Analysis of Soil Nutrients using Data Mining Techniques

Chiranjeevi M N¹, Ranjana B Nadagoudar²
¹PG Student, Dept. of CSE, VTU, Belagavi
²Asst. Professor, Dept. of CSE, VTU, Belagavi

Abstract: Soil nutrients analysis is useful for farmers to decide which type of crops to be cultivated in a particular soil condition. Analyzing soil nutrient will prove to be a major benefit to the farmers. Agricultural research has been profited by technical advances such as automation, data mining. Today, Data Mining Techniques are impressively appreciated in the analysis field of agriculture. The agriculture aspects are climate, rain, soil, pesticides and fertilizers stand the key responsible feature to increase the production of crops. A huge Dataset of Soil Nutrients grade Database was collected from the Department of agriculture, Belagavi. The database covers measure of soil nutrients for Belagavi District. The selected soil attributes pH, EC, OC, Nitrogen, Phosphorus, Potassium, Magnesium, Sulfur, Iron, Zinc, Copper and Boron for considering the soil increments using NB and J48 in WEKA. Soil Nutrients classification will help the farmer to cultivate suitable crops in a particular soil sample.

Keywords: Data Mining, Agriculture, Soil Nutrients, Classification, Naive Bayes, Decision Tree

I. INTRODUCTION

Data mining techniques are mightily, acclaimed in the research field of agriculture. It becomes popular in the field of agriculture for soil classification. The agriculture factors weather, rain, soil, pesticides and fertilizers are the main responsible aspect to raise the production of yields. The main aspect of agriculture is soil for crop growing.

Crop cultivation depends on the soil quality over a long distance, increasingly cultivation of land brings about loss of supplements present in the soil. In crop cultivation, great soil implies capacity of soil for high product yield. Good nature of soil guarantees us for maintenance and giving the normal outcome and opposes foulness. Soil is important part of natural environment. It is as important as plants, animals, rocks, landforms and living organisms. All these elements help in managing soil properly which helps in preserving its fertility to obtain a good yield.

Classification is one of the main problems in the field of data mining. Classification is a data mining technique based on machine learning which is used to categorize the data item in a data set into a set of predefined classes. Two classification algorithm J48 and Naive bayes are used. Naive bayes algorithm is based on probability and J48 algorithm is based on decision tree. We make comparative evaluation of Naive Bayes and J48 in the context of soil nutrients dataset. The results of comparison shown about classification accuracy and performance analysis. The result shows that performance analysis and accuracy of Naive Bayes is better than that of J48 and provides suggest to farmer which crop is suitable to cultivate in a particular soil sample.

II. RELATED WORK

In exiting system, classify the soil nutrients datasets based on classification algorithms using data mining.

In proposed system, provide suggest to farmers which crop is growing in particular soil sample.
III. METHODOLOGY

• Classification Algorithms
It is one of Data Mining. That is used to analyze the given dataset and takes each instance of it. It allocates this instance to a specific class. It is used to extract models. The important data classes within the given dataset. Classification is a two-step process.

During the first step, the prototypical is formed by applying classification algorithm. That is on training data set.

The second step, the removed prototypical is tested against the predefined test dataset. That is to quantity the prototypical trained performance and accuracy.

J48 Decision tree Classifier: J48 is improvement of C4.5 classifier. A choice tree is a flowchart like tree development, Where each internal hub demonstrates a test on a characteristic, each branch implies a consequence of the test and each leaf hub grasps a class tag. The most important hub in a tree is the root hub. All through tree structure, attribute increase techniques like information pick up, pick up proportion, Gini record are utilized to decision the characteristic that best dividers the tuples into disparate classes. Show settled on by choice tree predicts new occurrences of information.

Naive Bayes Classifier: - It is a humble probabilistic classifier based on bayesian theorem with tough naive individuality expectations, they can expect class association probabilities, such as probability that given tuple goes to a specific class. It usages prior probability of respectively class assumed no data almost an element. Classification harvests a posterior probability distribution complete the possible groups assumed a portrayal of a thing.

• Data Mining Process

1. Data Collection
Soil nutrient Dataset required for this work was collected from Department of Agriculture in Belagavi. These datasets contain various attributes and their respective values of soil samples taken from Belagavi District. Dataset has 12 attributes and a total 1676 instances of soil samples. Table 1 shows attribute description

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>pH value of Soil Data</td>
</tr>
<tr>
<td>EC</td>
<td>Electrical Conductivity</td>
</tr>
<tr>
<td>OC</td>
<td>Organic Carbon</td>
</tr>
<tr>
<td>N</td>
<td>Nitrogen</td>
</tr>
<tr>
<td>P</td>
<td>Phosphorus</td>
</tr>
<tr>
<td>K</td>
<td>Potassium</td>
</tr>
<tr>
<td>S</td>
<td>Sulfur</td>
</tr>
<tr>
<td>Zn</td>
<td>Zinc</td>
</tr>
<tr>
<td>Fe</td>
<td>Iron</td>
</tr>
<tr>
<td>Cu</td>
<td>Copper</td>
</tr>
<tr>
<td>Mn</td>
<td>Manganese</td>
</tr>
<tr>
<td>B</td>
<td>Boron</td>
</tr>
</tbody>
</table>

Table 1: Attribute Descriptions
2. Data Pre-Processing
In this step described about removing unwanted data from the dataset which helps to extract required result data from the dataset.

3. Data Conversion
If data mining uses WEKA tool, the data must be in ARFF format. All the data sheets converted into .arff file “Attribute Relation file format”. This ARFF file has sections those are Header and Data Information.

4. Classification
This technique of data mining is based on machine learning using concepts of algorithms. In this soil nutrient datasets are classified using Naive Bayes and J48 classification algorithms.

5. Prediction
The presentation of classification algorithm associated based on accuracy and performance analysis and provides suggest to farmer which crop is cultivate in a particular soil sample.

IV. RESULTS

![Fig. 1 Accuracy for Naive Bayes Classifier](image1.png)

![Fig. 2 Accuracy for J48 Classifier](image2.png)
Fig. 3 Execution Time of Classifiers

Fig. 4 Accuracy for Classifiers

Fig 5 Classification based on Naive Bayes Classifier
Fig 6 Classification based on J48 Classifier

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Match Instance</th>
<th>Unmatch Instance</th>
<th>Accuracy(%)</th>
<th>Execution Time in sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>1003</td>
<td>27</td>
<td>98%</td>
<td>14</td>
</tr>
<tr>
<td>J48</td>
<td>692</td>
<td>338</td>
<td>68.0%</td>
<td>34</td>
</tr>
</tbody>
</table>

Table 3: Comparative analysis of classifiers

V. CONCLUSION AND FUTURE WORK

In this proposed work, the comparative analysis of two algorithms like Naive Bayes and J48 is carried out. Among the available classification algorithm Naive Bayes gives better result for the dataset which we were used and is correctly classified into maximum number of instances comparing with remaining algorithms. Naive Bayes Algorithm can be recommended to predict crop growing in particular soil sample.

The future work may aim to create more efficient models using other data mining classification techniques such as support vector machine, principal component analysis, etc.

REFERENCES


II. Shailesh Gargade ,Gholap J, Anurag I, Jayesh G, vahida attar, “soil data analysis using classification techniques and soil attribute prediction”.


VIII. T. Mathavi Parvathi“Automated soil testing process using combined mining patterns”Research Scholar, Manonmaniam Sundaranar University.