IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Issue-6 E-ISSN: 2347-2693

Analysis of the Soil Data Using Classification Techniques for Agricultural Purpose

N. Hemageetha[1]* , G.M. Nasira[2]

[1] Research scholar, Department of computer Science, Periyar University, Salem, Tamil nadu ,

[2] Department of Computer Science , Chikkanna Govt. Arts College , Tirupur, India

www.ijcseonline.org

Received: May/28/2016 Revised: Jun/07/2016 Accepted: Jun/23/2016 Published: Jun/30/ 2016

***Abstract*—** India has more than sixty percent of its total area for agricultural purpose. Agricultural sector is the backbone for all the developing countries. Agriculture sector continue to play a vital role in the development of the economy. Data Mining supports decision making process and prediction. Agriculture needs the decision support system in variety of ways such as type of crop to be cultivated. There were many researches going on to support the agriculture using data mining. Analyzing soil provides major contribution to the support of the farmers. This paper analyzing Salem District soil data using data mining techniques.

***Keywords*—** Agricultural Soil, J48, Naïve Bayes Classification, Weka, Data mining.

I. INTRODUCTION

Data mining is the process of finding out interesting patterns and knowledge from large amount of data. There are various data analysis techniques available for agricultural research studies [1]. The motivation behind this paper is to explore data mining techniques, which are suitable for solving agricultural problems. The information about the previous work related to data mining in agriculture was collected [2]. Among the various trust areas of agricultural research as crop cultivation, market analysis, price prediction [3][4][5] and classification of soil plays an important role. This innovative application is useful for farmers to determine which crops should be produced in a particular type of soil. By analyzing the soil, guidance can be given on the type and amount of fertilizers that can be used in that soil [6].

In agriculture, with the help of soil test we can determine the fertility or the expected potential growth of soil that indicate the nutrient deficiencies from excessive fertility and inhibitions from the presence of non-essential trace minerals ("Soil Test", Wikipedia)..

Section II presents Materials and Methods. Section III presents a few Classification algorithms; Section IV presents Experiment and Results finally Section V presents Conclusion and Future Scope.

II. MATERIALS AND METHODS

Salem is one of the biggest districts of Tamil Nadu. Salem district is having administrative divisions of 9 Taluks, 20 Blocks, 376 Panchayats and 631 Revenue villages. The geographical area of the district is 5205.30 sq. kilometers. Salem is located at 11.669437°N , 78.140865°E, at an average elevation of 278 m (912 ft). The city is surrounded by hills. The soils of Salem District can be assorted into major type, Red Calcareous, Red non-calcareous, brown soil calcareous, Red colluvium calcareous, Red colluvium non-calcareous, Black soils, Alluvial calcareous, Brown soil non-calcareous.. Salem district receives major rainfall from the South West Monsoon followed by North East monsoon. Salem district receives maximum rainfall through northeast monsoon[7].

Soil consists of various nutrients present in organic and mineral forms. Soil testing is an important component of sustainable nutrient management in agriculture. Nitrogen, phosphorus, potassium are called major or macro nutrients because of their importance for a healthy and normal growth of a plant. Calcium , magnesium and sulfur are classified as secondary elements and rest are called trace nutrients or micronutrients. The secondary nutrients are not important components of plant tissue but are essential for growth. They include iron, manganese, zinc ,copper and boron but for all practical purposes, it is important to analyze the soil to find out the nutrient status in respect of Nitrogen, Phosphorus, Potassium . Both macronutrient and micronutrient availability are affected by soil pH[8].

Table 1 shows the macro and micro Soil Nutrients available in soil.

Table 1 Soil Nutrients description

Field	Description

Ph	PH value l
EC	Electric conductivity
OC	Organic Carbon
N	Nitrogen(pmm)
P	Phosphorous(ppm)
K	Potassium(ppm)
Fe	Iron(ppm)
Zn	Zinc (ppm)
Mn	Maganese (ppm)
Cu	Copper (ppm)

The soil analysis for the above parameters were carried out at Salem District. These parameters were taken for analysis due to the following reasons. The soil texture and fields productivity is determined by EC, if the soil is acidic or basic is indicates by PH. The macro nutrients N increases the crop yield , P and K increases the efficiency of fertilizers and yield. The micro nutrients such as Fe is used by plants during photosynthesis and respiration, Mn depletion leads to withering, Zn needed for a healthy growth, Cu prevents root damage , protects from fungus and Boron increases seed yield[8] for plants. Table 2 shows Nutrient rating of PH,EC,N,P and K [9] for plants growth.

Table 2 Nutrient rating of PH,EC,N,P and K

Parameter	Low	Medium	High
PH	<6.5 (Acidic)	6.5 - 7.5 (Neutal)	>7.5 (Alkaline)
Ec ($ds\ m^{-1}$)	<1.0 (Harmless)	1.0 -2.0 (Injurious)	>3.0 (Critical)
OC %	<0.5	0.5 – 0.75	>0.75
N ($Kg\ ha^{-1}$)	<280	280-450	>450
P ($Kg\ ha^{-1}$)	<11	11-22	>22
K ($Kg\ ha^{-1}$)	<118	118-280	>280

The soil is required to analyzed and measure its fertility. The soil by itself consists of various and minerals and salts. If the soil is analyzed and the deficiency of any of the minerals and salts are found. Then by making use of it the farmer can use the fertilizers. It is not required to use the fertilizers randomly, instead they can selectively use the fertilizers. The benefits of understanding of soils could improve productivity of crops, reduce reliance on fertilizers and create a best integrated soil management model for both the private and public sectors. Soil falling between PH – 6.5 to 8.5 is generally suitable for most of the common crops [11].

III. CLASSIFICATION ALGORITHMS

The classification of soil is difficult to study because it depends on the fertility class of the soil. The domain knowledge experts only determines which crops should be taken for the particular soil and which fertilizers should be used for the same. The following section describes Naive Bayes, J48, algorithms briefly

A. Naïve Bayes Classifier

Bayes theorem is developed by the British minister Bayes and it is named after him. The attributes are of two types as independent attributes and dependent attributes. In Naïve's approach it is assumed that all the attributes are independent of each other. Bayes theorem is based on the Naïve's assumption. In Naïve bayes Classifier, the value of the dependent attribute is calculated by using the values of the independent attributes.

The Naïve Bayes theorem is

$P(C_j \mid X) = [P(X \mid C_j)\, P(C_j)] / P(X)$

$P(C_j \mid X)$ is the probability of the object X belonging to class C_j

$P(X \mid C_j)$ is the probability of obtaining attribute values X if we know that it belongs to class C_j

$P(C_j)$ is the probability of any object belonging to class C_j

$P(X)$ is the probability of obtaining the attribute values of X

B. J48 (C4.5)

Decision trees are used in the data mining process. These decision trees are generated using the algorithm C4.5. This C4.5 algorithm is developed by Ross Quinlan. In this algorithm, the decision trees are generated by considering the set of labeled input data. In the data mining tool Weka, the C4.5 algorithm is implemented using JAVA and termed as J48. J48 is an open source tool.

C Random forest

The Random Forests algorithm is able to classify large amounts of data with accuracy. Random Forest is an Supervised learning method. Number of decision trees is constructed during training time [11]. Combination of tree predictors form the random forest, the each tree depends on the random vector values.

The basic idea behind Random forest algorithm is to group of "weak learners" can come together to form a "strong learner". It is a wonderful tool for predictions from large number of data. (Random forest "Wikipedia")

IV. EXPERIMENT AND RESULTS

A. *Dataset Collection*

Dataset required for this research was collected from Krishi Vigyan Kendra (Farm Science Centre) Tamil Nadu Agricultural University, Santhiyur, Salem. These datasets of soil samples taken from 11 block (Salem, Valapady, Panamarathupatti, Ayothiyapattinam, Attur ,Gangavalli,Omalur ,Sankari, Edappady, Kolathur, P.N.Palayam) out of 20 blocks in Salem District. Dataset has 11 attributes and a total 701 instances from 792 instances of soil samples.

B. *Data Formatting*

The WEKA 3.6.13 (Waikato Environment for Knowledge Analysis) workbench is the state of the art machine learning algorithms and data pre-processing tools. It is an open source software for Data Mining. WEKA is used to determine and predict result if any advantage would be gained in both interpretation of data set and time saving. The application of the data required for WEKA that some of pre-processing to be undertaken.

All the data are formatted into an Excel format based on various Blocks, soil types and relevant fields. Then converts all data sheets in to a single excel sheet. And the Excel data set is converted into .CSV file format to allow them to be applied to access in WEKA.

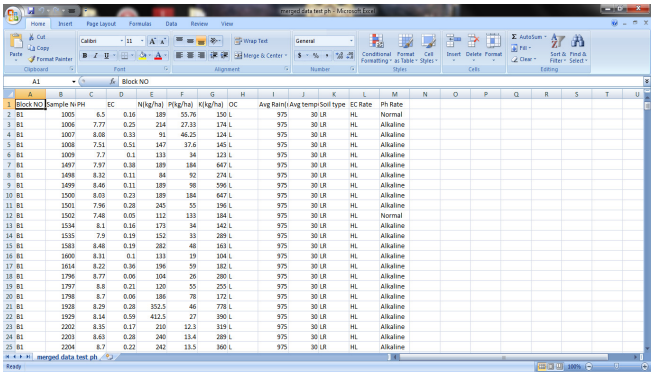

Fig 1 Excel data set

In this paper the Soil fertility is classified based on PH and EC values.

Random forest, J48 (C4.5), Naivebayes, BayesNet classifiers were evaluated and compared on basis of accuracy and mean absolute Error Rate. Tenfold cross-validation is also used in the experiment. Our studies showed that J48 model turned out to be best classifier for soil samples. In this paper Attribute selection also used. Attribute selection is reduce the dataset size by identify and removing redundant or irrelevant attributes Attribute selection finds minimum set of attributes.

Table3:Comparison of different classifiers (PH)

CLASSIFIER	Random Forest	Naivebayes	BayesNet	J48
CLASSIFIED INSTANCES	697	649	698	700
INCORRECTLY CLASSIFIED INSTANCES	4	52	3	1
ACCURACY	99.4294	92.582 %	99.572 %	99.8573 %
MEAN ABSOLUTE ERROR	0.0275	0.0579	0.0068	0.001

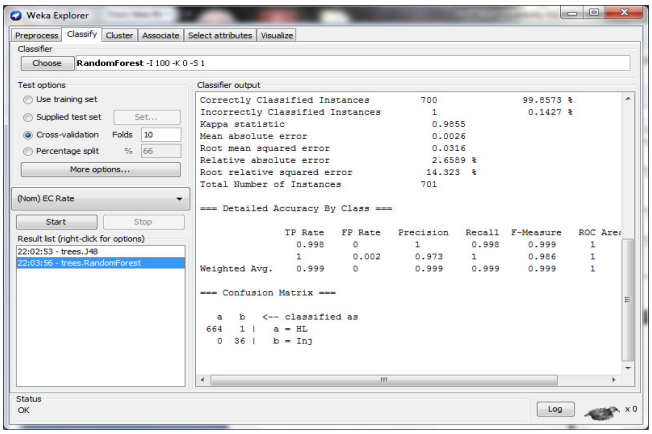

Fig 2 Weka J48 Classifier

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: merged data test ph-weka.filters.unsupervised.attribute.Remove-R9-10
Instances: 701
Attributes: 11
Block NO
Sample No
PH
EC
N(kg/ha)
P(kg/ha)

K(kg/ha)
OC
Soil type
EC Rate
Ph Rate

Test mode:10-fold cross-validation

=== Classifier model (full training set)

J48 pruned tree

PH <= 7.5
| PH <= 6.47: Acidic (15.0)
| PH > 6.47: Normal (114.0)
PH > 7.5: Alkaline (572.0)

Number of Leaves : 3
Size of the tree : 5
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 700 99.8573 %
Incorrectly Classified Instances 1 0.1427 %
Kappa statistic 0.9954
Mean absolute error 0.001
Root mean squared error 0.0308
Relative absolute error 0.4617 %
Root relative squared error 9.6353 %
Total Number of Instances 701

=== Detailed Accuracy By Class ===
TP FP Pre Recall F-mea ROC class
1 0.002 0.991 1 0.996 0.999 Normal 1 0 1
1 1 1 Alkaline
0.933 0 1 0.933 0.966 0.967 Acidic

Weighted Avg. 0.999 0 0.999 0.999 0.999
0.999
=== Confusion Matrix ===

a	b	c	<-- classified as
114	0	0	a = Normal
0	572	0	b = Alkaline
1	0	14	c = Acidic

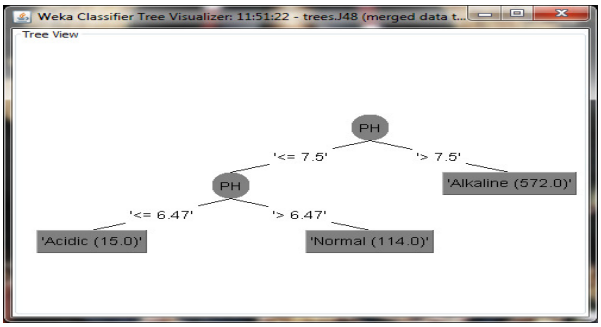

Fig 3 Tree View for PH Label

Test mode:10-fold cross-validation
Classifier model (full training set)
J48 pruned tree

EC <= 0.99: HL (665.0)
EC > 0.99: Inj (36.0)
Number of Leaves : 2
Size of the tree : 3
Time taken to build model: 0 seconds

Table 4:Comparison of different classifiers(EC)

CLASSIFIER	RANDOM FOREST	NAIVEBAYES	J48
CLASSIFIED INSTANCES	700	687	700
INCORRECTLY CLASSIFIED INSTANCES	1	14	1
ACCURACY	99.8573 %	98.002 %	99.8573 %
MEAN ABSOLUTE ERROR	0.0026	0.0334	0.0026

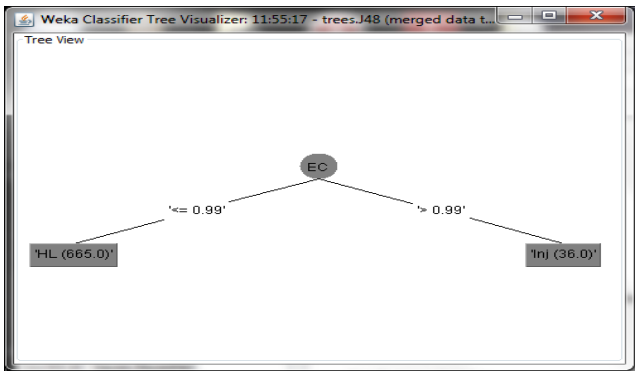

Fig 4 Tree view for EC

The results shows that the PH is High, EC is harmless , N and Zn is low, Mn,Fe,B,Cu values are high and P,K

values are medium or high in all most all the areas in Salem District.

V. CONCLUSION AND FUTURE SCOPE

Application of data mining techniques in agriculture an evolving field of research field there is a lot of work to be done. So the application of computer science in agriculture would help the agricultural field in improving the crop productivity and agricultural product marketing. Various decision tree algorithms can be used for classification of soil fertility. This paper shows that J48 gives 99.8573 % accuracy , hence it can be used as a base learner. In further research is to develop innovative model for agricultural soil classification based on the Salem District in Tamil Nadu. And to choose base crops for the farmers based on the available nutrients and soil condition.

ACKNOWLEDGMENT

The author gratefully acknowledges the support of Dr. Dr.N.Sriram, Programme Coordinator Krishi Vigyan Kendra (Farm Science Centre) Tamil Nadu Agricultural University Santhiyur, Salem, who help me more understanding of soil nutrients.

REFERENCES

[1] Mucherino.A, Petraq Papajorgji and P.M.Pardalos, "A survey of data mining techniques applied to agriculture". Published online **2009** © Springer-verlag.

[2] G.M. Nasira , N. Hemageetha ,"Perspective on Classification Techniques in Agriculture",International Journal of Computing Technology and Information Security Vol.1, No.2, pp. **40-46** , December, **2011**. ISSN: 2231-1998 © 2011.

[3] G.M. Nasira , N. Hemageetha, "Vegetable price prediction using data mining classification technique" Preceedings of the International Conference on pattern Recognition, Informatics and Medical Engineering (PRIME 2012), PP. **99-102** ISBN No:978-1-4673-1038-3. © **2012** IEEE.

[4] G.M. Nasira , N. Hemageetha, " Forecasting Model for Vegetable Price Using Back Propagation Neural Network" International Journal of Computational Intelligence and Informatics, Vol 2, no.2 sep **2012** PP **110-115**..

[5] G.M. Nasira , N. Hemageetha, " Radial Basis Function Model for Vegetable Price Prediction " Preceedings of the International Conference on pattern Recognition, Informatics and Mobile Engineering (PRIME 2013), PP. **424 – 428** ISBN No.978-1-4673-5843-9 © **2013** IEEE.

[6] N. Hemageetha, "A Survey on appliation of Datamining Techniques to Analyze the Soil for Agricultural purpose", Proceedings of the 10th INDIACom; INDIACom-2016; IEEE Conference ID: 37465 **2016** 3rd International Conference on "Computing for Sustainable Global Development", ISSN 0973-7529; ISBN 978-93-80544-20-IEEE.

[7] http://agritech.tnau.ac.in.

[8] Soil Testing Kit Hand Book.

[9] R.Santhi et at (**2014**) ,GIS based Soil map for salem district of Tamilnadu. TechInical Folder, TNAU,Coimbatore.

[10] Natesan et at(**2007**),. Technical Bulletinon "Soil test crop response based fertilizer prescription for different soils and crops in tamil nadu" ,AICRP-STCR TamilNadu Agricultural University, Coimbatore

[11] Random forest "Wikipedia".

.

AUTHORS PROFILE

.

N. Hemageetha. Received her MCA degree from University of Madras, Chennai, the M.Phil degree in computer Science from Bharathidasan University, Trichy and is currently pursuing the PhD degree in the research area of Data Mining under the guidance of Dr. G.M. Nasira. Currently she is working as Associate Professor in the Department of Computer Science, Govt. Arts College for Women, Salem 636008.

Dr. G M Nasira is working as Assistant Professor, Department of Computer Science, at Chikkanna Govt. Arts College , Tirupur. She has around 20 years of experience in teaching, at college level in various positions like Lecturer, Assistant Professor and Professor. She has published so far more than 40 research papers in referred journals and presented more than 40 papers in various conferences and seminars in addition she has also authored two books titled Fundamentals of Middleware Technologies and Web Technologies. Her research papers have won the Best paper award in two International conferences. She is a Recognized Supervisor to guide Ph.D in the area of Computer Applications for various Universities and also a Member of Board of Examiner to evaluate the Ph.D Thesis of various Universities. At present, 20 research scholars are doing their Ph.D. under her guidance. She has excellent track record in the administration of academic institutions in the capacity of Head of the Department and Vice-Principal. She got Best teacher award and her department got Best Department award twice under her headship.

