Soil Testing and Crop-Suggestion Using Data Mining

Priya Thomas
Dept. of Computer Applications
Amal Jyothi College of Engineering, Kanjirappallly, India
priyathomas@mca.ajce.in

Ms. Sruthimol Kurian
Dept. of Computer Applications
Amal Jyothi College of Engineering, Kanjirappallly, India
sruthimolkurian@amaljyothi.ac.in

***Abstract*— Agriculture is the most essential component of Indian economic system. To increase yield production, many factors are responsible. Data mining plays an important role in agriculture sector. Data mining is the process of mining important information from a collection of huge data set and transform this into an understandable form for future use. There for the aim of this work is to predict soil type by using data mining classification techniques.**

In this study, soil dataset containing soil test outcomes has been used to apply various classification strategies in data mining. Soil type deals with the categorization of soil into specific soil classes as "very low", "low", "medium", "high", and "very high" on the idea of % of nutrient determined within the soil and on the basis of these lessons crops are recommended for a soil portion.

Keywords—Data mining, Soil data set, Classification Techniques, Agriculture.

I. INTRODUCTION

Agriculture became a critical thing in the development of sedentary human civilization, where farming of domesticated species created food surpluses that nurtured the improvement of civilization. The study of agriculture is called agricultural technology. Soil is defined as a thin layer of earth's crust which serves as a natural medium for the increase of vegetation. The classification of soil can be done according to various factors like taxonomy, pH, liming, nutrients, and organic matter.

Data mining is a big part of agriculture these days. It uses various techniques to classify soil, such as J48, naïve Bayes' theorem, etc. WEKA is helpful in the learning process of basic concepts of data mining where we can apply different options and analyze the result that is being produced.

II. SOIL TESTING

Soil testing is the chemical analysis of soil to determine it's fertility state and to predict nutrients required by crops. It also involves determining other soil attributes like texture and density.

III. ANALYSIS OF SOIL TESTING USING CLASSIFICATION TECHNIQUES

In this section, we present soil datasets that can be useful when classifying soils into different classes using data mining techniques based on the percentages of nutrients contained in the soil.

A. Soil Dataset

It is essential for agricultural planning and development to have an in-depth understanding of soil properties. Survey data are collected from field sampling, and physical, chemical, and biological properties of a soil sample are studied.

Table I. Attribute Description

Field	Description
N	Nitrogen, ppm
P	Phosphorous, ppm
K	Pottassium, ppm
S	Sulfur, ppm
Fe	Iron, ppm
Cu	Copper, ppm
Zn	Zinc, ppm
B	Boron, ppm
pH	pH value of soil
EC	Electrical Conductivity, mmhos/cm

B. Soil Classification

The classification of soil is based on the amount of nutrients it contains, with different classes such as "very low", "low", "medium", "high", and "very high". Data mining algorithms consisting of Naïve Bayes and C4.5 are used for the categorization of soil.

Following segment provide explanation of classification algorithms such as Naive Bayesian classifier and J48 selection tree classifier.

B.1). Naïve Bayes

Naive Bayes is a probabilistic machine learning model that is used for classifying data, which is based on Bayes' theorem. Using Bayes theorem, we will discover the probability of A happening, given that B has occurred. The assumption made right here is that the predictors/capabilities are independent. That is, presence of one particular feature does not have an effect on the other. hence it is referred to as naive.

DOI: 10.5281/zenodo.6180525
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

B.2). J48(C4.5)

The J48 is one of the classification and decision tree algorithm and it is slightly modified from C4.5 in Weka. It can select the test as best information gain. J48 predicts dependent variable from available data. It builds a tree based on attribute values of training data. It classifies data with the help of feature of data instances that is said to have information gain. The significance of error tolerance is evolved using pruning concept.

III. RESULTS

A. *Data set Description*

The soil dataset has 10 attributes and 10 instances of soil samples. It is prepared in Excel and is saved in .CSV file to allow them to be applied to WEKA.

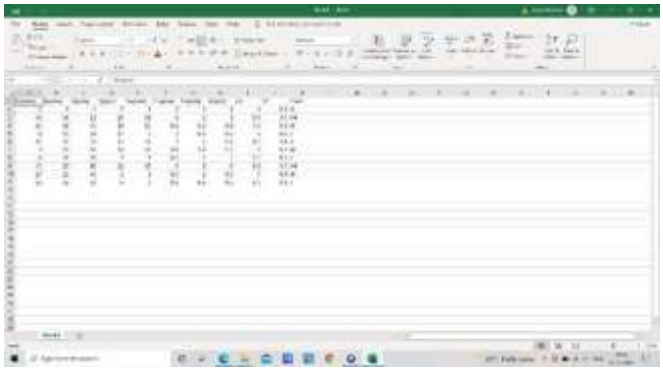

Fig. 1. Soil Dataset

The dataset is opened using WEKA.

Fig. 2. Open dataset using WEKA

B. *Soil Classification Using Various Classifiers*

Here we apply classification using the classify tab and classify the soil dataset using J48 classifier.

Fig. 3. J48 Tree

Soil dataset is classified by using J48 classifier into 5 fertility classes “very low”, “low”, ”medium”, ”high” ,”very high”. Out of 10 instances 6 instances classified successfully and 4 instances incorrectly in 0.02 sec.

Fig. 4. Classification by J48

Now classify the dataset using Naïve Bayes Classifier.

Fig. 5. Naïve Bayes Classification using WEKA

After applying Naive Bayes classification, soil attributes are classified into 5 fertility classes “very low”, “low”, “medium”, “high”, “very high” . This algorithm out of 10 instances classifies 3 instances correctly and 7 instances incorrectly in 0 sec.

Fig. 6. Classification by Naïve Bayes

DOI: 10.5281/zenodo.6180525
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

Table II. Comparison of Correctly and Incorrectly Classified Instances

Algorithm	Correctly Classified	Incorrectly Classified	Time(in seconds)
J48	60%	40%	0.02
Naïve Bayes	30%	70%	0

Here J48 gives better performance and high accuracy than Naïve Bayes.

Table III. Algorithm Error Rate

Algorithm	Mean Absolute Error	Root Mean Squared Error
J48	0.2267	0.414
Naïve Bayes	0.2881	0.5307

Table III shows the classifiers error rate for given sample dataset. Here, J48 has low error rate for Root Mean Squared Error. Naïve Bayes has low error rate for Mean Absolute Error.

IV. Conclusion

This segment encompass that after classification of soil into specific fertility classes expert decide what are the deficiencies in the soil, which kind of fertilizer need to be used for that sort of soil and which kind of crop is best in that specific sort of soil. This facilitates the most effective growth and obtaining the yield potential of the crop.

References

[1] Analysing Soil Data using Data Mining Classification Techniques, V. Rajeswari* and K. Arunesh.

[2] Agricultural Soil Lime Status Analysis Using Data Mining Classification Techniques, 1. Dr. K. Arunesh, 2. V. Rajeswari.

[3] Soil Classification and Fertilizer Recommendation Soil Classification and Fertilizer Recommendation using WEKA, 1. Suman , 2. Bharat Bhushan Naib.

DOI: 10.5281/zenodo.6180525
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam