Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 7, Issue. 3, March 2018, pg.20 – 24

A Novel Approach for Classification of Soil and Crop Prediction

Vrushal Milan Dolas[1], Prof. Uday Joshi[2]
[1] Department of Computer Engineering, K. J. Somaiya College of Engineering, Mumbai, India
[2] Department of Computer Engineering, K. J. Somaiya College of Engineering, Mumbai, India
[1] vrushal.d@somaiya.edu
[2] udayjoshi@somaiya.edu

Abstract— Decision tree is a well-known approach for classification in data mining. C4.5 and Classification and Regression Trees (CART) are two widely used decision tree algorithms for classification. The main drawback of C4.5 algorithm is that, it is biased towards attributes with more values while CART algorithm produces misclassification errors when the domain of the target attribute is very large. In view of these limitations, this paper presents a modified decision tree algorithm. The C4.5, CART and the proposed classifier are trained using data set containing soil samples by considering optimal soil parameters namely pH (power of Hydrogen), Ec (Electrical Conductivity) and ESP (Exchangeable Sodium Percentage). The model is tested with test data set of soil samples. The test proves that the modified decision tree algorithm has higher classification accuracy when compared to C4.5 and CART algorithms. Classification of soil is the separation of soil into classes or groups each having similar characteristics and potentially similar behavior. Classification of soil is needed so that farmer can know the type of soil and can plough the crops depending on the type of soil.

Keywords— Classification and Regression Trees (CART), Exchangeable Sodium Percentage, C4.5, Electrical Conductivity, Decision tree.

I. INTRODUCTION

Soils are complex mixtures of minerals, water, air, organic matter, and countless organisms that are the decaying remains of once-living things. It forms at the surface of land – it is the "skin of the earth." Soil is capable of supporting plant life and is vital to life on earth. Classification of soil is the separation of soil into classes or groups each having similar characteristics and potentially similar behavior. A classification for

engineering purposes should be based mainly on mechanical properties: permeability, stiffness, strength. The class to which a soil belongs can be used in its description. Classification of soil is needed so that farmer can know the type of soil and can plough the crops depending on the type of soil. Also if farmer want to yield particular crop then the improvement of soil can be done by adding the required nutrients in the soil.

Indian standard soil classification system

- Classification Based on Grain Size, The range of particle sizes encountered in soils is very large: from boulders with dimension of over 300 mm down to clay particles that are less than 0.002 mm. Some clay contains particles less than 0.001 mm in size which behave as colloids, i.e. do not settle in water.
- In the Indian Standard Soil Classification System (ISSCS), soils are classified into groups according to size, and the groups are further divided into coarse, medium and fine sub-groups.
- The grain-size range is used as the basis for grouping soil particles into boulder, cobble, gravel, sand, silt or clay.

Very coarse soils	Boulder size		> 300 mm
	Cobble size		80 - 300 mm
Coarse soils	Gravel size (G)	Coarse	20 - 80 mm
		Fine	4.75 - 20 mm
	Sand size (S)	Coarse	2 - 4.75 mm
		Medium	0.425 - 2 mm
		Fine	0.075 - 0.425 mm
Fine soils	Silt size (M)		0.002 - 0.075 mm
	Clay size (C)		< 0.002 mm

Table 1: Sizes of grain in different soils

- Gravel, sand, silt, and clay are represented by group symbols G, S, M, and C respectively.
- Physical weathering produces very coarse and coarse soils. Chemical weathering produces generally fine soils.

A. **Motivation**

The purpose of this work is to categorize the soil samples according to the nutrients present into it and to predict the crops which can be yield in the particular soil. Also this project provides suggestion to improve the soil if the farmer wants to yield specific crop in that soil.

B. **Scope**

- Classification of soil using CART algorithm.
- Prediction of crops for suitable land.
- Suggestions for improving the quality of land using various nutrients.
- Amount of water to be supplied to crop with respect to climatic conditions.
- Suggests suitable crop for a particular type of soil.

II. LITERATURE SURVEY

The literature for the classification of the soil and prediction of crops are as follows-

In 2014, K AdityaShastry, Sanjay H A and Kavya H discussed various classification techniques in the paper "A Novel Data Mining Approach for Soil Classification". The algorithms CART, C4.5 and proposed approach has been studied. Accuracy obtained for C4.5 is more as compared to CART. The proposed system uses gini index and gini ratio. The impurity measure gini index used in CART is biased towards attributes with higher range of values, while information gain used in C4.5 is biased towards attributes with high values. This drawback is removed using the proposed approach. From the paper it is also observed that the manual soil classification being done is very cumbersome and time consuming.

In 2015, Monali Paul, Santosh K. Vishwakarma and Ashok Verma in "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" introduced that Data mining in agriculture field is somewhat a novel research field. Data mining is the process of discovering unknown and likely impressive patterns in large datasets. Steps of data mining such as selection, preprocessing, transformation, data mining and interpretation has been discussed and Naïve bayes and K-nearest neighbor has been used for prediction and analysis. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting and works much better with complex real situations.K-Nearest Neighbor makes predictions based on the outcome of the K neighbors closest to that point. K- nearestNeighbour uses Euclidean distance formula for calculation.

In "Sensible approach for Soil Fertility Management using GIS Cloud" Decision Support System (DSS) using GIS enabled cloud technologies is implemented. The proposed approach of agricultural information development and integration system ensures that complete agricultural related data on cloud database can be integrated into spatial maps through GIS technologies, to organize, accumulate, and administer geospatial data in a cloud database according to individual farmer land information for improving data accuracy of digital agriculture fertilizer management. IaaS, PaaS and Saashas been discussed with the respective layers used. GIS Cloud server is used to incorporate, integrate, store, update and manage complete information of agriculture.

In 2015, Amol D. Vibhute, K. V. Kale, Rajesh K. Dhumal and S. C. Mehrotra in "Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data" has demonstrated use of support vector machine algorithm for identification, mapping and classification of various types of soil using high spectral resolution Hyperspectral data. Gaussian Radial Basis Function (RBF) kernel of SVM was used and the accuracy obtained is 71.18.

In 2010, Sofianita Mutalib, S-N-FadhlunJamian, Shuzlina Abdul-Rahman and Azlinah Mohamed explained In "Soil Classification: An Application of Self Organizing Map and k-means" The two unsupervised technique Kohonen Self Organizing Map (SOM) and k-means have been used to classify the soil. Characteristics of soil such as parent material, soil horizontal profile, color of soil, texture of soil and soil depth is listed. This paper predicts the classification of soil and gives information about the plants to be cultivated in specify type of soil.

III. PROPOSED SYSTEM

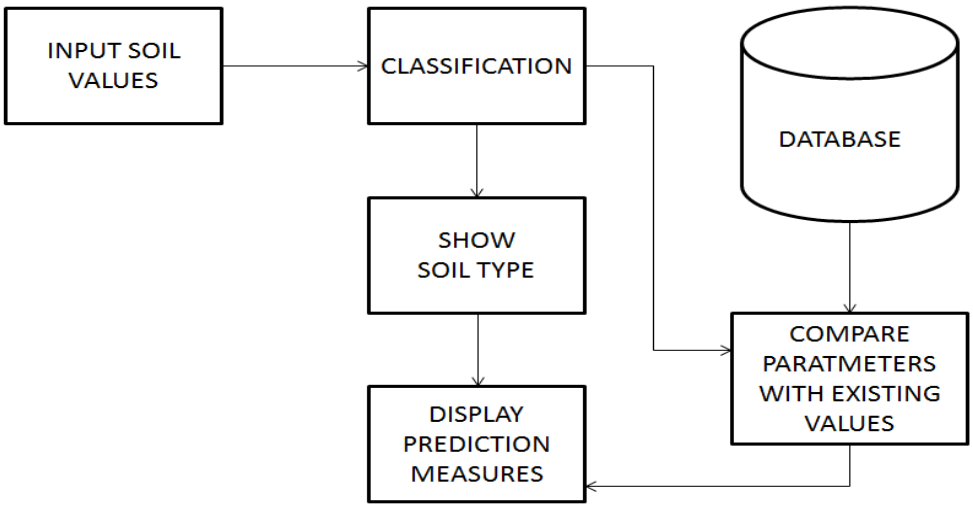

Figure 1: Proposed System of the project

The primary objective of this system is to classify the soil according to the nutrients into it. For this, we have taken datasets of soil samples. The soil will be classified using decision tree algorithm and type of soil will be displayed. Also we are going to predict the crops suitable for the particular type of soil. In addition to this, we are going to improve the soil if the farmer wants to yield particular crop in the same soil by suggesting the requirements of the nutrients for the same soil.

To overcome these drawbacks we propose a modified decision tree approach for soil classification. Here, we first calculate the gini index for different ranges of attribute values instead of computing for every successive pairs as was done in CART algorithm. Then we have used ratios of these computed gini indices to reduce the bias that was introduced by information gain in C4.5 algorithm. The proposed algorithm is described below.

Input: Experimental data set D

Output: A decision tree T which is created by giving experimental dataset.

1) Construct a root node which contains the whole data set

2) Compute Gini index for different ranges of attribute values of pH, Ec and ESP

3) Compute the ratio of the gini indices for each attribute

4) Select an attribute with the least gini ratio as the node to be split.

5) Split the examples of the current node into different subsets based on values of the selected attributes

6) Create a new node as a child of the current node for each subset and pass the examples in the subset to the node

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

IV. CONCLUSION

Classifying the soil using CART algorithm helps to know the overall health of the soil and the content of the nutrients in the soil. Also prediction helps to decide what all he can yield in particular soil. This can help the farmer to yield the variety of crops seasonally. In case we wish to yield particular crop, then we can improve the soil by adding the necessary nutrients in the soil as per required by that crop.

V. FUTURE WORK

In the current work we are using datasets, further works include:

1. Obtaining the requirement of water by the soil and scheduling the watering to the crops.
2. The use live data from Agriculture University.
3. Making the website or application for multiple languages.

ACKNOWLEDGEMENT

I would like to express my gratitude and appreciation to all those who helped me to complete this paper. A special thanks to my guide, **Prof. Uday Joshi** whose help, stimulating suggestions and encouragement helped me in writing this paper.

REFERENCES

[1] 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.

[2] 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.

[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] www.en.wikipedia.org

