DATA MINING TECHNIQUE TO ANALYZE SOIL NUTRIENTS BASED ON HYBRID CLASSIFICATION

*E.Manjula and S.Djodiltachoumy
Pachaiyappa's College, Chennai, Tamil Nadu India

Abstract: Data mining methods are greatly admired in the research field of agriculture. The agriculture factors weather, rain, soil, pesticides and fertilizers are the main responsible aspect to raise the production of yields. The fundamental basic key aspect of agriculture is Soil for crop growing. Examination of soil is a noteworthy part of soil asset management in horticulture. The soil investigation is exceptionally useful for cultivators to discover which sort of harvests to be developed in a specific soil condition. The main target of this work is to investigate soil supplements utilizing data mining classification techniques. A large data set of soil nutrients status database was collected from the Department of Agriculture, Cooperation and Farmers Welfare. The database contains measurement of soil nutrients for all different states. This work takes some district of Tamil Nadu in India to analyze the soil nutrients. Distinctive sort's soil has diverse variety of supplements. This paper chooses Nitrogen, Phosphorus, Potassium, Calcium, Magnesium, Sulfur, Iron, Zinc, and so forth, nutrients for investigating the soil supplements utilizing Naïve Bayes, Decision Tree and Hybrid approach of Naïve Bayes and Decision Tree. The performance of the classification algorithms are compared based on the following two factors: accuracy and execution time.

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

I. INTRODUCTION

Horticulture is for the most part reliant on the soil quality yet over the long haul increasingly agrarian generation brings about the loss of supplements present in the soil. It require recognizing methods that will back off this disposal of supplements and furthermore will restore the required supplements with the soil, so continue getting high caliber and great amount of crop productions [1]. In horticulture, great soil wellbeing implies capacity of soil to forces physical, compound and organic exercises for reliable profitability of high product yield. Great nature of soil guarantees us for maintenance and arrival of water and supplements, upgrade and steady root development while keeping up biotic condition, giving the normal outcome and opposes foulness [2].

Soil is very essential for the plant life. It is composed of solids (minerals and organic matters), liquids (water and dissolved substances), and gases (mostly oxygen and carbon dioxide) and contains living organisms. All these elements provide its physical and chemical properties. Managing the soil properly is necessary in order to preserve its fertility, obtain better yield and respect the environment. Testing the soil on the other hand is a must in order to manage it properly.

A soil test is the investigation of a soil example to discover supplement substance, composition and different attributes. Soil tests are normally performed to decide richness and demonstrate insufficiencies that should be cured [3]. Soil nutrients examination is valuable for agriculturist to figure out which kind of yields to be developed in a specific soil condition. In this work data mining classification methods are utilized to investigation the soil nutrients

Data mining is a process of extracting information from a data set and converts it into an understandable structure for further use. Different data investigation methods are accessible for farming exploration field. Classification is one of the data mining techniques that automatically create a model of classes from a set of records that contains class labels. Popular classification techniques include decision trees, neural networks, k-nearest neighbor, SVM, and Naïve Bayesian classifier etc.

Soil has three types of characteristics: physical, chemical, and biological. This paper mainly focuses on chemical characteristics. Chemical properties of soil involve the management of soil nutrients at the most basic level. Mineral nutrients within the soil have either a positive or a negative charge. The nutrients with a positive charge (including calcium, magnesium, potassium, ammonium, and sodium) are held by negatively charged clay particles and organic matter.

The soil supplies the following essential nutrients for proper plant growth and production. These nutrients can originate from weathered minerals or from decomposing organic matter. The nutrients are classified into three forms Primary, Secondary and Micro. The Nutrients and their plant-available forms are

Primary Nutrients: Nitrate, Ammonium Phosphorus and Potassium

Secondary Nutrients: Calcium, Magnesium and Sulfur

Micronutrients: Iron, Manganese, Boron, Molybdenum, Copper, Zinc, Chlorine, Nickel and Cobalt

The primary nutrients are used in the greatest amount by the plant. Analyzing nutrients is one of main topic in agriculture field. It is mostly helpful for agriculturists to figure out which sort of yields to be developed in a specific soil condition. This paper focuses on classification of the soil nutrients analysis based on the selected nutrients for some district (Ariyalur, Coimbatore, Karur, Salem, Thanjavur and Trichy) in Tamil Nadu.

II. RELATED WORK

Baskar et al.,[4] examination the soil data using distinctive algorithms and forecasting method. A report is exhibited using different classification algorithms i.e. Naïve Bayes, J48 (C4.5), JRip with the assistance of data mining tool WEKA.

Jay Gholap [5] predicts soil richness using decision tree algorithm. In [6], the author forecasted soil characteristics and examined soil data using classification techniques. Soil properties, for example, pH value, Electrical Conductivity (EC), Potassium, Iron, Copper, etc. were classified using classification algorithms like Naïve Bayes, J48 and JRip.

Suman et al., [7] use KNN, Naïve Bayes and J48 for analyzing soil data. This research suggests the fertilizer based on the level of nutrients found in the soil test set.

Bhuyar [2] focus on classification of soil richness rate using J48, Naïve Bayes, and Random forest algorithm. J48 algorithm gives preferable outcome over other algorithms. Decision tree form by J48 algorithm assists the cultivator and decision makers to identify the soil richness rate and on the premise of nutrients found in the soil specimen, different fertilizers can be suggested.

Hemageetha et al.,[8] investigates whether the Salem district soil is appropriate or non suitable for crops development in view of pH value using data mining classification techniques. The outcome demonstrate that accuracy of J48 is high compared to Naive bayes , Jrip and BayesNet. And it also shows that the major part of the Salem locale soil is appropriate for development of many harvests.

Bhargavi et al.,[9] argued GATree, Fuzzy Classification rules and Fuzzy C-Means algorithm for classifying soil surface in horticulture soil data. Classification based on Fuzzy rules gives much performance than GATree.

In [10], Naïve Bayes, JRip and J48 algorithms are compared. JRip classification algorithm gives improved outcome and it is efficiently classified into maximum number of instances comparing with the other two.

Ramesh et al., [11] explains comparison of different classifiers and the outcome of this research could improve the management and systems of soil users, throughout a large fields that include agriculture, horticulture, environmental and land use management.

Dildarkhan et al.,[12] gives an investigation of the soil data using different classification algorithms and forecasting methods. Soil testing research centers examination the soil and gives the sample data set. It will require a considerable measure of time to characterize the soil datasets manually.

Shivnath et al.,[13] an analysis of soil properties using Back Propagation Network. This research uses the gradient descent algorithm in training the samples. Back Propagation algorithm produces the better result.

This work focuses on a method that uses data mining technique to analyze the soil nutrients based on hybrid algorithm.

III. Soil Nutrients Analysis

Data Mining is crucial to determine the agricultural related facts such as soil fertility, yield prediction and soil nutrients analysis. This section analysis soil nutrients based on hybrid classification algorithm. Fig 1 shows the architecture of proposed work.

A. Overview of Dataset

To start with any data mining problem, initially bring all the data together. The data set contains 13 attributes. The attributes are District, pH, EC, OC, N, P, K, S, Zn, Fe, Cu, Mn, B. Table 1 shows the description of attributes.

B. Methodology

The proposed work starts with preprocessing step. In this step the collected data was preprocessed. Some records have missing attribute values, that records were removed from the dataset.

In the data conversion step, the preprocessed data was converted based on the nutrients values. Table 2 shows the value for three levels of nutrients.

After data conversion, the macro and micro nutrients are split into three types. The type-1 contains pH, EC, OC, N attributes. Type-2 contains P, K, S, Zn. And Type-3 contains Fe, Cu, Mn, B attributes.

Naïve Bayes

In machine learning, an undemanding probabilistic classification technique is Naïve Bayes classification. It depends on the bayes theorem with independence attributes [10]. The class labels are predicted based on the probability. For classification, a small amount of training data is expected to predict the class labels. [14].

Decision Tree (J48):

J48 is an upgraded version of C4.5 decision tree algorithm. A decision tree is really a predictive machine learning model, which decides the dependent variable (i.e, Target value) based on various attributes of the available training data set. The internal nodes of a decision tree denote varied attributes, the connecting branches of various nodes give us the likely values of the attributes and the terminal node states the classification of the dependent variable. J48 Decision Tree Classifier uses two phases: Tree construction and Tree Pruning. [15] J48 examine the normalized information gain that actually the outcomes the splitting the data by choosing an attribute. In order to make the decision, the attribute utmost standardized information gain is used. The minor subsets are returned by the algorithm. The splitting methods stop if a subset belongs to the same class in all the instances. J48 constructs a decision node using the expected values of the class. [10]

Hybrid Algorithm

The hybrid classifier consolidates the Decision Tree Classifiers property to isolate out dependent attributes, and the effective classification by the Naïve Bayes Classifier on independent attributes.

A large data set of soil nutrients status database was extracted from the Department of Agriculture, Cooperation and Farmers Welfare. The database contains measurement of soil nutrients for all different states. This work collects data for selected district of Tamil Nadu. The selected districts are Ariyalur, Coimbatore, Karur, Salem, Thanjavur and Trichy. It contains nearly 3000 records. Table 3 shows the sample dataset.

IV. EXPERIMENTAL RESULT

The collected data set was preprocessed and some records are removed from the data. After preprocessing step, the data set

was converted into Low, Medium and High based on nutrients level mentioned in Table 2.The converted data set is shown in Table 4.

Apply Naïve Bayes, Decision Tree and Hybrid classification algorithm to classify the soil nutrients as Very High, High, Medium, Low and Very Low.

The number of correctly classified instances and incorrectly classified instances are given in Fig 2, 3 and 4 for (pH, EC, OC, N), (P,K,S,Zn) and (Fe,Cu,Mn,B). The comparative analysis of classifiers is given in Table 5, 6 and 7. Fig 5 shows the execution time of classification algorithms with three types of nutrients. Fig 6 shows the accuracy rate of classification algorithms with three types of nutrients. It is observed from Fig 5 and Fig 6 that the Hybrid algorithm is able to classify the dataset in less time with better classification accuracy rate.

V. CONCLUSION

Data mining techniques in cultivation will help the agriculturist to enhance the crop productivity. The yield production is chiefly based on the soil nutrients. The investigation of soil nutrients present whether which type crop can be refined in specifically soil. This paper proposed an investigation of soil nutrients using distinctive algorithms. The comparative examination of three classification algorithms like Naïve Bayes, Decision and hybrid algorithm was anticipated. Hybrid classification algorithm gives enhanced result for this dataset and is correctly classified into maximum number of instances comparing with the other two. Hybrid algorithm can be suggested to dissect the soil nutrients.
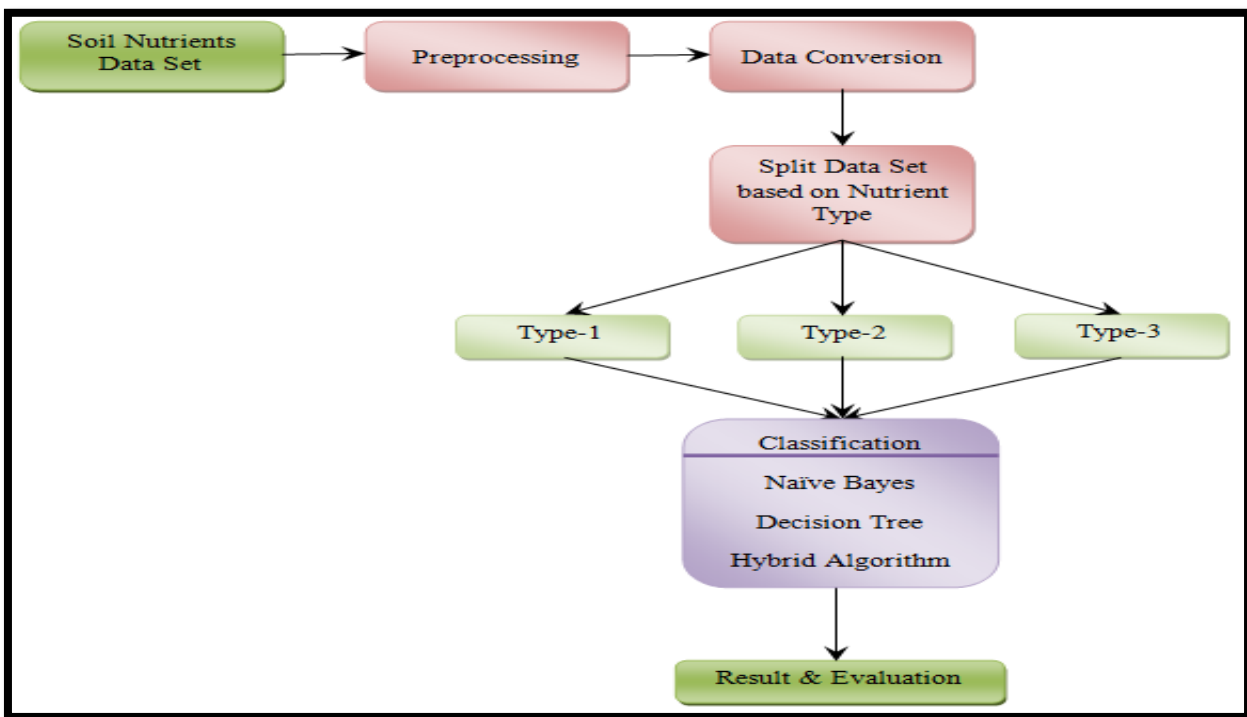
Figures and Tables

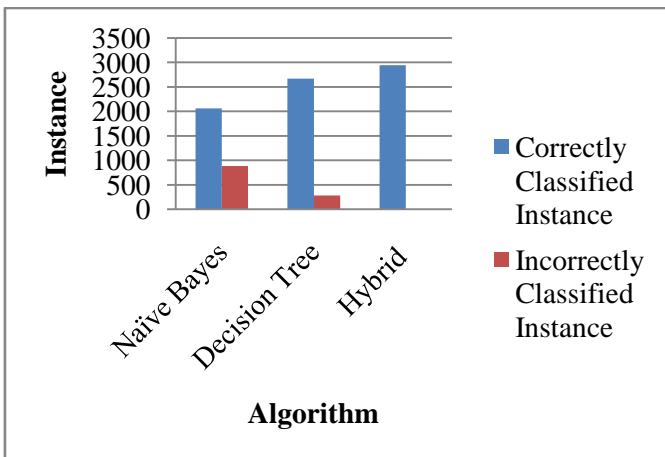Fig 1 Proposed Work Flow

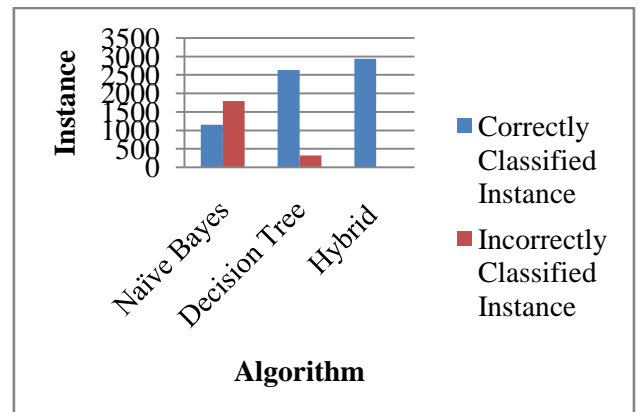Fig 2 Classification result based on pH, EC, OC, N

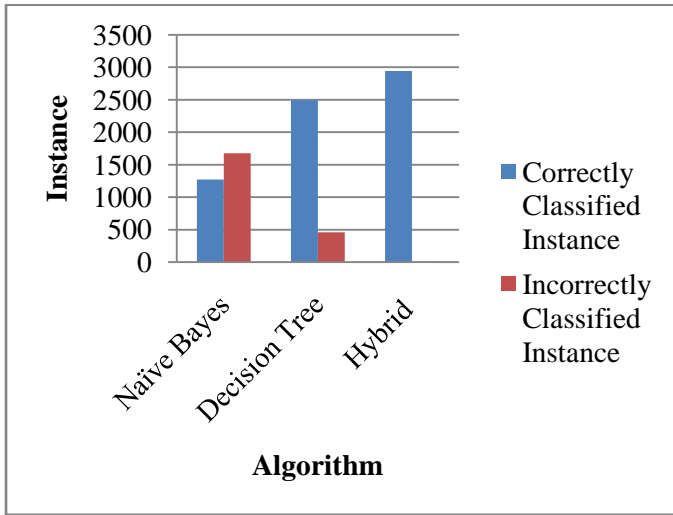Fig 3 Classification result based on P,K,S,Zn
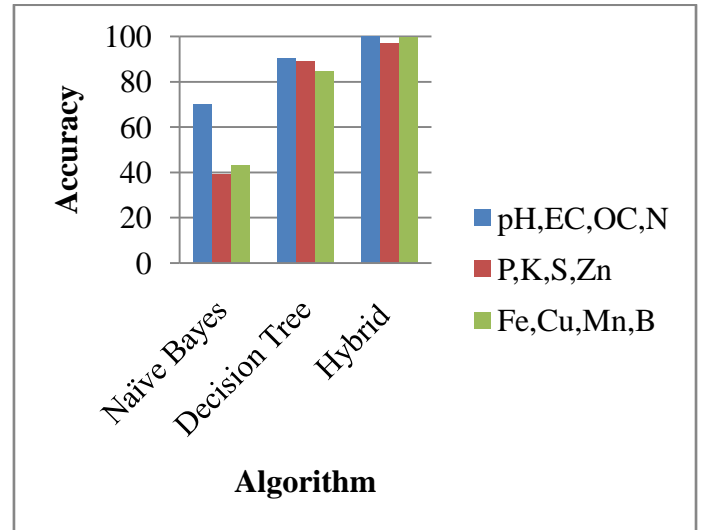
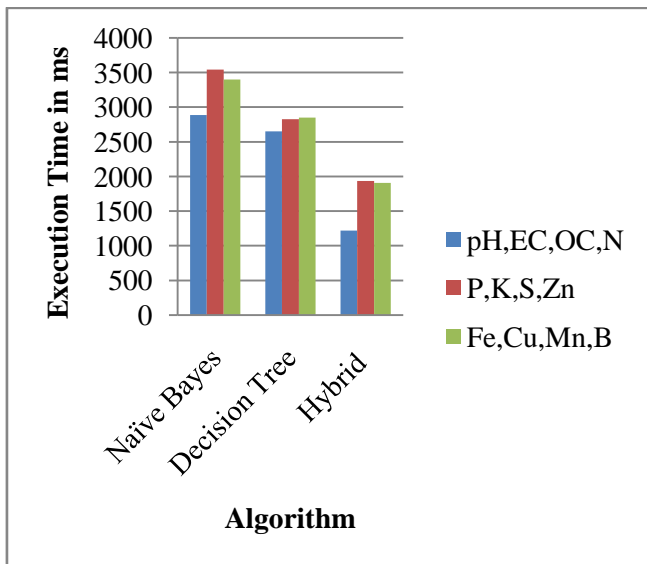Fig 4 Classification result based on Fe,Cu,Mn,B

Fig 6 Accuracy for classifiers

Fig 5 Execution Time of classifiers

Table 1 Soil Nutrients Data set Description

Attribute	Description
District	The data was collected from six district of Tamil Nadu in India.
pH	pH value of Soil Data
EC	Electrical Conductivity
OC	Organic Carbon
N	Nitrogen
P	Phosphorus
K	Potassium
S	Sulfur
Zn	Zinc
Fe	Iron
Cu	Copper
Mn	Manganese
B	Boron

Table 2 Nutrients Levels

Level / Attribute	Low	Deficient	Medium	Moderate	High	Sufficient
pH	< 6.5 (Acidic)	6.5 – 7.5 (Neutral)	> 7.5 (Alkaline)
EC	< 1.0 (Non Saline)	1.0 – 3.0 (Slightly Saline)	> 3.0 (Saline)
OC	< 0.5	0.5 – 0.75	> 0.75
N	< 280	280 – 450	450
P	< 11	11 – 22	> 22
K	< 118	118 – 280	> 280
S	< 10	10-15	> 15
Zn	< 1.2	1.2 – 1.8	> 1.8
Fe	< 3.7	3.7 – 8.0	> 8.0
Cu	< 1.2	1.2 – 1.8	> 1.8
Mn	< 2.0	2.0 – 4.0	> 4.0
B	< 0.46	0.46- 1.0	> 1.0

Table 3 Sample Nutrient Data set

District	pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B
Ariyalur	8.30	0.26	0.54	43.2	17.6	789	10.5	2.45	2.45	3.56	3.67	3.4
Ariyalur	5.10	0.17	0.40	60.60	16.25	60	17.50	0.53	5.23	0.17	3.38	1.0
Coimbatore	6.50	0.38	0.18	231.0	18.0	400	8.10	0.80	5.10	2.20	2.60	0.30
Coimbatore	7.80	0.18	0.18	173.0	13.6	284	17.24	1.0	5.30	1.40	2.60	0.41
Karur	7.90	0.6	0.35	80.00	2.70	189	8.10	0.84	1.40	0.02	1.08	1.10
Karur	8.10	1.50	0.15	45.00	2.70	233	0.11	0.10	2.60	0.03	1.70	1.10
Salem	7.90	0.10	0.28	162	7.50	315	29.2	0.23	10.5	0.72	7.23	2.30
Salem	7.80	0.10	0.11	189	10	378	33.20	0.32	13.5	0.74	0.32	1.0
Thanjavur	6.90	0.15	0.75	205	75	250	25	1.30	6.80	1.80	2.0	1.0
Thanjavur	7.70	0.17	0.69	189	75	300	20.60	1.60	7.40	0.90	2.0	1.0
Trichy	7.0	1.10	0.25	179	53.0	195.0	12.20	1.06	5.72	0.96	3.48	0.10
Trichy	6.80	0.40	0.19	147	25.0	185.0	12.60	0.95	6.27	0.91	5.76	0.10

Table 4 Converted Data set

District	pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B
Ariyalur	H	L	M	L	M	H	M	H	L	H	M	H
Ariyalur	L	L	L	L	M	L	H	L	M	L	M	M
Coimbatore	L	L	L	L	M	H	L	L	M	H	M	L
Coimbatore	H	L	L	L	M	H	H	L	M	M	M	L
Karur	H	L	L	L	L	M	L	L	L	L	L	H
Karur	H	M	L	L	L	M	L	L	L	L	L	H
Salem	H	L	L	L	L	H	H	L	H	L	H	H
Salem	H	L	L	L	L	H	H	L	H	L	L	M
Thanjavur	M	L	H	L	H	M	H	L	M	H	M	M
Thanjavur	H	L	M	L	H	H	H	M	M	L	M	M
Trichy	M	M	L	L	H	M	M	L	M	L	M	L
Trichy	M	L	L	L	H	M	M	L	M	L	H	L

Table 5 Comparative analysis of classifiers based on pH, EC, OC, N

Algorithm	Correctly Classified Instance	Incorrectly Classified Instance	Accuracy	Execution Time in ms(milli sec)
Naïve Bayes	2063	885	69.9 %	2886
Decision Tree	2666	282	90.43%	2651
Hybrid	2946	2	99.93%	1221

VI. REFERENCES

[1] Peter Gruhn, Francesco Goletti, and Montague Yudelman (2000), " Integrated Nutrient Management, Soil Fertility, and Sustainable Agriculture: Current Issues and Future Challenges" Food, Agriculture, and the Environment Discussion Paper 32,2000

[2] Bhuyar V. (2014), "Comparative analysis of classification techniques on soil data to predict fertility rate for Aurangabad District", International Journal of Emerging Trends and Technology in Computer Science. 2014 Mar-Apr; 3(2):200–3.

[3] "Soil test", Wikipedia, February 2012

[4] S.S.Baskar L.Arockiam S.Charles (2013), "Applying Data Mining Techniques on Soil Fertility Prediction", International Journal of Computer Applications Technology and Research Volume 2–Issue 6, 660 -662, 2013

[5] Jay Gholap (2012), "Performance Tuning of J48 Algorithm for Soil Fertility" ,2012. Asian Journal of Computer Science and Information Technology 2: 8 (2012) 251–252

[6] Gholap J, Lngole A, Gohil J, Shailesh, Attar V.(2012), "Soil data analysis using classification techniques and soil attribute prediction", 2012 Jun; 9(3):1–4

[7] Suman, Bharat Bhushan Naib (2013), "Soil Classification and Fertilizer Recommendation using WEKA" ,IJCSMS International Journal of Computer Science & Management Studies, Vol. 13, Issue 05, July 2013

[8] Hemageetha, G.M. Nasira (2016), "Analysis of Soil condition Based on pH value Using Classification Techniques", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 18, Issue 6, Ver. III (Nov.-Dec. 2016), PP 50-54

[9] Bhargavi P. & Jyothi S. (2011), "Soil Classification Using Data Mining Techniques: A Comparative Study", International Journal of Engineering Trends and Technology

[10] Rajeswari V, Arunesh K. (2016), "Analysing soil data using data mining classification techniques", Indian Journal of Science and Technology. 2016 May; 9(19).

[11] V.Ramesh and K. Ramar,(2011) "Classification of Agricultural Land Soils: A data Mining approach",Volume:6, Issue: 3,Sep. 2011

[12] Dildarkhan T. Pathan, Pushkar D. Joshi, Prof. S. U. Balvir,(2014), "Prediction of soil Quality for Agriculture", IRJSSE International Research Journals of sustainable Science & engineering, Vol. 2, Issues 3, 2014.

[13] Shivnath Ghosh, santanu koley,(2014), "Machine Learning for Soil fertility and Plant Nutrients Management using Back Propagation Neural Networks", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 2, Issues 2, 2014.

[14] Narain B. (2011), "Study for Data Mining techniques in classification of agricultural land soils", Journal of Advanced Research in Computer Engineering. 2011 Jan-Jun; 5(1):35–7.

[15] Ponmani S, Vidhu Priya P (2017), "Classification Algorithms in Data Mining –A Survey", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 1, January 2017