Automated Generation and Analysis of Soil Health Card and Calculation of the Village Soil Fertility Index

Aarti Kalekar[1], Prof. Sushma Vispute[2], Priti Kokane[3], Manali Kamble[4], Kajal Bokefode[5]
Department of Computer Engineering,
Pimpri Chinchwad College of Engineering,
Pune, Maharashtra, India

[1]aartikalekar01@gmail.com
[2]viputesushma@gmail.com
[3]priti.kokane2013@gmail.com
[4]manali96kamble@gmail.com
[5]kajalbokefode@gmail.com

Abstract: 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. In the state of Maharashtra, government officials known as 'Krushisahayaks' (Agro-assistants) collect soil samples from local farmers and submit them to a soil test laboratory, where they are evaluated by experts, after which the results are made available to every farmer. It is the job of the Agro-assistant to monitor the fertility of the soil in their locality and provide appropriate fertilizers to the farmers. This paper deals with the design and implementation of a system that automatically generates recommendations given the input of the soil test report. In addition, the system also helps the Agro-assistant to visualize the fertility of the soil in their village by mapping the information onto a Google map.

Keywords: agriculture, decision tree, clustering, recommendations

I. INTRODUCTION

Soil is an important component of agriculture. Nutrients in the soil move to the plants grown in it. Adding fertilizers is an important mechanism to keep agriculture production systems sustainable. In order to maintain the nutrient balance of the soil, it is important to apply fertilizers.

The Government of India has done a lot towards improving the fertility of soil in villages for agriculture. But there are some shortcomings. In the existing system, government officials known as Krushisahayaks (Agro-assistants) are assigned to each taluka (sub-district). It is the responsibility of these officials to collect the soil samples from the farmers in 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 manually generated by experts in the Department of Agriculture. The problem with this system is that the soil health cards are generated manually and on paper. These cards are distributed to the farmers and no computerized record is maintained for the same by the Agro-assistants. So, there is no system in place to monitor the condition of the soil, or to summarize the information succinctly so that the most prominent problems in a particular region can be determined.

Our system takes the soil test report as input and automatically creates a soil health card that provides fertilizer recommendations to farmers. Additionally, we use clustering techniques to compute a village-wise soil fertility index that can be used locally by decision makers to monitor the deficiencies in the soil of the village. We map the data obtained through clustering to a Google map to better visualize the information.

II. LITERATURE SURVEY

In a research carried out by Gholap, Ingole, et al[1], an automated system for soil classification based on its fertility was proposed. Under this, various classification algorithms like NBTree, SimpleCart, J48 have been studied, with the conclusion that the J48 decision tree algorithm works best with the soil dataset, showing an accuracy of 91.90%.

Hot and Popović-Bugarin analyze the problem of soil clustering based on the chemical characteristics of soil, and proper visual representation of the obtained results[2]. The paper focuses on clustering soil data using the K-means algorithm. Results obtained by using KM are presented on the Static Google map and dynamic Open Street Map of Montenegro.

A. Soil Attributes

Nitrogen(N): Plants absorb nitrogen in the form of ammonium or nitrate ions. Indian soils are almost universally deficient in N. It should be present in the right proportion in the soil for the growth of the plants. The optimum N concentration is 2-10 ppm.

Phosphorous(P): Phosphorous has been called the Master key to agriculture. It is essential in plant growth, fruit growth, cell division and early ripening. The optimum P concentration is 30-50 ppm.

Potassium(K): Potassium is an essential macro-nutrient and is associated with water movement, nutrients and carbohydrates present in the plant tissue. The optimum K concentration is 20K ppm.

Soil pH: Soil pH is an indicator of the acidity and alkalinity in soils. pH values range from 0 to 14, with 7 being neutral, below 7 acidic and above 7 alkaline. Excessively acidic or excessively basic soils can be toxic to the crop. A pH in the range of 6.5 to 7.5 is generally considered to be optimal, as most nutrients are available in this range.

B. Crop Seasons

Kharif: This season is from July to October during the south-west monsoon. The crops grown during this season include rice, maize, sorghum, bajra, soyabean, etc.

Rabi: This season is from October to March(winter). The crops grown during rabi are wheat, barley, oats, linseed, mustard, etc.

C. Algorithms

J48 Decision Tree

Decision tree J48 is the implementation of algorithm ID3(Iterative Dichotomiser 3) developed by the WEKA project team. R includes this work in the package RWEKA. J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool. C4.5 is a program that creates a decision tree based on a set of labeled input data. This algorithm was developed by Ross Quinlan. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

ALGORITHM 1: J48 Decision Tree

1. Check if algorithm satisfies termination criteria.
2. Compute information-theoretic criteria for all attributes.
3. Choose best attribute according to the information-theoretic criteria.
4. Create a decision node based on the best attribute in step 3.
5. Induce (i.e. split) the dataset based on newly created decision node in step 4.
6. For all sub-dataset in step 5, call J48 algorithm to get a sub-tree (recursive call).
7. Attach the tree obtained in step 6 to the decision node in step 4.
8. Return tree

NBTree

The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

ALGORITHM 2: NBTree

1. Calculate prior probabilities for each class in the data set.
2. Calculate the conditional probabilities for every attribute in the data set.
3. Classify each example in the data set with maximum posterior probability.
4. For misclassified examples, evaluate the utility of a split on each attribute.
5. Let j be the attribute with the highest utility.
6. If utility of j is not significantly better than that of the current node, create a Naive Bayes classifier for the current node and return.
7. Partition the training data set.
8. Call the algorithm recursively for each child on the portion of the data set that matches the test leading to the child.

SimpleCart

It is a non-parametric decision tree learning technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric, respectively. It is used for implementing minimal cost complexity pruning.

ALGORITHM 3: SimpleCart

1. Define the important variables.
2. Select a splitting criterion.
3. Create a tree in which a single node contains all data(Initialization).
4. Split the one terminal node resulting in the greatest increase in the splitting criterion(Splitting).
5. If each leaf contains samples belonging to the same class, or if some preset threshold is not satisfied, stop. Otherwise continue(Stopping).
6. Use a test set to prune the tree(Pruning).

From the algorithmic survey, it is found that J48 Decision Tree algorithm works the best for the soil data set.

III. METHODOLOGY

A. Dataset

The dataset required for the project was collected:

- Agriculture Officer - Mr. Vijay Kolekar
- Agro-assistant (Khed sub-district)
- College of Agriculture, Pune
- www.soilhealth.dac.gov.in

The dataset contains nine soil attributes including macro-nutrients (N, P, K) and micro-nutrients (OC, EC, pH, B, Mn Mg).

We define five class labels (VL, L, M, H, VH for Very Low, Low, Moderate, High, and Very High respectively) that describe the soil's fertility.

B. Design

The system is targeted towards Agro-assistants. An agro-assistant can register and login to view data specific to their area. Pertinent information needs to be entered to generate soil status generate recommendations with respect to fertilizers that can be used and their respective proportions.

Furthermore, a map of the soil fertility across the village allows the agro-assistant to better understand the persistent problems with the soil in the region under their control and enable an optimized use of fertilizer.

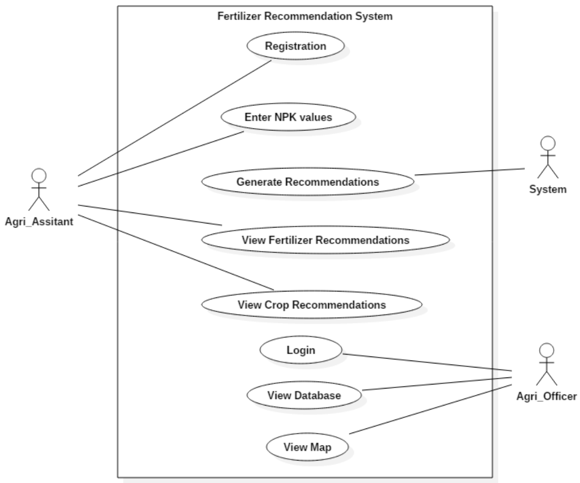

Figure 1: Use Case Diagram

III. RESULTS

Our findings indicate that the J48 decision tree algorithm is 87.5% accurate in classifying the soil according to its fertility. The clusters found out through K-means algorithm are mapped onto a Google map.

The SI unit for magnetic field strength H is A/m. However, if you wish to use units of T, either refer to magnetic flux density B or magnetic field strength symbolized as $\mu_0 H$. Use the center dot to separate compound units, e.g., "$A \cdot m^2$."

Table 1: Performance Evaluation

label(NPK)	Recall	Precision	F-measure	Accuracy
VH-M-H	1	1	1	1
VH-L-H	1	1	1	1
VH-L-VH	1	0.6	0.75	0.95
M-L-H	1	1	1	1
H-VL-H	1	1	1	1
M-L-VH	1	1	1	1
H-L-H	1	1	1	1
H-L-VH	1	0.5	0.67	0.95
VH-L-M	1	1	1	1
H-M-VH	0	1	1	0.75
Overall	0.9	0.91	0.94	0.88

Table 1: Confusion Matrix

label(NPK)	TP	FP	TN	FN
VH-M-H	1	0	39	0
VH-L-H	11	0	29	0
VH-L-VH	3	2	35	0
M-L-H	1	0	39	0
H-VL-H	2	0	38	2
M-L-VH	9	0	31	0
H-L-H	4	0	36	0
H-L-VH	2	2	36	0
VH-L-M	2	0	38	0
H-M-VH	0	0	38	0

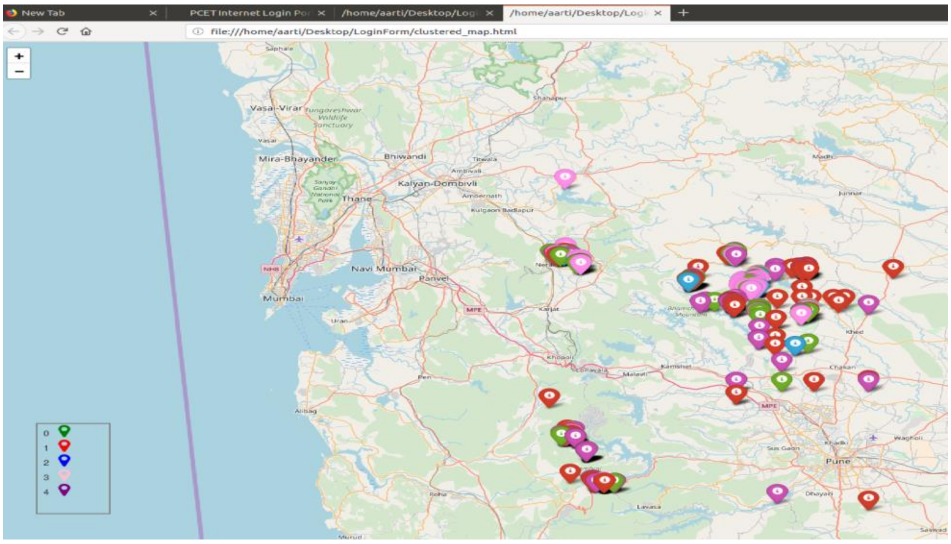

Figure 2: Mapping Soil Fertility onto Google Maps

IV. CONCLUSION

An increase in agricultural yield can only be attained by ensuring that the soil provides a balanced and an adequate supply of nutrients. This system will be useful to decision makers to understand problems in soil and resolve them locally. The J48 decision tree algorithm shows an accuracy of 87.5% in classifying the soil according to its fertility, and can be used to recommend appropriate fertilizers. Furthermore, the visualization helps provide better understanding of the problem at hand.

V. ACKNOWLEDGEMENT

We express our sincere thanks to our Guide Prof. S. R. Vispute for her constant encouragement and support throughout our project, especially for the useful suggestions given during the course of project and having laid down the foundation for the success of this work. We would also like to thank Mr. Vijay Kolekar, Agronomist, Department of Agriculture, Govt. of Maharashtra for providing us invaluable guidance.

REFERENCES

[1] 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

[2] Elma Hot, Vesna Popović-Bugarin, “Soil Data Clustering by Using K-means and fuzzy K-means Algorithm”, Telfor Journal, Vol. 8, No. 1, 2016

[3] K.K.Deshmukh, “Evaluation Of Soil Fertility Status From Sangamner Area, Ahmednagar District, Maharashtra, India”, Rasayan Journal of Chemistry, 2012

[4] 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

[5] Department of Agriculture & Cooperation Ministry of Agriculture Government of India, “Methods Manual-Soil Testing in India”, 2011

[6] Bhargavi P. & Jyothi S., “Soil Classification Using Data Mining Techniques: A Comparative Study”, International Journal of Engineering Trends and Technology, 2011

[7] Dr. S.Hari Ganesh and Mrs. Jayasudha, "Data Mining Technique to Predict the Accuracy of the SoilFertility", IJCSMC,Vol. 4, Issue. 7, July 2015, pg.330 –333.

[8] 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

[9] Rachana P. Koli, V. D. Jadhav, "Agriculture Decision Support System As Android Application", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

[10] Govt. of India [Online]. Available: http://soilhealth.dac.gov.in/

First Author

1. Name : Aarti Kalekar
2. Date of Birth : 22/01/1996
3. Gender : Female
4. Address : Chinchwad, Pune-411033
5. E-mail : aartikalekar01@gmail.com
6. Contact : 8600386941
7. Placement Detail : Placed in Mindtree
8. Paper Published : Automated System for Soil Health Card Generation and Analysis to Satisfy Local Agriculture Needs : An Overview

Second Author

1. Name : Priti Kokane
2. Date of Birth : 06/09/1996
3. Gender : Female
4. Address : Sajjangad colony, Rahatani, Pune-17
5. E-mail : priti.kokane2013@gmail.com
6. Contact :7387124140
7. Placement Detail : Placed in Excellon Software
8. Paper Published : Automated System for Soil Health Card Generation and Analysis to Satisfy Local Agriculture Needs : An Overview

Third Author

1. Name : Manali Kamble
2. Date of Birth : 24/10/1996
3. Gender : Female
4. Address : Thergaon, Kalewadi, Pune-411033
5. E-mail : manali96kamble@gmail.com
6. Contact : 7030703138
7. Placement Detail : Placed in KPIT
8. Paper Published : Automated System for Soil Health Card Generation and Analysis to Satisfy Local Agriculture Needs : An Overview

Fourth Author

1. Name : Kajal Bokefode
2. Date of Birth : 02/02/1997
3. Gender : Female
4. Address : Thergaon, Pune-411033
5. E-mail : kajalbokefode@gmail.com
6. Contact : 7066757281
7. Placement Detail : Not Placed
8. Paper Published : Automated System for Soil Health Card Generation and Analysis to Satisfy Local Agriculture Needs : An Overview