iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4309

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Soil Quality Analysis using Data Mining (Comparison of Two Algorithm ID3 and Naive Bayes)

Miss Komal Varadkar[1], Miss Ankita Kharat[2], Mrs. Anushree Deshmukh[3], Miss Samiksha Dhulap[4], Miss.Prachi Gaikwad[5]

[1, 2, 4, 5]*BE Informaton Technology, Rajiv Gandhi Institute of Technology, Affiliated to Mumbai University, Mumbai, India*
[3]*Assistant Professor, Rajiv Gandhi Institute of Technology, Affiliated to Mumbai University, Mumbai, India*

Abstract: Farmers must be aware of macronutrients and micronutrients present in soil to gain maximum yield of particular crop and to know which fertilizer to use. This system will help farmers in yield prediction. Yield prediction is very popular among farmers these days, which particularly contributes to the proper selection of crops for sowing.
With the help of this system type of soil will be displayed. Also it will predict the crops suitable for the particular type of soil. In addition to this, it will 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. It is useful to farmers to get right insight to perform their activities with less cost and to improve the crop yields, such as by measuring soil properties the farmers can decide what kind of crops to be adopted and use of fertilizers etc. The soil analysis may use in many dimensions such as to protect the environment, diagnosis of crop culture troubles, to identify nutrient deficiencies, energy conversation, and so on

I. INTRODUCTION

Globally, day to day the need for food is escalating, therefore the agricultural scientists, farmers, government, and researchers are trying and using varied techniques in agriculture for improvement in production. As an impact, the information generated from the agricultural data is increasing day by day. Because the volume of information enlarged, it needs a spontaneous method for this information to be mined and analyzed once required. Data Mining techniques will be used for prediction the longer term trends of agricultural processes.
Data Mining techniques are of two types, one is descriptive which considers the existing data and another is predictive which depends on probability for future analysis. Data Mining process involves

A. Collect, clean and load the data into data warehouse system
B. Stores the data in multidimensional format
C. Provides information access to analysts and decision makers
D. Analyzation of data using different applications
E. Presents the data using different patterns (ISSN 2017)

II. EASE OF USE

From last few years, Data mining has attracted a great attention in the information industry due to large amount of data and the imminent need for using such data into useful information and knowledge. The information and knowledge gained can be used for application ranging from market analysis, fraud detection, production control, disaster management and science exploration. Data mining technique plays an important role in the analysis of data. Data mining is the process of analyzing patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database system. (Journal of bigdata 2017) In this digital era, analysts have colossal amounts of data available in data stores that make them difficult to be managed, processed or analyzed by using the typical database software tools & technologies. Data mining technologies are used to find relationship between data to analyze the large unstructured data which make it easier for the analyst to make a decision & generate knowledge from this data store according to business needs. To mine information from this large data store a data mining software tool called Waikato Environmental for knowledge Analysis (WEKA) is used. This tool has option to select appropriate algorithms and techniques for knowledge discovering process. Weka supports several data mining tasks, data preprocessing, clustering, classification, regression, visualization and features selection.

Knowledge discovery as a process consists of following steps:

A. Data cleaning, where it eliminate noise and inconsistent data.
B. Data integration, where many data sources are integrated.
C. Data selection, where data suitable for the analysis task are retrieved from the database
D. Data transformation, where data consolidated into forms.
E. Data mining is a process where intelligent methods are applied in order to bring out the data patterns.
F. Knowledge presentation, where visualization and knowledge representation techniques are used to present the mined knowledge to the user. (Journal of bigdata 2017)

Unsupervised (clustering) and supervised (classifications) are two different types of learning methods available in the data mining. Clustering is the process of examining a collection of "data points," and then categorize the data points into "clusters" according to some distance measure. Classification is the process that assigns data points in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data.
These are the following techniques used in data mining:

1) *Genetic Algorithm:* a genetic algorithm is a search method that imitate the process of natural selection. Genetic algorithms which produce solutions to optimization problems using techniques inspired by natural growth, such as inheritance, mutation, selection and crossover.
2) *Artificial Neural Network (ANN):* An Artificial Neural Network which is inspired by the method biological nervous systems, such as the brain, process data. It is composed of a large number of highly integrated processing elements (neurons) that's works to solve specific problems.
3) *Nearest Neighbor:* Nearest neighbor search, also called as proximity search, similarity search or closer point search, is a methodology problem for discovering the closest points..
4) *Rule Induction:* Rule Induction reasoning is a logical process in which multiple premises, all believed true are found true most of the time, are merged to secure a specific end. Inductive reasoning is often used in applications that involve prediction, forecasting, or behavior.
5) *Memory Based Reasoning:* MBR finds neighbors similar to a new record and uses the neighbors for classification and prediction. (International Journal of Engineering Trends and Technology 2011)

III. PROPOSED METHADOLOGY

There are 2 algorithms being compared in the project.

A. ID3(iterative Dicotomizer)Algorithm

Its is a non incremental algorithm means it derives its classas from a fixed set of training instances.so Iin our system we have dataset which contains attributes like p,n, with class labels which are m,h,l
By using this training data set we are going to classify the sample of soil into different fertility levels by taking its attribute. So for this we need to create a decision tree based on training dataset by using id3 algorithm.Adecision tree is a binary tree or multiple nodes tree which contains nodes which are all of given dataset branches values leaf node which are the class labels.while creating decision tree first we need to decide root node.for this the information gain of all attributes are calculated and the attributes with highest IG is selected as root node then the value becomes branches of tree and repeatedly using this formula DT is construced.
After constructing the DT the attributes of the soil which we want to test for fertility and finally the cass label for that unclassified record will obtained.

B. Naïve Bayes Algorithm

The main aim in the naïve bayes algorithm is to calculate the condition probability of an object with a feature vetor belongs to a particular class.So in this systemdifferent attribute of soil are considered to obtain the probability of samplesoil for different class with highest priority is selected as a class lable for that sample soil. So to find the probability.First need to find the frequency of each attribute against the different clses Then calculate the likehood table for the attribute against the classes
And then the use of N, B equation to calculate a propability for each class the class with highest probability is the outcome of the prediction.

Bayes theorem formula to calculate probability:

$$P(c|x) = \frac{P(x|c)\,P(c)}{P(x)}$$

$$P(c|x)=P(x1|c)xP(x2|c)x\ldots..xP(xn|c)xP(c).$$

P(c|x)=Posterior Probability.

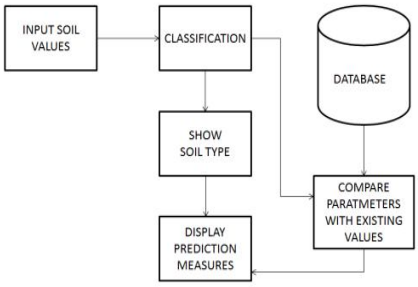
P(x|c)=Likelihood.

P(c)=Class Prior Probability.

P(x)=Predictor Prior Probability.

IV. PROPOSED SYSTEM

Using classification algorithms like ID3 and Naive Bayes with different attributes were predicted. According to these results the values of Phosphorous attribute was found to be most accurately predicted and it depends on least number of attributes.



A. ID3

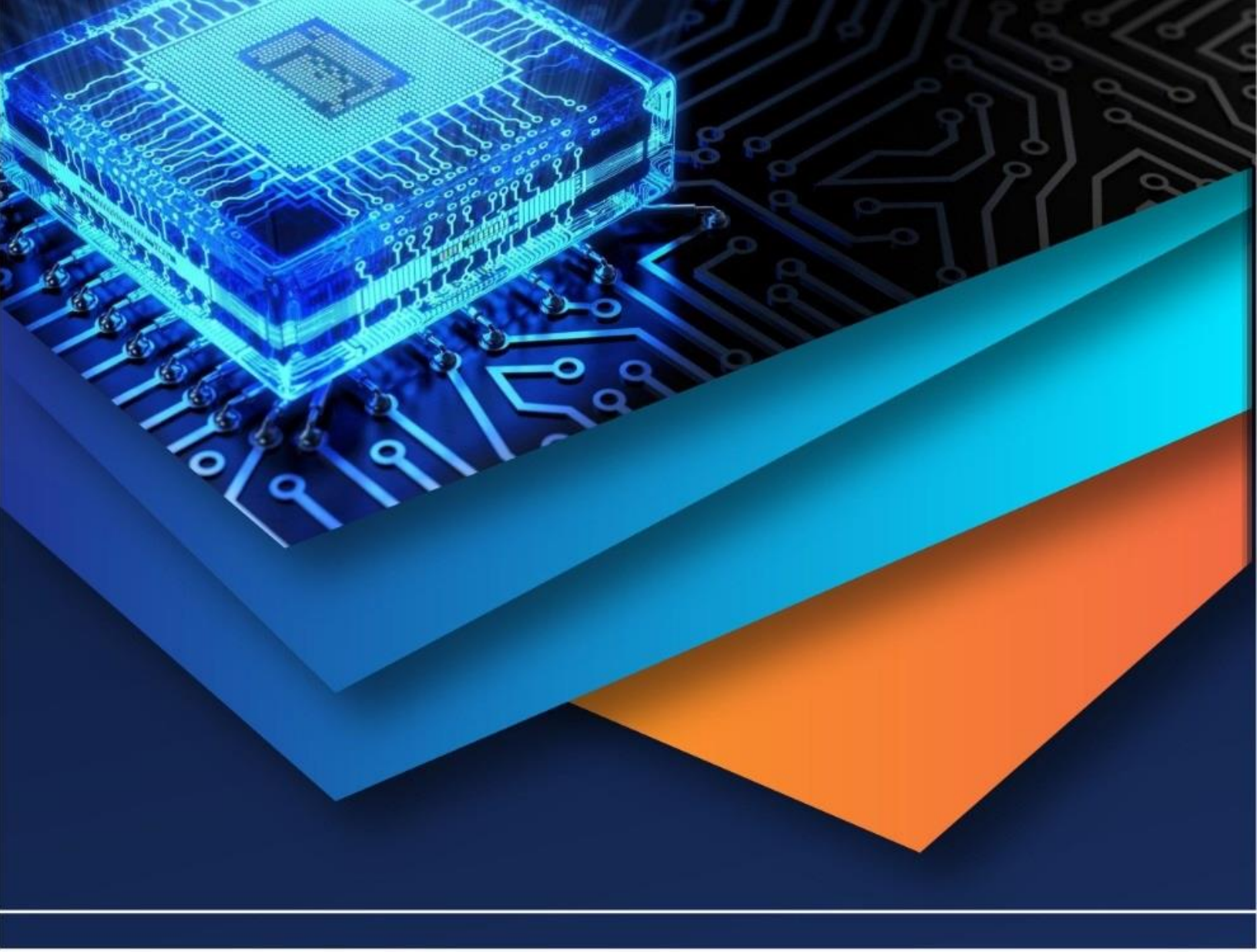
ID3 is a non-Incremental algorithm, meaning it derives its classes from a fixed set of training instances. The classes created by ID3 are inductive, that is, given a small set of training instances, the specific classes created by ID3 are expected to work for all future instances.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an analysis of the soil data using different algorithms and also classificaton technique. In spite the fact that the ID3 and Naïve Bayes algorithm which is known as classification technique from the given set of attributes, the most accurately predicted attribute was Phosphorous content of algorithm Id3 and Naive Bayes which was determined using the set of attribute.A comparative study of 2 classification algorithms i.e. Naive Bayes and ID3 with the help of data mining tool WEKA. ID3 is very simple classifier to make a decision tree, but it gave the best result in the experiment. In future, we contrive to build Fertilizer Recommendation System which can be utilized effectively by the Soil Testing Laboratories. This System will suggest appropriatesoil sample and cropping pattern.

REFERENCES

[1] Kumar (2011), “A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining “, IJCSI International Journal of Computer Science Issues.
[2] “Soil test”, Wikipedia, February 2012
[3] “Naïve Bayes”, Wikipedia, February 2012
[4] ”C4.5 (J48)”, Wikipedia, February 2012
[5] Cunningham & Holmes, (1999), ”Developing innovative applications in agriculture using data mining”,
[6] Vamanan & Ramar, (2011), “Classification Of Agricultural Land Soils A Data Mining Approach”, International Journal on Computer Science and Engineering, ISSN: 0975-3397, Vol. 3
[7] Bhargavi & Jyothi, (2011), “Soil Classification Using Data Mining Techniques.International Journal of Engineering Trends and Technology
[8] Han & Kamber, (2006), “Data Mining: Concepts and Techniques” Second Edition, San Francisco: Morgan Kaufmann Publishers



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY