Classification of Soil Based on Fuzzy Logics

Fathima Rubi M.H[1], Dr. Sharmila kumari[2]

M. Tech Student, Department of Computer Science & Engineering, P.A College of Engineering, Mangalore, India[1]

HOD, Department of Computer Science & Engineering, P.A College of Engineering, Mangalore, India[2]

Abstract: The quick height in innovation has made information in plenty which has both required just as undesirable. So as to truncate the undesirable information and order, plentiful of data mining methods has been actualized. The primary objective of information mining is to gather, pre-process and order the information dependent on specific highlights of soil, for ex, PH, Nitrogen, Oxygen and Carbon percent of soil and foresee the harvests developed for the particular soil type. The application provides a deep insight of the behaviour of the soil in a particular region. It provides a feature extraction method to sort the data set according to the users need and thereby determining the necessary values of certain attributes. The decision makers can easily focus on the content they are interested on under certain context in the sub-classifications by grouping the attributes. Information from different nations are assembled and broke down and were determined based predefined Fuzzy Rules which has returned the precise and recall as result at the end of an output.

Keywords: Data Mining, PH, Nitrogen, Carbon, Oxygen, Fuzzy Rule, Precise, Recall

I. INTRODUCTION

Data mining has developed both in commercial as well as research centers. These techniques are useful in various fields like scientific, commercial etc. These mining techniques are used to take out the information out from datasets. The data mining learning is utilized for the point of higher psychological procedure in fluctuated fields. the information mining is relevant to fields like businesses to broaden the benefit by making some future strategies, information handling could be a horrendously vital investigation area in ongoing examination world. The procedures are useful to inspire significant and utile data which might be seen by a few people, So as to truncate the undesirable information and group, plentiful of information mining strategies has been executed[1]. Effective mining techniques have been used to extract information. The accuracy and effectiveness can be calculated using this technique navie bayes technique is used here to calculate the effectiveness of by using the method of maximum likelihood. Primary method is to gather information from soil data[2], pre-process and order the information dependent on specific highlights of soil, for example, PH, Nitrogen, Oxygen and Carbon level of soil and foresee the harvests developed for the particular soil type. Soil arrangement manages the efficient classification of soils dependent on recognizing characteristics just as criteria that direct decisions being used. It impacts numerous different properties and hugeness of land use and the executives[3]. The arrangement standards can be connected to the new information tuples if the precision is viewed as worthy. Information Mining is a significant piece of Knowledge revelation Process that can dissect an enormous set of information and get covered up and helpful learning. Arrangement and bunching is the real procedure in information mining to settle AI process. Information from different nations are assembled and investigated and were determined dependent on pre-characterized Fuzzy Rules which restored the exact result toward the end as a yield[4].

II. RELATED WORK

The work proposed by Awasthi N, Bansal study on "Application Of Data Mining Classification Techniques On Soil Data Using R" Fuzzy algorithm and comparing it with the GA Tree which in turn resulted in 99% accuracy for fuzzy classification rules whereas the GA Tree algorithm doesn't meet the requirements. It has applied C-means fuzzy classifier for soil data, which has several advantages such as simple and systematic structure. In the cluster generation process, a clustering algorithm based on C-means is applied to cluster the soil data such that the training data can be classified by the fuzzy classifier. The properties of attributes in the Soil database are specified with the specified quality standards. Quality has the property of fuzziness. If same type of texture occurs with different composition of attributes of Soil data, ambiguity may appear[1]. The survey made by Anand V. Saurkar proposed on "A Review Paper on Various Data Mining Tech- niques" soil classification with the data mining technologies such as USDA soil taxonomy and the UNESCO system FAO. The survey involved different mining techniques which is relevant to the subject and thus comparing the results of all possible outcomes of those algorithms such as Decision tree, K-means clustering and also includes a study of ANN[2]. Work done by Bhargavi, P. and Dr.Jyothi, S "Soil Classification Using Data Min- ing Techniques: A Comparative Study" on the different techniques of soil classification which involved Classified soil

types such as Red Gravels, Black Cotton, Sandy Alluvial, Saline alkali. The role of the methods for classification is on the improvement of the soil fertility and the nutrient analysis[3]. n one of the early research by Baskar, S, S. Arockiam,L. and Charles, S "Soil Classification Using Data Min- ing Techniques A Comparative Study" used classification techniques called JRip & J48, Naive Bayes which are acted on the data sets of the soil. Skyline shrewd examples were gathered and portrayed for significant physical and concoction following standard strategies[4]. To identify the fertility prediction of the soil Bhuvaneswari, S., Pramananda Perumal, T., Jagadhesan, B. (2016) in their work "An analysis and impact factors on Agriculture field using Data Mining Tech- niques" the team has compared the result of algorithms such as Naive Bayes, j48 and JRip which enriches the cultivation of crops. In dislike the way that the least middle squares relapse is known to create preferable outcomes over the established straight relapse system, from the given arrangement of characteristics[5]. On AI methods for prompting space models or examining datasets has utilizing there to actualize application. the work is done by Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E. D., & Goldschmitt, M in "Digital soil mapping using artificial neural networks. Mining of data from existing agrarian datasets delivered and to perform fundamental research in information mining by growing new AI calculations. The objective may be to utilize a model gainfully, to give mechanized grouping of new examples[6].Emotional traits for mushroom reviewing may not be helpful practically speaking. Freund Y. and Schapire R. in his work "A Short Introduction to Boosting", a comparison of two data mining techniques of Artificial Neural Network and Support Vector Machine is presented. This study is done with the help of data analytics tool R. More techniques for the classification and prediction will be done in the future, and will improve the performance of these classifiers with other available methods. A fertility recommendation system in the future is developed. It will need more data to be collected from different parts of India and will prepare a digital map of soil. This will become input of new system and will help in identifying different profiles of soil[8]. Gholap, Jay in his work "Performance Tuning Of J48 Algorithm For Prediction Of Soil Fertility" certain Data Mining procedures were received so as to estimate crop yield examination with existing information. The applications that utilization the K-Means approach, use just the essential calculation, while numerous other improvements are accessible. A few Data Mining procedures have not yet been connected to agrarian issues. For instance, Bi-grouping methods might be employed for finding significant data from rural related arrangements of information. The K-Means calculation can parcel the examples in groups, however no contemplations are made on the intensifies that are in charge of this segment. Bi-grouping can gives this sort of data[9]. In this proposed work "A Review on Data Mining Techniques for Fertilizer Recommendation" by Jignasha, M.,Jethva, Nikhil Gondaliya, Vinita Shah, found the similar examination of two calculations like Naive Bayes and J-48 is carried out. Among the available classification algorithm Naivee Bayes gives better result for the dataset which we were used and accurately ordered into greatest number of examples contrasting and remaining calculations [10].Naïve Bayes Algorithm can be recommended to predict crop growing in particular soil sample.

III.   EXISTING PRACTICAL METHODS

Standard penetration test: Tube is of thick walled sample, which is driven to the ground by borehole by side hammer. Purpose: To identify density and also check the strength, it is very simple and cost is less.
Cone penetration test: Popularly used method. This has a cone which is facing downwards; it has a pressure sensor which will calculate the pressure of underground water
Pressure meter test: Here it calculates the pressure of world horizontally. It is used to test very hard clay
Vean shear test: This is for very sticky soil; rod has 4 rotating blades at the end. Majorly used for the classification of clay.[7]

IV.   PROPOSED SYSTEM AND METHODOLOGY

To classify the soil based data completely by using Fuzzy Classification algorithm to produce rules for the characterized enrolment elements of the in-put properties. We have changed over each preparation information into a fuzzy standard in java language. Hierarchical Fuzzy Classification decomposes multi-dimensional classification space, and thus reduces complexity, By using navie bayes classifier we have calculated the training set value as well as probability values for each classifications. By gathering qualities of a given setting in sub-characterizations, it enables decision makers to concentrate on zone of intrigue. The participation work maps from the arrangement of propositional capacities PF and the universe of U into the arrangement of truth esteems T. In traditional rationale reality esteems are sure. In this manner, a characterization is exact since true esteems are either precisely obvious or precisely wrong. Naive Bayes classifier expect that the nearness or nonappearance of a specific element is random to the appeared or nonattendance of some other element, given the class variable. It tends to be prepared in all respects effectively in a regulated learning. Contingent upon the exact idea of the likelihood model, Naive Bayes classifiers can be prepared in all respects proficiently in a regulated getting the hang of setting. In numerous functional applications, parameter estimation for naive Bayes models utilizes the technique for most extreme probability. Figure given below shows the Collection soil data from an online consultancy services which provides information of the soil data from various

countries in the world. and then Generating Fuzzy rules from preparing information to classify by characterization by de-fining the participation capacities for the info qualities of the information. The Generated Fuzzy grouping rules are connected on the gathered soil information and the accomplished exhibitions were looked.
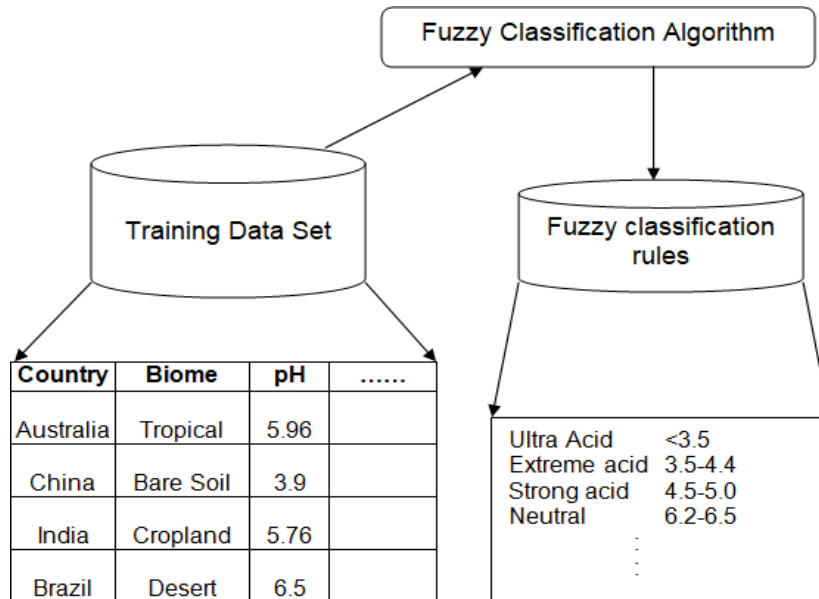

Fig 1. Methodology of Proposed system

V. FUZZY LOGICS

Fuzzy rationale is a type of many-esteemed rationale in which reality estimations of variables might be any genuine number somewhere in the range of o and 1 comprehensive. It is utilized to deal with the idea of incomplete truth, where reality esteem may go between totally evident and totally false. In the given figure below it shows Boolean rationale, reality estimations of factors may just be the whole number qualities o or 1.
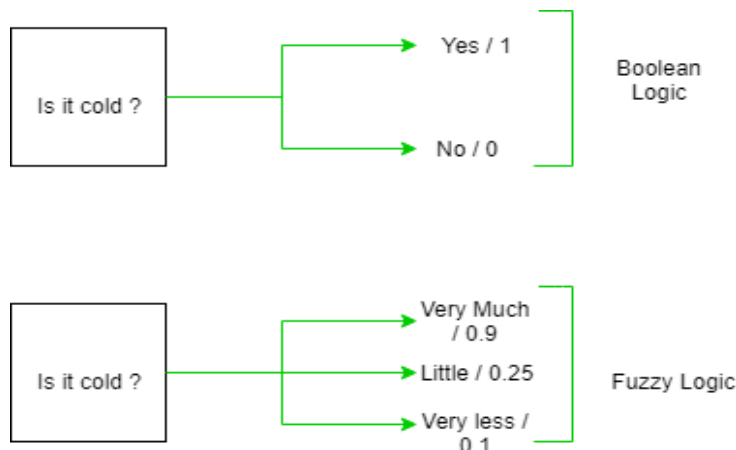

Fig.2 Comparison of Boolean and Fuzzy logic.

A fundamental application may describe different sub-scopes of a nonstop factor. For example, a soil information mining may have a few separate participation capacities characterizing specific pH ranges required for the development of the harvests. Each capacity maps a similar pH incentive to reality esteem in the o to 1 territory. This reality esteems would then be able to be utilized to decide how the soil information can be characterized regarding its qualities.

Table below shows the values that are used to check the acidity of the soil. This includes many intermediate values accordingly based on its concentration. These values are based on the pH value of the soil sample. The acidic concentration is high when the pH is less than 3.5 and further the pH is very high then the soil is considered to be alkaline. These rules are called pH rules and are accommodated to find the percentages of the minerals such as Nitro-gen, Phosphorous & Carbon. The process of the algorithm is to initially fuzzily all info esteems into fuzzy participation

capacities. At that point execute every single material guideline in the standard base to register the fuzzy yield capacities. De-fuzzily the fuzzy yield capacities to get "fresh" yield esteem.

Ultra acidic	<3.5
Extreme Strong acidic	3.5-4.4
Very Strong Acid	4.5 - 5.0
Strong acidic	5.1-5.5
Moderate Acidic	5.6-6.0
Slight Acidic	6.1-6.5
Neutral	6.6-7.3
Slight alkaline	7.4-7.8
Moderate alkaline	7.9-8.4
Strong alkaline	8.4-9.0
Very alkaline	>9.1

Table 1. PH rules defined in the system

VI. METHODOLOGY AND MODULES DESCRIPTION

To classify the soil based data completely by using Fuzzy Classification algorithm to produce rules for the characterized enrolment elements of the in-put properties. We have changed over each preparation information into a fuzzy standard in java language. Hierarchical Fuzzy Classification decomposes multi-dimensional classification space, and thus reduces complexity, By using naive bayes classifier we have calculated the training set value as well as probability values for each classifications. By gathering qualities of a given setting in sub-characterizations, it enables decision makers to concentrate on zone of intrigue. The participation work maps from the arrangement of propositional capacities PF and the universe of U into the arrangement of truth esteems T. In traditional rationale reality esteems are sure. In this manner, a characterization is exact since true esteems are either precisely obvious or precisely wrong.

Figure shown below shows the complete data set of the soil contains the information of the soil attributes such as pH, biome, minerals etc. are loaded into the database. The loaded soil data is pre-processed and retrieved the useful information from the dataset. The pre-processed data is subjected to the fuzzy rule sets that are already defined. Extraction of the data is done according to the certain attributes and constraints. This is called as feature extraction. The fuzzy rules are specified for the input attributes of the extracted data. The fuzzy rules in this system specify the amount of acidic content and the organic level of certain minerals which are compared with the existing one. The data satisfying the constraints are given to classify the soil. This might include separating the types of lands based on the soil type, organ- ic level of minerals or any other constraints. The prediction of the soil is done with respective to the cultivation, vegetation and pH rules. Based on this, the precision and recall percentage is calculated according to the predicted values and checked for its accuracy. The functions and computations that are undergone by the system during the process of soil classification. The activities include collecting the data, computing the relevance of those data, computing similarity, finding the representatives. The data is sorted out according to their respective attributes by computing its similarity. The attributes that gives similar values and redundant data is eliminated by finding similarity. Only the required attributes are selected for the classification of the dataset. These processes come under the feature extraction process. Further the data is represented with the attributes and these attributes are clustered according to the user needs. To classify the soil data and its attributes we are using fuzzy approach which generates rules for the characterized enrolment elements of the information traits and we have changed over each preparation information into a fuzzy standard and this is shown as the outcome. The system is divided into 5 modules for a superior comprehension of the working of the framework and simple usage of the equivalent. The modules are recorded as underneath.

Soil data Collection: Collecting different kind of soils with their specification is precisely blend of mineral and natural materials in addition to air and water. The substance of soil shift by area and are always showing signs of change.

Input can be gathering as text file or excel file format to train the further progress.

The entire data updating soil values has been discovering in the basis of machine learning techniques further moving on for classification work.

Input data pre–processing: Pre-process is the process of training the dataset for the purpose of interrupting into the core work. Training the data will get the process of removing the missing content, ir-relevant information from the valuable data. Accordingly, fuzzy order is the way toward gathering people having similar attributes into a fuzzy set. Fuzzy rationale has been reached out to deal with the idea of incomplete truth, where reality esteem may extend between totally evident and totally false.

Performance evaluation: This modules enables users get to know graphically the value of entire progress. By showing the mathematical evaluation for the entire classification work by concluding the better performances. Reducing time on current working instances will show the integrity of project.
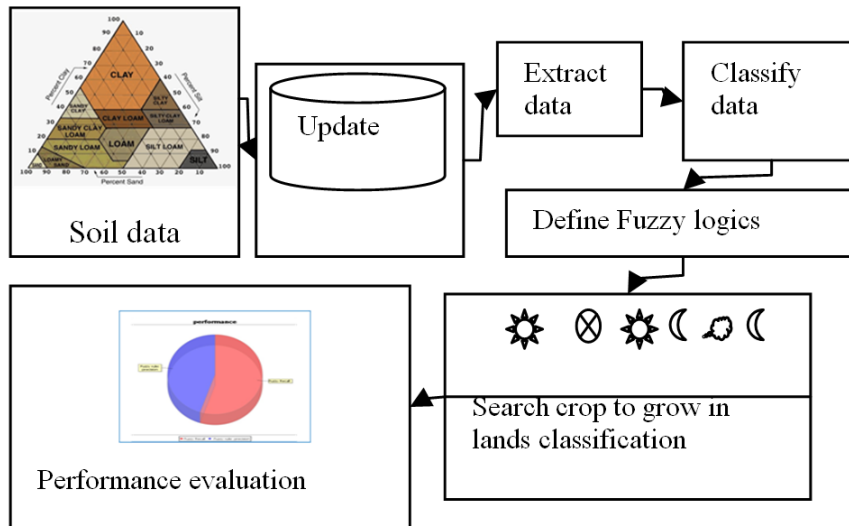

Fig 3. Flow of processes for the classification & predictions of the soil

VII.   OUTCOMES AND CONCLUSION

The graph is plotted based on the precision and recall method where the percentage of previously estimated features such as carbon, nitrogen and phosphorous are considered. The percentage is expressed as the ratio of the total number of the retrieved records that are relevant to the context to the number of relevant records in the database. From the study, it is seen that, recall value is less if there is no much relevant record to be retrieved from the database. For different data sets, different precision and recall values are obtained in the form of percentages. In our project, the present data set has the precision value of 7% and recall value as 75% which is highly accurate. The performance of implemented fuzzy classification algorithm is screenshot below.
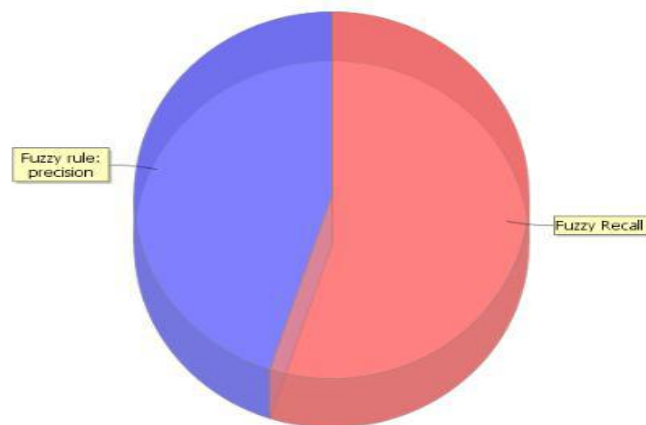

Fig 4. Results: precision and recall.

Modules had introduced another strategy to create emotional fuzzy principles from preparing information to manage the soil grouping issue. Initially, we convert the preparation information to fuzzy standards, and after that, we combine those

fuzzy principles so as to diminish the number of fuzzy guidelines. At that point, we compute the heaviness of each info variable showing up in the produced fuzzy standards by the connections of information factors. Weighted fuzzy principles strategy gets a higher normal characterization exactness rate than the ones. This is benevolent for the researchers to help them estimate the type of soil required for cultivation of crops on the different varieties of land.

ACKNOWLEDGEMENT

REFERENCES

[1]. Awasthi Nikhita, Bansal Abhay, "Application Of Data Mining Classification Techniques On Soil Data Using R" International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835, 2017

[2]. Anand V. Saurkar, "A Review Paper on Various Data Mining Tech- niques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4,Pp 98-1o1, 2014

[3]. Bhargavi, P. and Dr.Jyothi, S."Soil Classification Using Data Min- ing Techniques: A Comparative Study", International Journal of Engineering Trends and Technology, July to Aug Issue .2011

[4]. Baskar, S, S. Arockiam,L. and Charles, S. "Applying Data Mining Techniques on Soil Fertility Prediction", International Journal of Computer Ap- plications Technology and Research, Volume 2– Issue 6, 66o - 662, 2o13,ISSN: 2319–8656 2013

[5]. Bhuvaneswari, S., Pramananda Perumal, T., Jagadhesan, B. "An analysis and impact factors on Agriculture field using Data Mining Tech- niques" International Journal of Business Intelligents Volume: o5 Issue Page No.41-44 ISSN: 2278-24oo, 2016

[6]. Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E. D., & Goldschmitt, M. "Digital soil mapping using artificial neural networks. " Journal of plant nutrition and soil science, 168(1),21-33. 2005

[7]. Chandan, Ritula Thakur , "Recent Trends Of Machine Learning In Soil Classification: A Review" International Journal of Computational Engineering Research (IJCER) ISSN (e): 225o – 3oo5 || Volume, o8 || Issue, 9|| September – 2018 ||

[8]. Freund Y. and Schapire R., "A Short Introduction toBoosting",Journal of Japanese Society for Artificial Intelligence, 14(5):771-78o 1999

[9]. Gholap,Jay "Performance Tuning Of J48 Algorithm For Prediction Of Soil Fertility" Asian Journal of Computer Science and Information Technology ISSN 2249-5126.

[10]. Jignasha,M.,Jethva,Nikhil Gondaliya, Vinita Shah,"A Review on Data Mining Techniques for Fertilizer Recommendation" International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2o18 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-33o7-2o18

[11]. Khan, Huma, Navaz, Shahista Dr Ghosh., S., M. "A Survey on Various Data Mining Techniques in Field of Agriculture for Prediction of Crop Yield" In- ternational Journal of Science and Research (IJSR) ISSN (Online): 2319- 7o64.

[12]. M. Kovacevic, B. Bajat, B. Gajic "Soil Type Classification and Es- timation of Soil Properties using Support Vector Machines", Geoderma 154(3– 4), 34o–347,-2010.

[13]. Madhuri Kommineni, Someswari Perla, Divya Bharathi Yedla "A Survey of using Data Mining Techniques for Soil Fertility", International Journal of Engineering & Technology, 7 (2.7) -2o18,917-918

[14]. Murugesh Kumar, B. , Dr.Ananda Kumar, K and Dr.Bharathi "A survey on soil classification methods using data mining techniques" , Inter- national Journal of Current Trends in Engineering & Research (IJCTER) e- ISSN 2455–1392 Volume 2 Issue 7, July 2016

[15]. Neha Sharma, Damayanthi Sharma "Classlification & Prediction based Data Mining Techniques" Volume 4, Issue 11, November 2o16 International Journal of Advance Research in Computer Science and Management Studies ISSN: 2321-7782-2016 (Online)

[16]. Palepu,Ramesh Babu, Rajesh Reddy Muley "An Analysis of Agricultural Soils by using Data Mining Techniques" Volume 7 Issue No.1o -2017

[17]. Pooja.M, Sangeetha, Shreyaswi Salian,J, Veena Kamath, Mithun Naik "Implementation of Crop Yield Forecasting using Data Mining" International Research Journal of Engineering & Technology (IRJET) e-ISSN: 2395-oo56 Vol: o5 Issue: o4 | Apr-2o18 p-ISSN: 2395-oo72-2018

[18]. Rajeshwari, V. and Arunesh, K "Analyzing Soil Data using Data Mining Classification Techniques", Indian Journal of Science and Technology, Vol 9(19), DOI: 10.17485/ijst/2016/v9i19/93873, May 2016

[19]. Ramesh,D., Vishnu Vardhan,b "Data Mining Techniques and Ap- plications to Agricultural Yield Data" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013, ISSN (Print) : 2319-594o, ISSN (Online) : 2278-1o21

[20]. Samundeeswari,K., Dr.Srinivasan.,K "Data Mining Techniques In Agriculture Prediction Of Soil Fertility" International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2o17 ISSN 2229-5518-2017

[21]. Solanki,J, Mulge,.Y, "Different techniques used in data mining in agricul- ture", International Journal of Advanced Research in Computer Science and Software Engineering. 2015 May; 5(5):1223–45

[22]. Vamanan,R.,Ramar,K., "Classification Of Agricultural Land Soils A Data Mining Approach"; International Journal on Computer Science and Engineering (IJCSE); ISSN: o975-3397 Vol. 3- 2011

[23]. Velide Phani kumar and Lakshmi Velide, "Data mining plays a key role in soil data analysis of Warangal region", International Journal of Scientific and Research Publications, Volume 4, Issue 3, March 2014