IJCTA, 9(40), 2016, pp. 57-66

Availability of Macro Nutrients Status in Salem District Soil Using Data Mining Classification Techniques

N. Hemageetha* and G.M. Nasira**

Abstract : Agriculture sector plays a vital role in developing countries. Main goal of this work is to classify the soil macro nutrients using classification techniques in data mining for Salem District Soil dataset. Soil macro nutrient availabilities are predicted with the help of classification techniques like Naive Bayes, J48, Bayesian networks. These techniques are used to extract useful information from the soil dataset. The J48 classifier produces the best result for Macro nutrients availability in soil. This Analyses will help the farmers to use which type of fertilizers their agriculture field.

Keywords : J48, Naive Bayes, Bayesian networks, Soil Macro Nutrients.

1. INTRODUCTION

Data mining is the process of discovering knowledge from large amount of data. There are various data analysis techniques used for agricultural research studies[1]. previous papers were collected related to data mining in agriculture [2].

In[3][4] the vegetable price prediction is analyzed for using data mining techniques. The analysis of soil plays an important role in Agriculture which is useful for farmers to determine the type of crops to be cultivated in a particular type of soil[5]. In[6] Soil type is Classified using data mining techniques. Soil test is used to determine the nutrient level of soil. It examines the extent of contaminated composition and measures the pH level to find whether soil is acidic, saline or sodic [7].

The motivation behind this paper is to explore data mining techniques, which are suitable for solving agricultural problems. Salem is one of the biggest Districts in Tamil Nadu. There are nine Taluks, twenty blocks, Three hundred and seventy six Panchayats and six hundred and thirty one revenue villages in salem district. The geographical area of Salem district is 5205.30 sq. kilometers. Salem 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[8].

Soil testing is an important part of nutrient management in agriculture sector. Sixteen nutrients are available in soil which is absorbed by the plants for their growth. They are divided into macro primary, macro secondary and micro nutrients[9]. Nitrogen, phosphorus, potassium are called major or macro nutrients because of their importance for a healthy and normal growth of a plant. Secondary nutrients such as Calcium, magnesium and sulfur are not major components but they are also essential for the growth of the plant. Iron, Boron, zinc, manganese, molybdenum, nickel, cobalt, chlorine are micro nutrients. In and around the Salem District many Macro nutrients such as Nitrogen, Phosphorus, Potassium and

* Department of computer Science Periyar University Tamil Nadu, India *geekani2010@gmail.com*

** Department of Computer Application Chikkanna Govt. Arts College Tirupur, Tamil Nadu, India *nasiragm99@yahoo.com*

Micro nutrients such as Iron and Zinc are available. pH value of the soil will affect the availability of macronutrients and micronutrients[10].

A. Macro Nutrients Funtions

1. Nitrogen (N) is required by plants for greenish and leaf growth. It leads to seed and fruit production. All plants cells have nitrogen. The source of nitrogen for the plant is the atmospheric nitrogen.
2. Phosphorous (P) help the plants to make use of the light energy to produce its food. It is very important for formation of protein, germination and also all process of growth of the plant.
3. Potassium (K) - plants absorb a large amount of potassium than other minerals. It helps in production of good quality of fruits and the reduction of diseases. It is very important for formation of starch, protein synthesis, sugar and carbohydrate.
4. Calcium (Ca)- Calcium is essential for plants cell elongation , transport processing and development of roots.
5. Sulphur (S)-Vitamin's structured component is sulphur . It is very essential for plants chloroplasts manufacturing.
6. Magnesium(Mg)- it is used in plants photosynthesis process. Which gives green color to the leaves. Magnesium deficiency in soil can be solved by application of rich organic compost.

From the above macro nutrient list, the primary macro nutrients NPK are sufficient to determine the soil fertility.

Table 1 shows the Nutrient rating[10] of N, P and K for plants growth.

Table 1
macro nutrient ratind of N, P and K

S. No	*Level*	*Low*	*Medium*	*High*
1.	N(Kg ha^{-1})	<280	280–450	>450
2.	P(Kg ha^{-1})	<11	11–22	>22
3.	K(Kg ha^{-1})	<118	118–280	>280

If NPK levels are low then the soil fertility level is low so the farmers are recommend to give more fertilizer to soil for enrich the soil fertility. In this paper the NPK level is analyzed for the salem district soil.

Soil can be classified with the help of the available Nutrients. Generally pH value of Soil lies between 6.5 and 8.5 which is suitable for most of the common crops growth[8]. 7.8 to 9.4 is the pH range of soil in Salem District. In[9] the pH and EC levels are analyzed and find out that the soil pH level is high and EC level is harmless in Salem district. The main aim of this paper is the analysis of soil macronutrients availability in Salem district in Tamil Nadu with the help of data mining techniques.

2. MATERIALS AND METHODS

A. Dataset Colletion

In this paper soil dataset for Eleven blocks in Salem district is collected from Farm Science center , Santhiyur, Salem. Attributes in the dataset are sample no, block no, soil type, pH , Electric conductivity ,organic carbon, phosphorus rate, Nitrogen rate and Potassium rate. The level of Nitrogen, Phosphorous and the potassium is used to determine whether the soil is fertile soil or not for the crops cultivation. Based on the N, P ,K level the farmers can use the fertilizer for their crop cultivation.

B. 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 an 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. Figure 1 shows the flow chart of the Process.

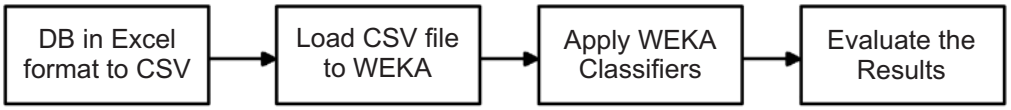
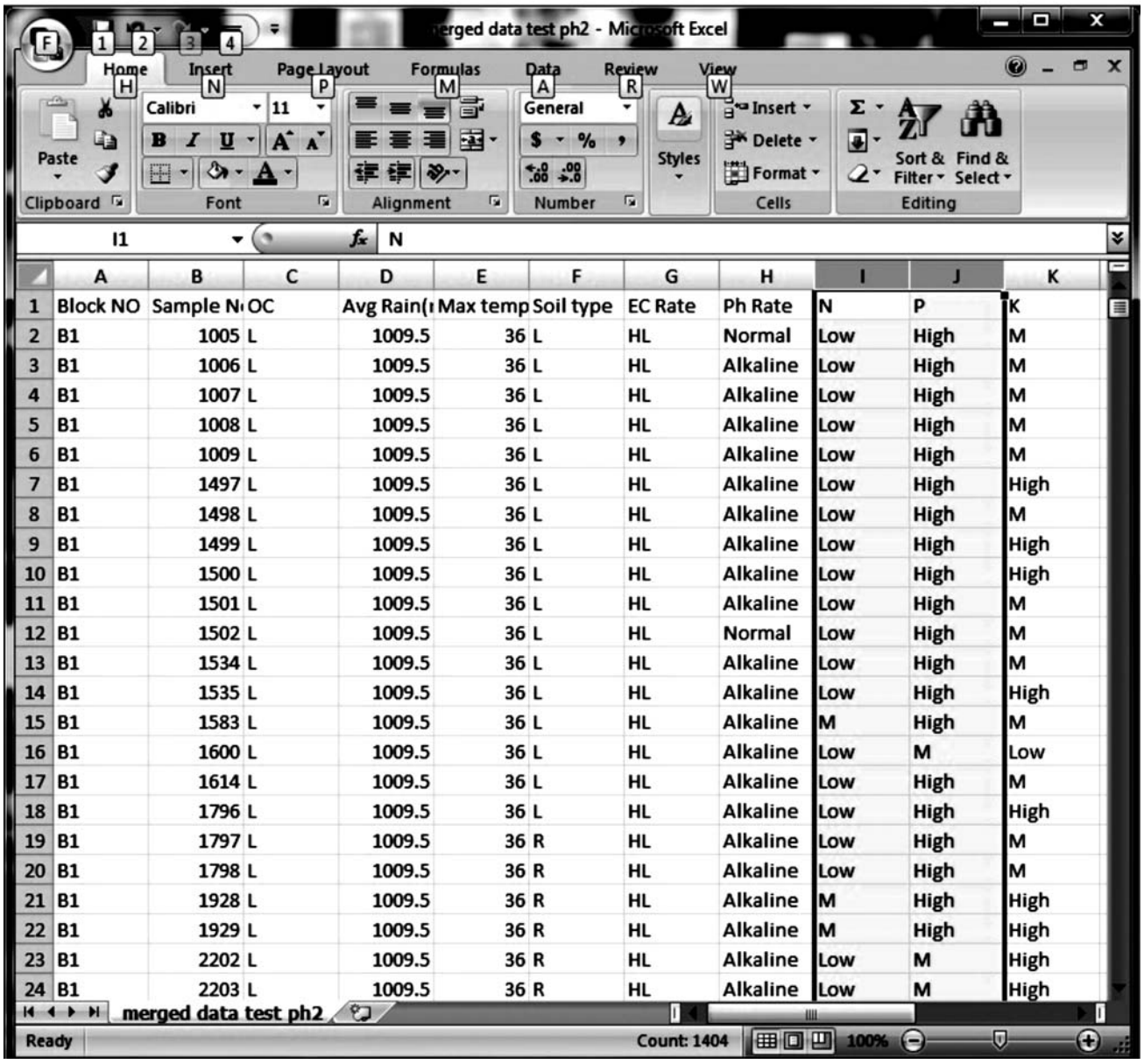


Figure 1: Proposed Setup and its Flow

From the data base collected out of 792 instances of soil samples 701 instances has been considered for our proposed methodology. The instance which had missing attributes values, noisy data and miss match. They are filtered using Weka filters like InterqartileRange, Removewithvalue. The Preprocessed Soil data set is shown in the Fig 2.



	A	B	C	D	E	F	G	H	I	J	K
1	Block NO	Sample N	OC	Avg Rain(r	Max temp	Soil type	EC Rate	Ph Rate	N	P	K
2	B1	1005	L	1009.5	36	L	HL	Normal	Low	High	M
3	B1	1006	L	1009.5	36	L	HL	Alkaline	Low	High	M
4	B1	1007	L	1009.5	36	L	HL	Alkaline	Low	High	M
5	B1	1008	L	1009.5	36	L	HL	Alkaline	Low	High	M
6	B1	1009	L	1009.5	36	L	HL	Alkaline	Low	High	M
7	B1	1497	L	1009.5	36	L	HL	Alkaline	Low	High	High
8	B1	1498	L	1009.5	36	L	HL	Alkaline	Low	High	M
9	B1	1499	L	1009.5	36	L	HL	Alkaline	Low	High	High
10	B1	1500	L	1009.5	36	L	HL	Alkaline	Low	High	High
11	B1	1501	L	1009.5	36	L	HL	Alkaline	Low	High	M
12	B1	1502	L	1009.5	36	L	HL	Normal	Low	High	M
13	B1	1534	L	1009.5	36	L	HL	Alkaline	Low	High	M
14	B1	1535	L	1009.5	36	L	HL	Alkaline	Low	High	High
15	B1	1583	L	1009.5	36	L	HL	Alkaline	M	High	M
16	B1	1600	L	1009.5	36	L	HL	Alkaline	Low	M	Low
17	B1	1614	L	1009.5	36	L	HL	Alkaline	Low	High	M
18	B1	1796	L	1009.5	36	L	HL	Alkaline	Low	High	High
19	B1	1797	L	1009.5	36	R	HL	Alkaline	Low	High	M
20	B1	1798	L	1009.5	36	R	HL	Alkaline	Low	High	M
21	B1	1928	L	1009.5	36	R	HL	Alkaline	M	High	High
22	B1	1929	L	1009.5	36	R	HL	Alkaline	M	High	High
23	B1	2202	L	1009.5	36	R	HL	Alkaline	Low	M	High
24	B1	2203	L	1009.5	36	R	HL	Alkaline	Low	M	High

Figure 2: Salem District pre proceed Soil Data set

C. Classifiation Algorithms in Agriculture

Classification of soil is critical to our proposed 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. 701 samples were processed through three classifiers (Naive Bayes, J48, Bayesian networks) for obtaining better Accuracy and less Error rate. The results of the three classifiers are better than other classifiers like ZeroR, Decision Table, oneR and BFtree. So the three classifiers only taken for this paper.

Other existing classifiers used among various researchers were also compiled with soil data which lead to less accuracy and more Error rate[11]. Naive Bayes, J48, Bayesian networks algorithms as follows :

1. Naïve Bayes Classifier

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

2. J48(C4.5)

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

The steps of J48 :

1. Create a node K
2. If instances are all the same class C1 then
   Return K as a leaf node with C1 as the label.
3. If list of attribute is null then
   Return K as a leaf.
4. The highest information gain attribute is selected for the test-attribute
5. For each j_i of test –attribute generate a branch from K.
6. If S1 is null then
   Attach the leaf node with the common class in instance
   Else
   Attach the node return by tree generate function

3. Bayesian networks

A Bayesian network is based on graphical model. Bayesian network is used to learn relationships between the items and can be used to understand the problem domain and consequences of intervention

3. RESULTS AND DISCUSSION

In this paper the Soil macro nutrients are classified based on the available parameters. J48 (C4.5), Naivebayes, BayesNet classifiers have been generated. The results of classifiers are compared based on its Accuracy and Mean absolute Error Rates. This is implemented in Weka data mining tool. Analysis of the level of N, P and K are as follows

A. Analysis of Nitrogen (N) in soil

Nitrogen (N) level in soil is classified based on the N value in soil. If N value is less than 280 then the content of N in soil is Low, If N value lies between 280 to 450 then the content of N in soil is Medium otherwise it is High. J48 (C4.5), Naïve bayes, BayesNet classifiers have been generated. The Weka Screen shot of J48 for Nitrogen is given in Figure 3.

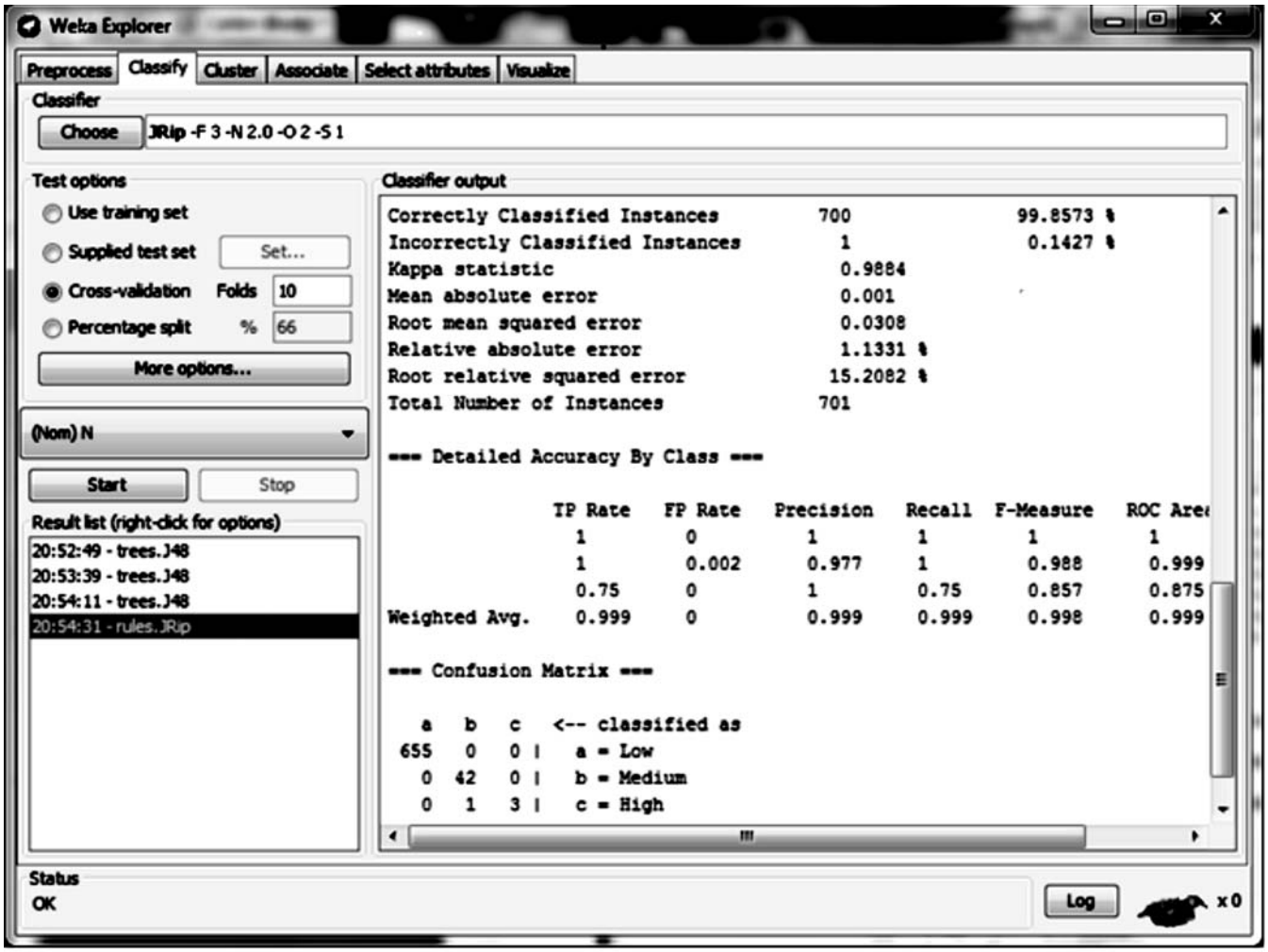


Figure 3: J48 -Weka Screen Shot for Nitrogen(N) classification

The results of all the classifiers are compared based on its Accuracy and Mean absolute Error Rates. The J48 classifier shows the best results compared with other classifiers. Comparison of different classifiers results for Nitrogen (N) is given in Table 2.

Table 2

Comparison Results of clasifier for nitrogen

S.No	*Classifier*	*Naive bayes Simple*	*BayesNet*	*J48*
1.	Correctly Classified Instances	680	693	699
2.	Incorrectly Classified Instances	21	8	2
3.	Accuracy	97.00%	98.56%	99.71 %
4.	Kappa Statistics	0.77	0.91	0.98
5.	Mean Absolute Error	0.02	0.01	0.00

Table 2 shows that accuracy of J48 classifier result is high compared with Naive bayes and BayesNet. The Mean absolute Error Rates is less compared with other classifiers. The kappa statistics are compared on the basis of Tenfold cross-validation. Out of 701 instances J48 classify 699 instances correctly, the Accuracy is 99.71% which is high compared with other classifiers. The number of instances currently classified and incorrectly classified is also given in the Figure 4.

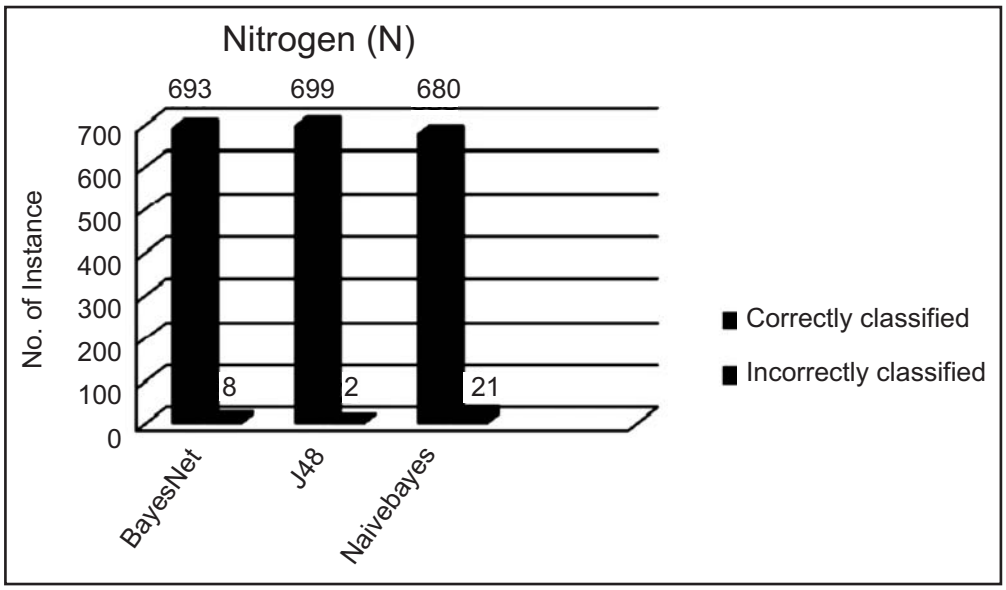


Figure 4: The number of instances currently classified for N

The Confusion Matrix in the result shows that the Nitrogen (N) availability is low in the Salem Districts.

B. Analysis of Phosphorus (P) in soil

Phosphorous Rate in soil is classified based on the P value in soil. If P is less than 11 then the content of P in soil is Low, If P value lies between 11 to 22 then the content of P in soil is Medium otherwise it is High. J48 (C4.5), Naivebayes, BayesNet classifiers have been generated. The Weka Screen shot of J48 for Phosphorus is given in Figure 5.

The results of all the classifiers are compared based on its Accuracy and Mean absolute Error Rates. The J48 classifier shows the best results compared with other classifiers. Comparison of different classifiers results for P is given in Table 3.

Table 3

Comparison Results of clasifier for phosphorus

S.No	*Classifier*	*Naive bayes Simple*	*BayesNet*	*J48*
1.	Correctly Classified Instances	650	695	700
2.	Incorrectly Classified Instances	51	6	1
3.	Accuracy	92.72%	99.14%	99.86 %
4.	Kappa Statistics	0.82	0.98	0.97
5.	Mean Absolute Error	0.07	0.01	0.00

Table 3 the results for P shows that accuracy of J48 classifier is high compared with Naive bayes and BayesNet. The Mean absolute Error Rates is less compared with other classifiers. The kappa statistics are compared on the basis of Tenfold cross-validation. Out of 701 instances J48 classify 700 instances correctly, the Accuracy is 99.85% which is high compared with other classifiers. The number of instances currently classified and incorrectly classified is also given in the Figure 6.

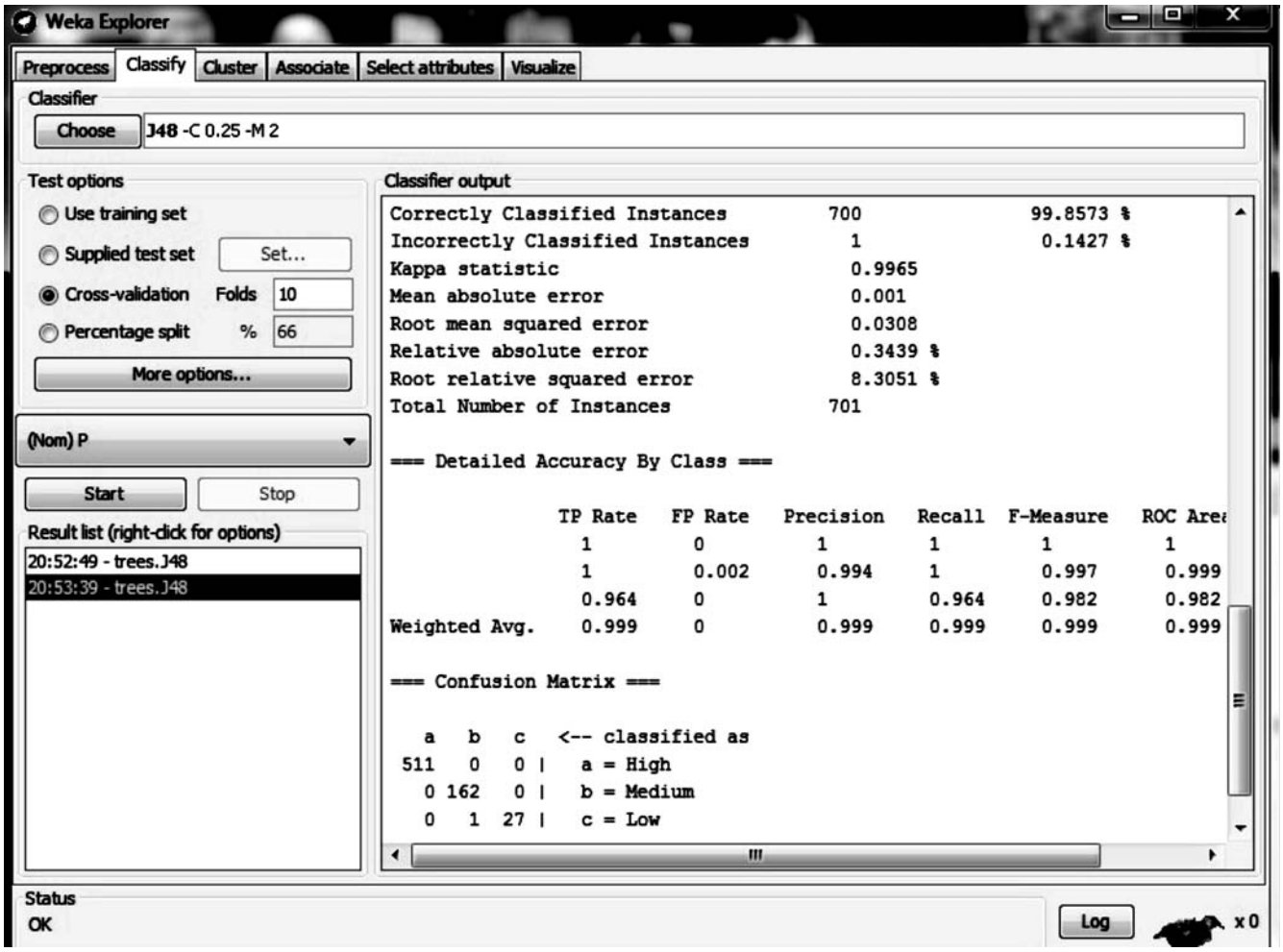


Figure 5: J48 -Weka Screen Shot for Phosporos(P) classification

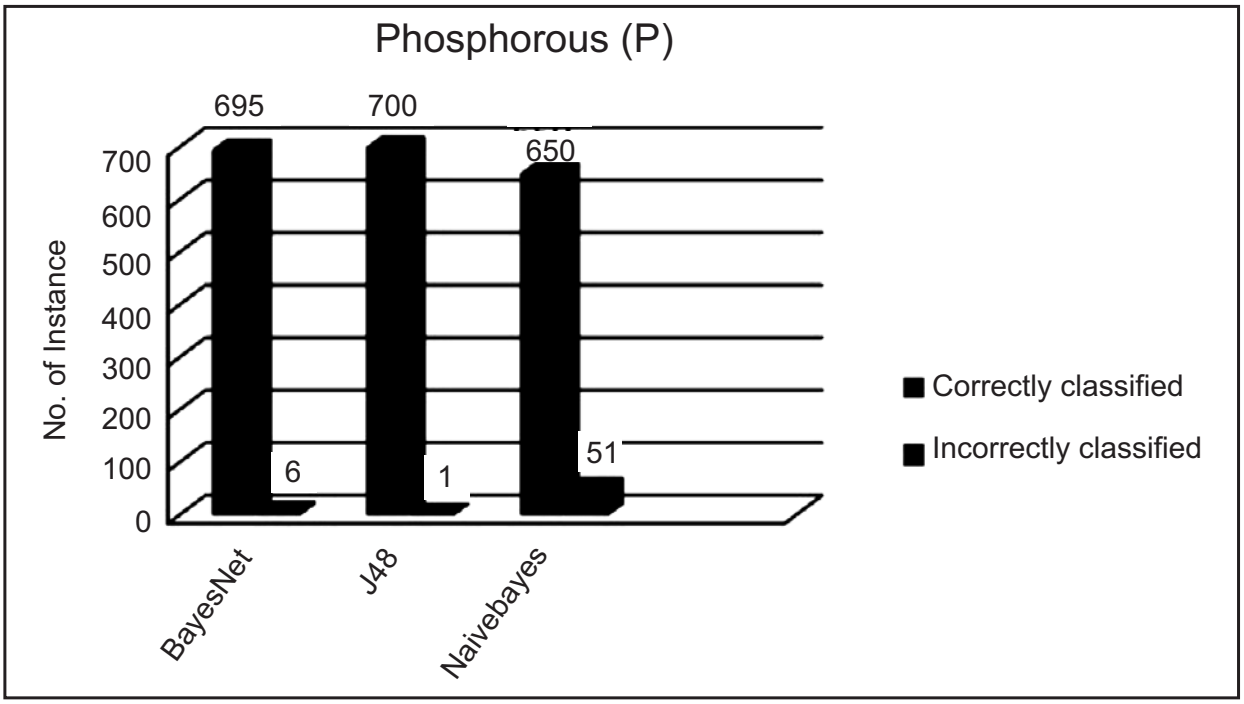


Figure 6: The number of instances currently classified for P

C. Analysis of Potassium (K) in soil

Potassium rate in soil is classified based on the K value in soil. If K is less than 118 then the content of K in soil is Low, If K value lies between 118 to 280 then the content of K in soil is Medium otherwise High. J48 (C4.5), Naivebayes, BayesNet classifiers have been generated. The Weka result shot of J48 for Potassium is given in Figure 7.

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

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

K(kg/ha) <= 279
|   K(kg/ha) <= 117: Low (41.0)
|   K(kg/ha) > 117: Medium (314.0)
K(kg/ha) > 279: High (346.0)

Number of Leaves  :   3

Size of the tree  :   5


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 square 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
Medium
                   1         0          1         1         1         1
High
                   1         0          1         1         1         1
Low
Weighed Avg.       1         0          1         1         1         1

=== Confusion Matrix ===

   a   b   c   <-- classified as
 314   0   0 |   a = Medium
   0 346   0 |   b = High
   0   0  41 |   c = Low
```

Figure 7: J48 -Weka result for Potassium(K) classification

The results of all the classifiers are compared based on its Accuracy and Mean absolute Error Rates. The J48 classifier shows the best results compared with other classifiers. Comparison of different classifiers results for K is given in Table 4.

Table 4

Comparison Results of clasifier for potassium

S.No	*Classifier*	*Naive bayes Simple*	*BayesNet*	*J48*
1.	Correctly Classified Instances	626	698	701
2.	Incorrectly Classified Instances	75	3	0
3.	Accuracy	98.56%	99.57%	100 %
4.	Kappa Statistics	0.91	0.99	0
5.	Mean Absolute Error	0.01	0.00	0

From the table IV the results for Potassium shows that accuracy of J48 is high compared with Naive bayes and BayesNet. The Mean absolute Error Rates is less compared with other classifiers. The kappa statistics are compared on the basis of Tenfold cross-validation. Out of 701 instances J48 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 Figure 8.

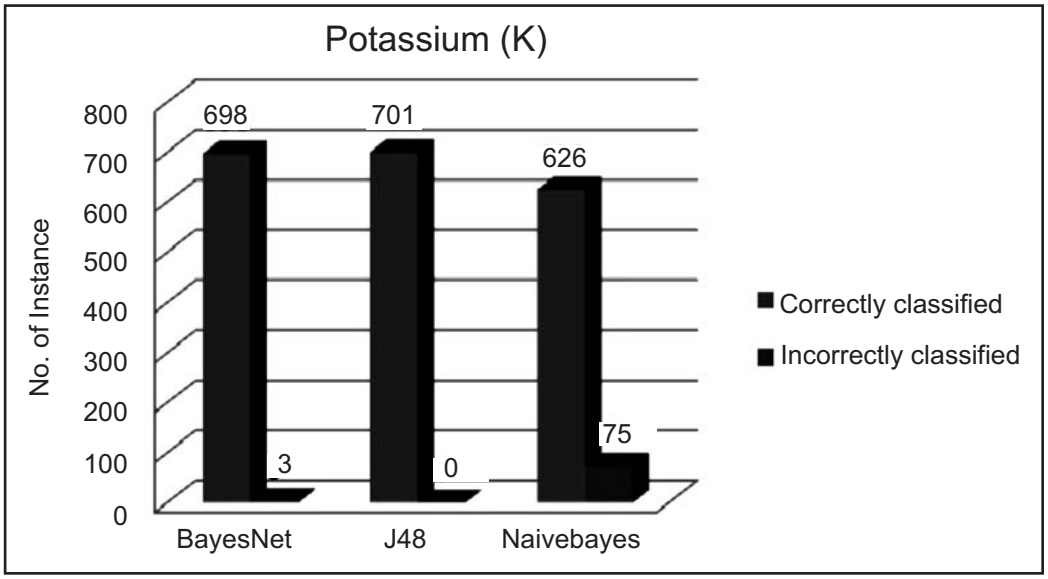


Figure 8: The number of instances currently classified for K

The Confusion Matrix in the classifier results shows that the Nitrogen availability is low, Phosphorus availability is high to medium and Potassium is medium to high all most all the areas of Salem District soil.

4. CONCLUSION

The application of Data Mining techniques will help the agricultural field to improve the crop productivity. Various decision tree algorithms are used for classification of the soil. In this paper the Soil Macro nutrients like N,P,K are classified based on the available attributes. J48 (C4.5), Naivebayes, and BayesNet classifiers have been generated. The Accuracy and Mean absolute Error Rates are compared. The J48 classifier shows the best results compared with other classifiers. The classifier results shows that the Nitrogen availability is low, Phosphorus availability is high to medium and Potassium is medium to high at all most all the areas of Salem District soil. The farmers can use the fertilizers based on the NPK level for their fields.

5. ACKNOWLEDGMENT

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

6. 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. 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
3. 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
4. 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.
5. 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
6. V.Rajeswari, K.Arunesh,"Analysis Soil data using Data mining Classification Technique ", "Indian Journal of science and Technology, vol9(19), May 2016. ISSN (print):0974-6846.
7. Hemageetha,, Dr. G.M. Nasira , "Analysis of the Soil data Using Classification Techniques for Agricultural Purpose "International Journal of Computer Sciences and Engineering IJCSE E-ISSN: 2347-2693 Vol-4 Issue -6 June 2016
8. R.Santhi et at (2014) ,GIS based Soil map for salem district of Tamilnadu. TechInical Folder, TNAU,Coimbatore.
9. Soil Testing Kit Hand Book
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. Vrushali Bhyar, "comparative Analysis of Classification Techniques on soil data to predict Fertility rate for Arangabad District", International Journal of Emerging Trends and Technology in computer science , Vol 3, Issue 3 , March 2014