data.

5. A Study on Various Data Mining Techniques for Crop Yield Prediction

This paper presents a study on the different data mining techniques used to predict crop yield. The performance of any crop yield forecasting program relies heavily on the precision of the features extracted and the correct use of the classifiers. This paper summarizes the results of various algorithms used by different authors to forecast crop yields, with their accuracy and recommendations.

IV. PROPOSED SYSTEM

The primary purpose of this method is to identify the soil according to the nutrients it contains. We have collected data sets of soil samples for this purpose. The soil will be categorized using the decision tree algorithm and the soil type will be shown [8]. We can also predict crops suitable for a specific form of soil. In addition, we can improve the soil if the farmer decides to plant a specific crop in the same soil by recommending the nutrient requirements for the same soil. To order to resolve these disadvantages, we suggest an updated decision tree approach to soil classification. Here, we first calculate the gini index for different sets of attribute values instead of calculating for each successive pair, as was done in the ID3 algorithm [8]. Instead we used the ratios of these calculated gini indices to

 DOI : 10.35291/2454-9150.2020.0280

reduce the bias caused by the information gain in the C4.5 algorithm. The algorithm proposed is listed below.

1) Create a root node that contains all the data set
2) Compute Gini index for various levels of pH, Ec and ESP attribute values.
3) Calculate the ratio of the gini indices for each attribute.
4) Select an attribute with the least gin ratio as the node to be broken.
5) Split the examples of the current node into different subsets based on values of the selected attributes.
6) Develop a new node as a child of the current node for each subset and move the examples to the node in the subset.
7) Repeat steps 2 to 5 recursively until further splitting is not feasible (only one instance remains in the node).

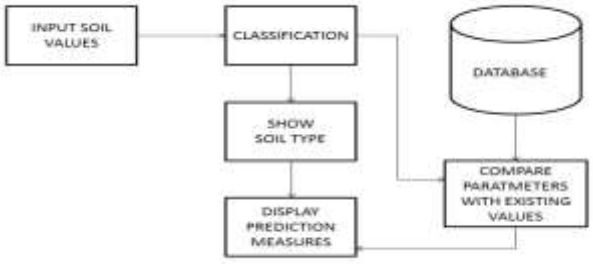

Figure 1: System Architecture

V. RESULT ANALYSIS

The website implemented for this project has login for the administrator to check if the farmer has posted any queries and login for the farmer to check the quality of the soil and see if the soil is fertile enough to grow crops.

1. Administrator Login: In this, admin register into the system.

Figure 2: Administrator Login

2. Admin Dashboard: Here, admin can delete the user, admin can upload training dataset, add farmers, check queries raised by the farmer, and check details of soil, crop and weather.

Figure 3: Admin Dashboard

3. User Login: Farmers can register with the help of the Aadhar card or phone number.

Figure 4: User Login

4. Soil and Rainfall Prediction: In this module, the uploaded dataset will test using ID3 algorithm and display the results [6].

Figure 5: soil and rainfall prediction

VI. ALGORITHMS AND METHODOLOGIES

The classification of the soil is the primary objective of this project. This can be done by using decision tree algorithm. The learning of decision trees is a method commonly used in data mining. A decision tree consists of an internal flow-chart structure, with a trial on an attribute for each (non-leaf) node, a test outcome for each branch (or terminal node) and a class mark for each leaf (or terminal). There are few decision tree algorithms available as follows:

- **C4.5 ALGORITHM:**

A decision tree T built through the offerings of the experimental data collection.

Step 1: Create the node N.

 DOI : 10.35291/2454-9150.2020.0280

Step 2: if instance belongs to the same class then return node N as the leaf node marked

with a class label.

Step 3: if attribute List is null, then return the node N as the leaf node signed with the

most common class.

Step 4: Select the attribute with highest information gain in the attribute list for splitting

purpose.

Step 5: Repeat the steps I to 4 until further splitting is not possible. [2]

- **CART ALGORITHM:**

Classification and regression trees (CART) are a non-parametric decision tree learning methodology that either generates classification or regression trees, depending on whether the dependent variable is either categorical or numerical. A decision tree T which is created by offering an experimental dataset.

Step 1: Construct a root node which contains the whole data set.

Step 2: Select an attribute that benefits the task most according to some criterion

computed using the data within the current node (gini index).

Step 3: Split the examples of the current node into different subsets based on values of

the selected attributes.

Step 4: Create a new node as a child of the current node for each subset and pass the

examples in the subset to the node.

Step 5: Recursively repeat steps 2 to 5 until further splitting is not possible (only one instance remains in the node). [2]

- **DATABASES USED:**

The various databases [6] used for this project are as follows:

	username	password
▶	admin	admin123
*	NULL	NULL

Table 1: Admin Login Credentials

2	Rice	500	1000	2
3	Wheat	450	650	2
4	Sorghum	450	650	1
5	Maize	500	800	1
6	Sugarcane	150	550	2
7	Cotton	400	700	1

Table 2: Crop Details

VII. FUTURE SCOPE

In the current work we are using datasets. In future, we plan to test the methodology by increasing the size of the field data with a greater number of parameters in soil to identify more details.

1. Obtaining the requirement of water by the soil and scheduling the watering to the crops.
2. The use of live data from Agriculture University.
3. Making an application for the farmers and then making it available in various languages as it can be more feasible to use.
4. Making it easy for farmers by incorporating this web application into a smartphone application.
5. The data will be preloaded so that farmers can access the application without the internet
6. Application of field boundary sensors to prevent animals from breaching the field.
7. Such sensors will be connected to the farmers mobile device to alert them.

VIII. CONCLUSION

Agriculture is of the utmost importance, particularly in a warming country like India. The use of information technology in agriculture will change the decision-making scenario, and framers will achieve better results. Classification of the soil using CART and C4.5 algorithm helps to know the overall soil quality and nutrient content in the soil. By using Predictive Analysis, we have determined what kind of soil the farmer can produce. This will help the farmer grow seasonally a variety of crops. When the farmer wants to grow a specific crop, he can enhance the soil by adding the required nutrients to the soil as the crop needs. Decision-making on overall problems relevant to agriculture; data mining plays a critical role. The survey addressed the role of data mining in the field of agriculture. We have addressed various types of soil, a variety of data mining techniques in agriculture and soil containment. With the aid of our program by using pH meter and algorithm, it will be useful for farmers to obtain estimation values at present. To decrease the need of the farmer to go to the laboratories and getting their soil checked they can just enter the pH value of the soil using the pH meter and then by the algorithm it gives the almost approximate

 DOI : 10.35291/2454-9150.2020.0280

N(Nitrogen), P(Phosphorous) and K(Potassium) values which can then be used to determine the fertility of the soil.

REFERENCES

[1] Monali Paul, Santosh K. Vishwakarma and Ashok Verma "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" 2015 International Conference on Computational Intelligence and Communication Networks.

[2] K Aditya Shastry, Sanjay H A and Kavya H "A Novel Data Mining Approach for Soil Classification" The 9th International Conference on Computer Science & Education (ICCSE 2014) August 22-24, 2014. Vancouver, Canada.

[3] Amol D. Vibhute, K. V. Kale, Rajesh K. Dhumal and S. C. Mehrotra "Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data" 2015 International Conference on Man and Machine Interfacing (MAMI).

[4] Sofianita Mutalib, S-N-Fadhlun Jamian, Shuzlina Abdul-Rahman and Azlinah Mohamed "Soil Classification: An Application of Self Organising Map and k-means" 2010 10th International Conference on Intelligent Systems Design and Applications.

[5] Y. Gandge and Sandhya, "A study on various data mining techniques for crop yield prediction," 2017 International Conference on Electrical, Electronics, Communication, computer, and Optimization Techniques (ICEECCOT), Mysuru, 2017, pp. 420-423.S.

[6] M. Pawar and G. Chillarge, "Soil Toxicity Prediction and Recommendation System Using Data Mining In Precision Agriculture," 2018 3rd International Conference for Convergence in Technology (I2CT), Pune, 2018, pp. 1-5.

[7] P. S. Vijayabaskar, R. Sreemathi and E. Keertanaa, "Crop prediction using predictive analytics," 2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC), Melmaruvathur, 2017, pp. 370-373.

[8] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system for precision agriculture," 2016 Eighth International Conference on Advanced Computing (ICoAC), Chennai, 2017, pp. 32-36.

[9] D. Vibhute, K. V. Kale, R. K. Dhumal and S. C. Mehrotra, "Soil type classification and mapping using hyperspectral remote sensing data," 2015 International Conference on Man and Machine Interfacing (MAMI), Bhubaneswar, 2015, pp. 1-4.

 DOI : 10.35291/2454-9150.2020.0280