International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

AgroTech: Soil Analysis and Crop Prediction

Rishikesh Gundla[1], Jaswinder Singh[2], Prajakta Gaikwad[3], Priti Bhole[4], Jyoti Dange[5]

[1,2,3,4]*Student, Dept. of Electronics and Telecommunication, Atharva College Of Engineering, Mumbai, India*
[5]*Professor, Dept. of Electronics and Telecommunication, Atharva College Of Engineering, Mumbai, India*

Abstract - *Data Mining is a methodology that has data sets as its focal point to extract information for predicting and uncovering the unseen patterns. Data Mining has a wide range of applications in industries like healthcare, marketing, retail, and agriculture. Agriculture is an important aspect of the economy of India. Along with food, it also provides employment opportunities for a large number of people. Agriculture depends on topography, climate, biology and soil. Agricultural Mining has the capability to bring knowledge to agriculture development. In agriculture, data mining plays the vital role in forecasting the yield, fruit grading, weather prediction, fertilizer usage, soil fertility and weed detection. Current studies present the different data mining techniques and their role in the much important factors of soil fertility and nutrient analysis. Decision tree is a technique for classification in data mining. C4.5, Classification and Regression Trees are most used algorithms for classification in decision tree. The major negative of the C4.5 algorithm is that it is inclined towards attributes with more values whereas the ID3 algorithm produces inadmissible errors if domain of the target attribute is huge. Considering these negatives, this paper presents an altered decision tree algorithm. The ID3, proposed classifier and the C4.5 trained on data sets containing soil samples by considering optimal soil parameters like the EC (Electrical Conductivity), ESP (Exchangeable Sodium Percentage) and pH (power of Hydrogen). The model is tested with the test data of soil samples. The test proves that the altered decision tree algorithm has greater and better accuracy when compared to ID3 and C4.5 algorithms. Classification of soil means separation of soil into groups or classes each having similar behavior and potentially similar characteristics. Classification of soil can help the farmer to know the types of crops and fertilizers best for that soil and hence to provide the best yield.*

***Key Words*: Data Mining, decision tree algorithms, soil fertility, nutrient analysis, crop prediction, machine learning**

1. INTRODUCTION

On a global scale, food needs are increasing day by day. As a result, agronomists, farmers, government and researchers are trying out and using various techniques in agriculture to improve production. As a result, the information generated from agricultural data is increasing constantly. The increased volume of information needs an unconstrained method for this information to be extracted and analysed when necessary. Data mining techniques would be used to predict long-term trends in agricultural processes.

Data mining techniques are basically of two types, first one is descriptive which takes into account existing data and the other is predictive which depends on the probability of a future analysis.

The data mining process includes:-

- Collecting, cleaning and uploading data to the data warehouse system
- Storing data in a multidimensional format
- Providing access to information for analysts and decision-makers
- Analyzing data using different applications
- Presenting data using different models

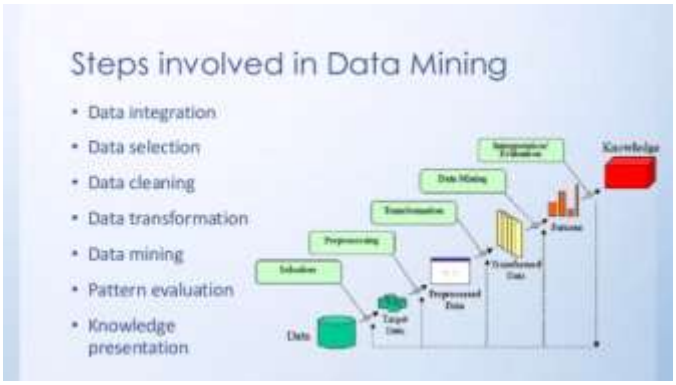

Figure 1: Data Mining Process

Agricultural analysts and farmers use sensors on their remote fields to obtain data from temperature, soil, moisture, etc. Automated collection of this data greatly facilitates their analytical work. Currently, they rely on network service providers, for example, the cell phone network to perform such automated collection from their remote locations, which is tough for most farmers due to operating costs. Examination and meta-analysis of yield data approximating organic and conventional farming displayed that current organic produce of each crops represent an average of 80% conventional yields. The analysis of 362 datasets too showed a large variation in the yield of organic farming (standard deviation of 21%). Part of this variation seems systematic. For example, Soybeans, some other Legumes, Rice and Corn score above 80% and Wheat, Barley and Potatoes score below 80%. Most regions have relative yields fairly close to the total average. In research by Leisa J. Armstrong, a comparative study of current data mining

techniques such as statistical methods and cluster analysis was carried out to establish the most effective technique. They used a large set of data extracted from the Western Australian Ministry of Agriculture and Agriculture (AGRIC) soil database to conduct this research. Experiments analyzed a small number of traits contained in the data set to determine their effectiveness compared to standard statistical techniques. The objectives of this study were to characterize and classify the soil to provide more detail on the soil quality information to search for approximate crops that could grow on the field. This study was to characterize and classify the soil to provide more details on the morphological information of the subsoil. An additional objective of this study was to search for an approximate value of soil loss from the Lower Moshi irrigation scheme using the universal soil loss equation.

Data mining is largely popular in the field of agriculture. Data mining involves the processing of data in agricultural soil, and the systematic analysis of huge sets of information datasets is an exciting space for analysis.

The production capacity of soil relies on its fertility. Data processing is used in vast areas and many ready-to-use data processing software and domain-centric data rendering and processing application software are available, but data processing in terms of agricultural soils can be a relatively young area of analysis. In the current scenario, there is no such recommendation system available online and the farmer must physically seek recommendations based on his report on the soil by quality frameworks. Each sector of this digital world is undergoing a drastic change due to the impact of the IT sector. But so far, little work has been done in the agricultural sector. The use of various data mining techniques in the agricultural sector will be an ongoing area of research. It will provide a basic Internet response for soil testing labs, free messages for farmers containing data such as soil test code, a chemical important for the crop. The results support the classification of the containers. The ultimate goal is to increase the outcome of the agricultural sector and to provide the best of produce to the farmer.

2. LITERATURE SURVEY

2.1. Automated Crop Prediction based on Efficient Soil Nutrient Estimation using Sensor Network

In this paper, the estimation of soil nutrients has been done. For the estimation of nutrients, NPK sensors have been used. These sensors detect the nutrient value and predict the suitable crops that can be grown in the soil. Farmer first need to list the main server with their NPK sensor. This sensor gathers nutrient values from the soil and updates this data to the main server with the help of raspberry pi. Depending on the values gathered from sensor, the algorithm makes predictions based on previous data. The proposed method estimates soil nutrients based on sensor network that helps in forecasting suitable crop for the soil under test. The requirements of this model are Admin, user, GSM, NPK sensor and raspberry pi. NPK sensors can detect the value anytime the farmer wants and predict the best crops suitable for that soil. Hence it ultimately saves time of farmers and also improves the quality of agriculture.

Advantages:-

- Varios samples of soil collected from different locations
- can be analysed.
- The device is mobile and handy.
- Lesser time consumption.

Disadvantages:-

- NPK sensor detects only potassium, nitrogen and phosphate in the soil.
- Parameters such as pressure, humidity and pH (Potential of Hydrogen) values cannot be recognised.

2.2. Crop Predition using Predictive Analytics

In this paper, Predictive Analysis has been used in order to measure the soil nutrients and predict suitable crops. Predictive analytics includes data mining and predictive modelling which can analyse the current and past data to make the predictions that help plan future ideas. Predictive analysis plays a vital role in business management. In business management, these analytics gives the most effective amalgamation of product versions, marketing material, and communication. Hence predictive analytics can be implemented in the case of agriculture as well which gives the most effective combination of soil and fertilizers.

Advantages: -

- Predicts soil and fertilizer pair making farmer's jobs easier.
- Works efficiently with different types of soil.
- Since it predicts future results, chances of failures are minimized.

Disadvantages:-

- User needs to have complete and proper knowledge about how to use predictive analytics.
- The model is difficult to implement

2.3. Crop Prediction using Predictive Analytics

In this paper, only evaluation of the data is done. It provides the software part of the system for the data evaluation and this will be used for further analysis and crop prediction. By this method production and also be predicted.

The technologies used in this are neural network and principal component analysis. The NPK values are used from the data set for the evaluation purpose. The principal component analysis is a procedure that would lead to the conversion of observation of correlated variables into orthogonal variables and these are called principal components. This is obtained by the orthogonal transformation. The first component has the highest variance and the succeeding components will have the highest variance with the limitation of it to be uncorrelated with previous components.

Advantages:

- The similarity of the soil nutrients and soil fertility level is shown more efficiently.
- Mathematical principles are used so the output is more efficient.

2.4. Two-Phase Fuzzy based Prediction model to predict Soil Nutrient

Soil nutrients help us to find out the crop productivity. The 'pH' value is used to forecast the concentration of nutrients in soil. In this article, an author proposes a two-phase methodology to forecast soil nutrients. Phase I predicts the pH value of soil depending on the preceding time-series pH value of the soil with the help of a fuzzy model. Phase II employs a regression model to forecast soil nutrients such as Phosphorus (P), Nitrogen (N) and Potassium (K) based on forecasted pH value. Regression and the fuzzy-based model is used to predict crop production. This model can be made more efficient

3. SYSTEM ANALYSIS

3.1. Problem Definition

Farmer should have data regarding soil fertility for higher crop yield. Farmers should remember macronutrients and small nutrients gifts in the soil to achieve the most yield of the specific crop and to grasp that plant food to use. Thus Soil Testing is a major think about Farming. In the current state of affairs, soil testing in Central government Labs takes two to three months to check specific soil samples. There are unit probabilities that Soil samples might get misplaced and farmers might get wrong reports. Thus we have a tendency to area unit proposing a model to automatize manual soil testing method.

3.2. Scope

Our system can analyze the soil parameters and nutrients gift in the soil like NPK which can facilitate to work out the fertility level of that soil. Our system won't solely give soil analysis however additionally predict the crops. The system also will recommend a list of fertilizers for that crop per NPK values. Farmers will check the soil multiple ranges of times throughout the cultivation method and take necessary precautions to induce smart yield. At the top reports are generated thus farmers will keep a record of their fertility.

4. PROPOSED SYSTEM

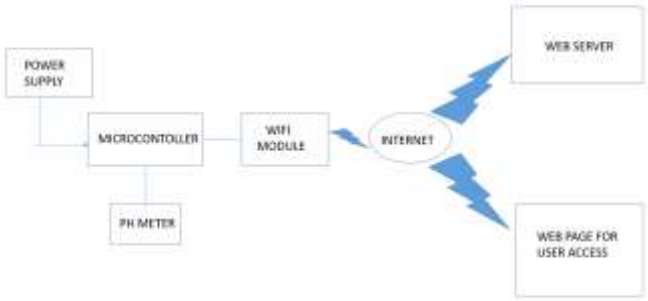

Figure 2: System Architecture

In this system hardware and software are integrated to form an IOT based system. The microcontroller ATMEGA328P is responsible for the conversion of the analog values from the PH meter to digital values for its further processing. The pH meter measures pH values of the soil from that Phosphorus (P), Nitrogen (N) and Potassium (K) values are estimated. These input values are then sent to a web server through an HTTP link, for this we have a Wi-Fi module ESP8266-01.

Here the data is stored. A web page is also made for the user to access it. Here we will use a Machine learning algorithm so that the user can know which crop can be grown in that area. The decision tree algorithm will be used for the prediction of crops.

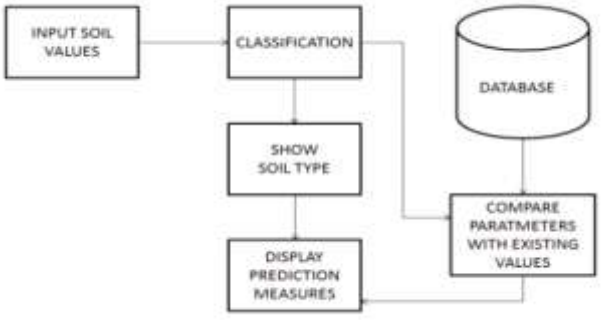

Figure 3: Conversion System

The objective of this system is to classify the soil according to the nutrients into it. For this data sets of soil samples are taken beforehand. The soil is classified using a decision tree algorithm and which type of soil is used will be displayed. Conjointly we have a tendency to predict the crops suitable for the actual type of soil. In addition to this, we are going to help the framer by improving their soil if the farmer

wants to yield a particular crop on that soil by suggesting them nutrients for the soil. To overcome these negatives we use a different and improved decision tree approach for soil classification. In this, first we evalulate the gini index for varied ranges of attributes instead of every consecutive pair. Then we used ratios of these calculated gini indices to lower the bias introduced by information gain in C4.5. The proposed algorithm is described below.

Input: Experimental available data set D

Output: A decision tree T is generated by giving experimental datasets.

- Create a node root which consist of the complete data set
- Calculate gini index for varied ranges of values of EC, ESP and pH (Potential of Hydrogen).
- Calculate the ratio of the gini indices for every single attribute.
- Choose an attribute with the smallest gini index as the node to split.
- Split the examples of the current node into various subsets based on values of the selected attributes.
- Create a new node as a child of the current node for every subset and pass the examples in the subset to the node
- Repeat steps 2 to 5 until only one instance remains in the node.

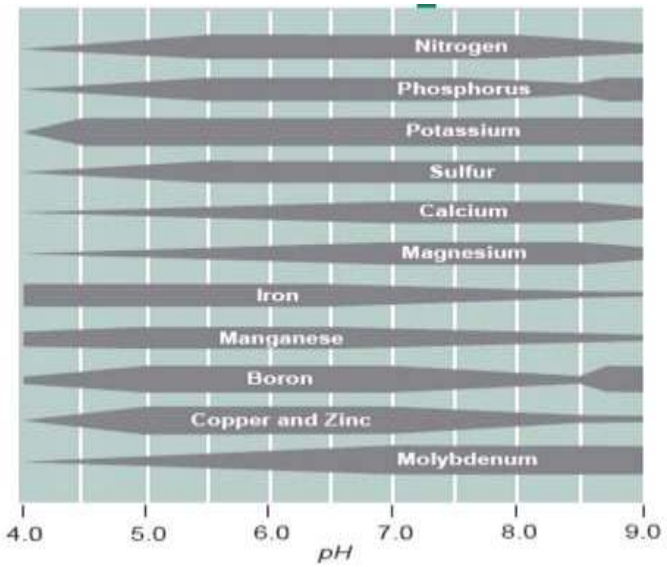

Figure 4: Nutrients present at particular pH value

5. WORKING

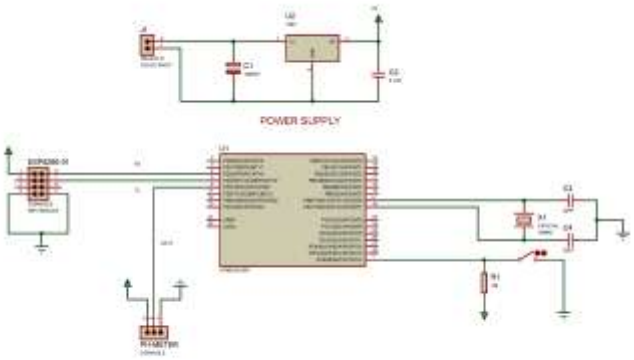

Figure 5: Internal Connections

Our system is a handheld device in which we are using:-

- 12V/1 amp adapter as an input power supply
- The board have 12V to 5V converter Regulator with a filtering capacitor
- We are using ATMEGA328P micro-controller which uses 8051 board architecture with 14 digital pins and 6 analog pins
- pH meter whose one wire is connected to ground and another to pin number 6 of microcontroller ATMEGA328P. It gives three value Soil pH value, Soil Moisture, and Surrounding Light Intensity
- It has the option to select any one at a time, and have three-pin Vcc/Gnd/Data
- pH meter measure pH values of the soil which is in analog format, so to send to the server we need to convert, from analog to digital values
- One of this analog Pin is used to capture the analog data of pH Meter and convert to the digital value
- pH values of the soil are used to estimate Nitrogen (N), Phosphorus (P) and Potassium (K) values
- After this, we have to create HTTP link with NPK data and send it to server
- For server communication, we are using Wi-Fi module ESP8266-01.
- In the Wi-Fi module (ESP8266) the receiver and transmitter are connected to pin numbers 4 and 5 respectively of microcontroller ATMEGA328P and the other wire is grounded.

- We have programmed in such a way that Wi-Fi automatically is connected to our mobile phone's hotspot.
- With the help of this internet connectivity, we sent the recently updated data to the server every 30 seconds.
- At server side we will use Machine learning algorithm so that user can know which crop can be grown in that area and depending on these values we will decide fertility level of soil and divide them into LOW, MEDIUM, HIGH class.
- The decision tree algorithm will be used for the prediction of crops.

Nutrients	Low	Medium	High
pH	1% Acidic	99% Neutral	0% Alkaline
EC- Electrical Conductivity	90% Normal	6% Slightly Saline	4% Saline
OC- Organic Corbin	100%	0%	0%
N- Nitrogen	100%	0%	0%
P- Phosphorous	0%	0%	100%
K- Potassium	0%	3%	97%
S- Sulphur	1%	5%	94%
Zn- Zinc	1% Deficient	99% Sufficient	
Fe- Iron	0% Deficient	100% Sufficient	
Cu- Copper	0% Deficient	100% Sufficient	
Mn- Manganese	100% Deficient	0% Sufficient	
B-Boron	0% Deficient	100% Sufficient	

Table 1: Soil Nutrient Status

6. CONCLUSION

Agriculture's the most important sector particularly in a country such as India. Around 65 – 75% of the population (directly or indirectly) depends on the agriculture sector. Hence it plays a very important role in the Economic growth of India. Therefore, Use of IT in agriculture can transform the picture of decision making and farmers can produce in an efficient way. In deciding the overall problems associated with the agriculture field, data mining has an important role. The part of data mining in terms of agriculture has been discussed in the survey. Also, we have discussed about different types of soils, fertilizers to improve the fertility of soil and various data mining techniques in soil containment and agriculture. Hence the project has been done in such a way that it succeeds all the requirements for improved quality of agriculture.

ACKNOWLEDGEMENT

We would like to take this opportunity to express our sincere thanks towards our guide Prof. Jyoti Dange for helping with the technical guidelines and propositions regarding the line of this work. We would like to express our gratitude towards her constant encouragement throughout the development of the project.

REFERENCES

[1] Ahmed, A.U., R.V. Hill, L.C. Smith, and D.M. Wiesmann. 2007. The World's Most Deprived: Characteristics and Causes of maximum financial condition and Hunger.

[2] Allen, J.C., and N. Mar. 1985. Soil Response to Forest Clearing within the u. s. and also the Tropics : geologic and Biological Factors Soil Response to Forest Clearing within the u. s. and also the Tropics : geologic and Biological Factors1. 17(1): 15–27.

[3] Andersson, J. a., and S. D'Souza. 2014. From adoption claims to understanding farmers and contexts: A literature review of Conservation Agriculture (CA) adoption among granger farmers in southern continent. Agric. Ecosyst. Environ. 187:116–132

[4] Anthofer, J. 2004. The Potential of the SRI for financial condition Reduction in Kampuchea.Consultancy Report. FSNPSP

[5] Asfaw, S., B. Shiferaw, F. Simtowe, and L. Lipper. 2012. Impact of contemporary agricultural technologies on granger welfare: proof from Tanzania and African nation. Food Policy 37(3): 283–295

[6] Baker, C.J., K.E. Saxton, W.R. Ritchie, W.C.T. Chamen, D.C. Reicosky, and F.R.S.E. Justice. 2006. No-tillage seeding in conservation agriculture (CJ Baker, KE Saxton, WR Ritchie, WCT Chamen, DC Reicosky, MFS Ribeiro, SE Justice And PR cricketer,Eds.).

[7] Barrett, C.B., C.M. Moser, and O. V Mchugh. 2004. higher Technology , higher Plots or higher Farmers ? distinguishing Changes In Productivity And Risk Among Malagasy Rice Farmers higher Technology , higher Plots or higher Farmers ? distinguishing Changes In Productivity And Risk Among Malagasy Rice Farmers. (687): 0–37.

[8] Bindraban, P.S., M. van der Velde, L. Ye, M. van den Berg, S. Materechera, D.I. Kiba, L. Tamene, K.V. Ragnarsdottir, R. Jongschaap, M. Hoogmoed, W. Hoogmoed, C. van Beek, and G. van Lynden. 2012. Assessing the impact of soil degradation on food production. Curr. Opin. Environ. Sustain. 4(5): 478– 488Available at http://linkinghub.elsevier.com/retrieve/pii/S1877343512001224 (verified ten July 2014).

[9] Chaain, a. K., and a. Y. Hoekstra. 2010. The green, blue and gray water footprint of rice from each a production and consumption perspective. Ecological Economics(40): 1–62 page