IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 6, Ver. III (Nov.-Dec. 2016), PP 50-54
www.iosrjournals.org

Analysis of Soil condition Based on pH value Using Classification Techniques

Mrs. N. Hemageetha[1], Dr. G.M. Nasira[2]
[1](Department of Computer Science, Periyar University, India)
[2](Department of Computer Application, Govt. Arts College Tirupur, India)

Abstract: *Agriculture is the backbone of developing countries. . Soil analysis is useful for farmers to determine which type of crops to be cultivated in a particular type of soil. Analyzing soil will prove to be a major benefit to the farmers. This paper analyzes whether the Salem district soil is suitable soil or non suitable soil for crops cultivate on based on pH value using data mining classification techniques.*

Keywords: *Agricultural Soil, Bayesian networks, Data mining, Naive Bayes , Jrip, J48.*

I. Introduction

The Finding out the interesting patterns and knowledge from large amount of data is the process of Data Mining. There are various research studies available for agriculture using data analysis techniques [1]. The aim of this paper is to explore data mining techniques, which are suitable for solving agricultural sector problems. In Tamil Nadu Salem is one of the biggest districts. This district has nine taluks, twenty blocks, three hundred and seventy six panchayats and six hundred and thirty one revenue villages. The geographical area of Salem district is 5205.30 sq. kilometers. It is located at 11.669437°N 78.140865°E, at an average elevation of 278 m (912 ft). This city is surrounded by hills. The soils of Salem District can be classified as Red soil, Black soil, Alluvial soil and Loamy soil. Salem district receives major rainfall from the South-West Monsoon followed by North--East monsoon. Salem district receives maximum rainfall through northeast monsoon [2].

Soil sample Analysis is called soil testing. It is used to determine the micro or macro nutrients level in soil, examine the extent of contaminated composition and measure the pH level to find out whether soil is acidic, saline or sodic. This paper focuses on classification of the soil condition based on the available attributes of soil database for Salem district.

II. Literature review

The information about the previous work related to data mining in agriculture was collected [3]. Among the various trust areas of agricultural research such as crop cultivation, market analysis and price prediction [4][5][6] classification of soil plays an important role. By analyzing the soil, guidance can be given on the type and amount of fertilizers that can be used in that soil [7].

Organic and mineral forms of various nutrients are present in soil. Soil testing is an important component of nutrient management in agriculture. Nitrogen, Phosphorus, Potassium are called major or macro nutrients. So they are very useful for a healthy and normal growth of a plant. Secondary nutrients like Calcium , Magnesium, Iron and Sulfur are not major components but they are also essential for plants growth. pH value of the soil will affect the availability of macronutrients and micronutrients [8].The soil analysis for the above parameters were carried out at Salem District. Table 1 shows the Nutrient rating of pH,EC,N,P and K [9] for plant growth.

Table 1 Nutrient rating of pH,EC,N,P and K

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

pH is the negative logarithm of the hydrogen ion activity otherwise it is the logarithm of the reciprocal of the H ion concentration.

PH = long 10 (H+) = log 10 (H)-1

Soil can be classified with the help of the available Nutrients. Normally pH value of Soil falls between 6.5 and 8.5 which is suitable for the of most of the crops [10]. pH range of soil in Salem District lies between 7.8 to 9.4. In and around the Salem District many Macro nutrients such as Nitrogen, Phosphorus, Potassium and Micro nutrients such as Iron, Manganese, Zinc are available. Sodic soils are characterized by the predominance of sodium in the complex with the exchangeable sodium percentage exceeding 15 per cent and the pH is more than 8.5.

The high amount of soluble salts (in saline soils) and of sodium present in the exchange complex (in sodic soils) hinder crop growth and rendered them barren [11].

Salt affected soils are the soils that contain considerable amounts of soluble salts and /or sodium on the exchange complex. In saline soils, the problem is high concentration of soluble salts in the rooting zone. Despite many limitations, the sodic soils, once ameliorated partially, by using gypsum technology, are used successfully for growth sodium- tolerant or semi tolerant crops, such as rice, sugar beet, wheat, barley etc.

III. Agriculture Data mining Algorithms

Classification of soil is critical to the study because depending upon the fertility class of the soil, the domain experts determine the type of crops to be cultivated in particular soil and also determine the type of fertilizers to be used for the same.

Naive Bayes, J48, *Bayesian networks* and Jrip classifiers as follows

3.1. Naïve Bayes Classifier

Bayes theorem developed by the British minister Bayes has been named after him. There are two types of attributes. They are 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.

3.2. J48 (C4.5)

Decision trees are used in the data mining process. Using the algorithm C4.5 the decision trees are generated. Ross Quinlan developed the C4.5 algorithm. The decision trees are generated using set of labeled input data. In the data mining tool Weka, the C4.5 algorithm is implemented and termed as J48 by using JAVA.

3.3. Bayesian Network

Graphical model is used in Bayesian network. It is used to learn relationships between the items and can also be used to understanding the problem domain and consequences of intervention.

3.4. JRip

JRip algorithm implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction, it was proposed by William W. Cohen. It is an optimized version of IREP. It divides the dataset into two sets named as growing set and pruning set. Then it reduces the errors in both the sets and generates rules for the sets.

IV. Experiments and results

4.1. Data Set Collection

Soil Dataset was collected from Krishi Vigyan Kendra (Farm Science Centre) Tamil Nadu Agricultural University, Santhiyur, Salem. These datasets contain various attributes. Soil samples have been taken from eleven blocks (Salem, Valapady, Panamarathupatti, Aothiyapattinam, Attur, Gangavalli, Omalur ,Sankari, Edappady, Kolathur, P.N.Palayam) out of 20 blocks in Salem District. Dataset has the following attributes Sample no., Block no, Soil Type (Red soil, black soil, Alluvial, and loamy soil),PH value, Electric conductivity(EC), Organic Carbon(OC), Phosphorous (P), Potassium(K),Nitrogen (N) , ProbSoil and a total of 701 instances from 792 instances of soil samples.

4.2. Data Formatting

The WEKA 3.6.13 (Waikato Environment for Knowledge Analysis) workbench is a state of art for machine learning algorithms and data pre-processing tools. It is open source software for Data Mining. All the data are formatted into an Excel format based on various Blocks, soil types and relevant related fields. All sheets are converted into a single excel sheet, which are again converted into .CSV file format to be accessed in WEKA. The Methodology flow is given in Fig 1.

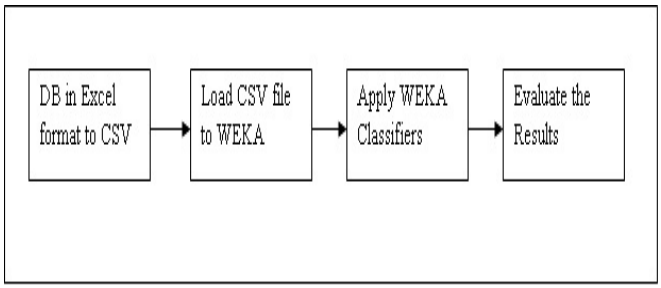
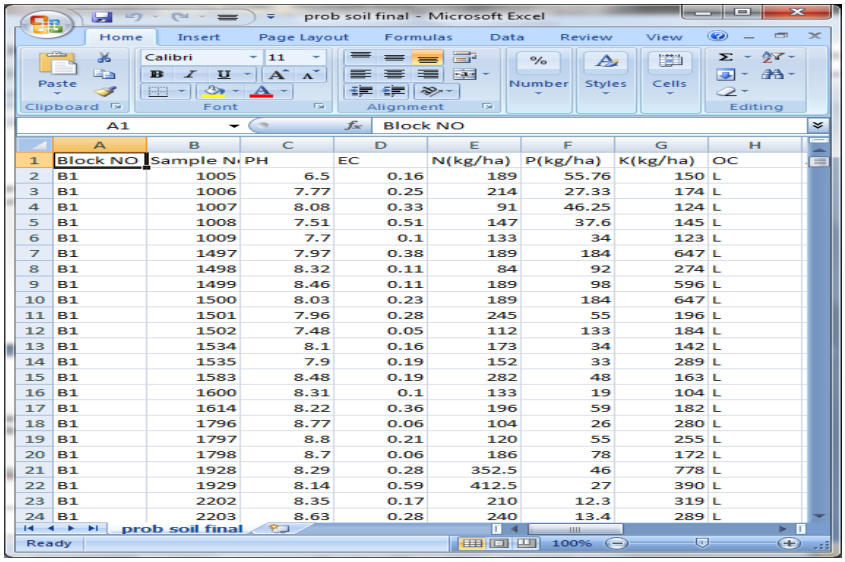


Fig 1 Methodology Flow Chart

From the data base collected out of 792 instances of soil samples 701 instances have been considered for proposed methodology. Instances with missing attributes values, noisy data and miss match. They are filtered using Weka Filters. The Preprocessed Soil data set is shown in the Fig 2.



	A	B	C	D	E	F	G	H
1	Block NO	Sample N	PH	EC	N(kg/ha)	P(kg/ha)	K(kg/ha)	OC
2	B1	1005	6.5	0.16	189	55.76	150	L
3	B1	1006	7.77	0.25	214	27.33	174	L
4	B1	1007	8.08	0.33	91	46.25	124	L
5	B1	1008	7.51	0.51	147	37.6	145	L
6	B1	1009	7.7	0.1	133	34	123	L
7	B1	1497	7.97	0.38	189	184	647	L
8	B1	1498	8.32	0.11	84	92	274	L
9	B1	1499	8.46	0.11	189	98	596	L
10	B1	1500	8.03	0.23	189	184	647	L
11	B1	1501	7.96	0.28	245	55	196	L
12	B1	1502	7.48	0.05	112	133	184	L
13	B1	1534	8.1	0.16	173	34	142	L
14	B1	1535	7.9	0.19	152	33	289	L
15	B1	1583	8.48	0.19	282	48	163	L
16	B1	1600	8.31	0.1	133	19	104	L
17	B1	1614	8.22	0.36	196	59	182	L
18	B1	1796	8.77	0.06	104	26	280	L
19	B1	1797	8.8	0.21	120	55	255	L
20	B1	1798	8.7	0.06	186	78	172	L
21	B1	1928	8.29	0.28	352.5	46	778	L
22	B1	1929	8.14	0.59	412.5	27	390	L
23	B1	2202	8.35	0.17	210	12.3	319	L
24	B1	2203	8.63	0.28	240	13.4	289	L

Fig 2 Salem District Soil Data Set

Soil condition is classified as suitable soil or non-suitable soil based on pH value of soil. The condition is if the pH value is greater than 8.5 then the soil is Non-suitable soil otherwise it is suitable soil for crop cultivation J48 (C4.5), Naivebayes, BayesNet classifiers, Jrip have been generated for the preprocessed soil dataset. The screen shot of J48 classifier for soil condition is given in Fig 3.

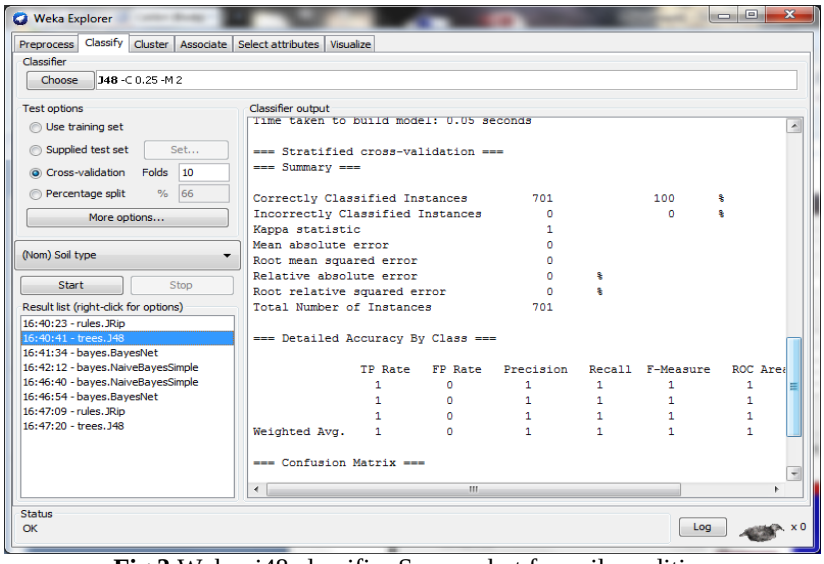


Fig 3 Weka- j48 classifier Screen shot for soil condition

The J48 classifier result for the soil is shown in Fig 4.

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

PH <= 8.5
|   PH <= 6.63
|   |   Ph Rate = Normal: NO (6.0)
|   |   Ph Rate = Alkaline: Acidic (0.0)
|   |   Ph Rate = Acidic: Acidic (15.0)
|   PH > 6.63: NO (604.0)
PH > 8.5: Problm soil (76.0)

Number of Leaves  :     5

Size of the tree :     8


Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         701               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances              701

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC
Area  Class
                 1         0          1         1         1          1
NO
                 1         0          1         1         1          1
Problm soil
                 1         0          1         1         1          1
Acidic
Weighted Avg.    1         0          1         1         1          1

=== Confusion Matrix ===

   a   b   c   <-- classified as
 610   0   0 |   a = NO
   0  76   0 |   b = Problm soil
   0   0  15 |   c = Acidic
```

Fig 4 J48 classifier result

The accuracy and Mean absolute Error Rates are compared. The J48 classifier shows the best results compared with other classifiers. The Correctly classified Instances, Incorrectly classified Instances ,Accuracy and Mean Square Error are given in the Table 2.

Table 2: Comparison of different classifiers results for Soil pH condition

CLASSIFIER	SimpleNaive bayes	BayesNet	J48	Jrip
Correctly Classified Instances	667	700	701	699
Incorrectly Classified Instances	34	1	0	2
Accuracy	95.14%	99.84%	100 %	99.71%
Kappa Statistics	0.78	0.99	1	0.99
Mean Absolute Error	0.04	0.00	0	0.00

The results show that accuracy of J48 is high compared with Naive bayes , Jrip and BayesNet. The Mean absolute Error Rates is less compared with other classifiers. The kappa statistics is compared on the basis of Tenfold cross-validation. Out of 701 instances J48 has classify 701 instances correctly, the Accuracy is 100% which is high compared with other classifiers. The number of instances currently classified and incorrectly classified is also given in the fig 5.

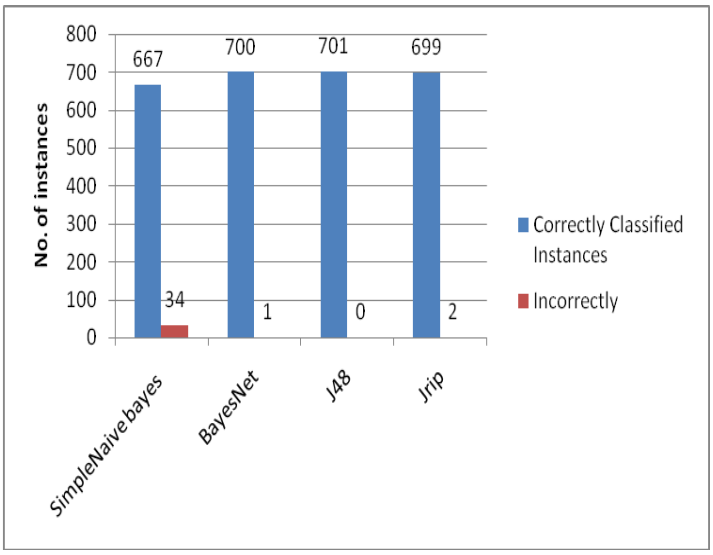


Fig 5 Instances are Correctly or incorrectly classified

This result shows that the major part of the Salem district soil is suitable for cultivation of many corps like Paddy, Cholam, Ragi, Maize, Green gram, Cowpea, Mochai, Horse gram, Groundnut and Sugarcane. J48 classifier results show more accuray than the other classifier.

V. Conclusion And Future Scope

Data mining in agriculture will help the agricultural field to improve the crop productivity. Various decision tree algorithms are used for classification of soil type. Salem district soil is analyzed based in its pH value and find out whether it is suitable or not suitable for crops cultivation. The results show that accuracy of J48 is high compared with Naive bayes , Jrip and BayesNet. And also it shows that the major part of the Salem district soil is suitable for cultivation of many corps. In future an Innovative model has been developed to choose base crops for the farmers based on the soil type in Salem District.

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

[3]. Dr. 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.* *www.ijctis*

[4]. Dr. 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.*

[5]. *Dr.* 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.*

[6]. Dr. G.M. Nasira , N. Hemageetha, Radial Basis Function Model for Vegetable Price Prediction, Proceedings 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.*

[7]. 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.

[8]. 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.

[9]. Soil Testing Kit Hand Book

[10]. National Agricultural Development Programme (NADP)

[11]. Gupta, R.K. and I.P. Abrol. 1990. Salt–affected soils – Their reclamation and management for crop production. Advances in soil science **12**,223-275.