

## International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X



*IJCSMC, Vol. 4, Issue. 7, July 2015, pg.330 – 333*

### **RESEARCH ARTICLE**

# Data Mining Technique to Predict the Accuracy of the Soil Fertility

**Dr. S.Hari Ganesh**

Assistant Professor, Department of Computer Science, Bishop Heber College (Autonomous), Tiruchirapalli, India  
[hariGANESH17@gmail.com](mailto:hariGANESH17@gmail.com)

**Mrs. Jayasudha**

Mphil. Scholar, Department of Computer Science, Bishop Heber College (Autonomous), Tiruchirapalli, India  
[Jsudha15@gmail.com](mailto:Jsudha15@gmail.com)

*Abstract-- The techniques of data mining are very popular in the area of agriculture. Today, data mining is used in a vast areas and many off-the-shelf data mining system products and domain specific data mining application software's are available, but data mining in agricultural soil datasets is a relatively a young research field. In this paper, we provide web base solution for the soil testing laboratories as well as free messages for the farmer which contains information like soil testing code, fertilizer which is necessary for the crop and also the expert advice. Also farmers specify there next crop while they give their sample to scantiest so according to next crop the fertilizer will suggest. The result is based on the classification of contains which must be present in soil and according to result report will be generated.*

**Keywords:** Data Mining, Soil Testing, classification.

## I. INTRODUCTION

Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). A soil test is the analysis of a soil sample to determine nutrient content, composition and other characteristics. Tests are usually performed to measure fertility and indicate deficiencies that need to be remedied. The soil testing laboratories are provided with suitable technical literature on various aspects of soil testing, including testing methods and formulations of fertilizer recommendations.

## II. LITERATURE SURVEY

The outcome of this research will result into substantial diminution in the price of these tests, which will save a lot of efforts and time of Indian soil testing laboratories.

A data mining approach” In this research paper V. Ramesh and K. Ram explains comparison of different classifiers and the outcome of this research could improve the management and systems of soil uses throughout a large fields that include agriculture, horticulture, environmental and land use management.[4]

From “Generalized software tools for crop area estimation and yield forecast” Roberto Benedetti and others describes the procedure that leads to the estimates of the variables of interest, such as land use and crop yield and other sampling standard deviation, is rather tedious and complex, till to make necessary for statistian to have a stable and generalized computational system available. The SAS is also often the ideal instrument to face with these needs, because it permits the handling of data effectively and provides all necessary functions to manage easily surveys with thousands of micro data. This paper focus on the use of this system in different steps of the survey: sample design, data editing and estimation. The information produced is however available for one user only, the manager of the survey [5].

A study of crop yield distribution and crop insurance” by Narsi Reddy Gayam in his research study examines the assumption of normality of crop yields using data collected from INDIA involving sugarcane and Soybean. The null hypothesis (Crop yield are normally distributed) was tested using the Lilliefore method combined with intensive qualitative analysis of the data. Result show that in all cases considered in this thesis, crop yield are not normally distributed.[5]

Dr. Bharat Misra, et al., [6] observed the research studies on application of data mining techniques in the field of agriculture. Some of the techniques, such as ID3 algorithms, the k-means, the k nearest neighbor, artificial neural networks and support vector machines applied in the field of agriculture were presented. Data mining in application in agriculture is a relatively new approach for forecasting / predicting of agricultural crop/animal management. This article explores the applications of data mining techniques in the field of agriculture and allied sciences. Historical crop yield information is important for supply chain operation of companies engaged in industries that use agricultural produce as raw material. Livestock, food, animal feed, chemical, poultry, fertilizer pesticides, seed, paper and many other industries use agricultural products as intergradient in their production processes. An accurate estimate of crop size and risk helps these companies in planning supply chain decision like production scheduling. Business such as seed, fertilizer, agrochemical and agricultural machinery industries plan production and marketing activities based on crop production estimates.

yashovardhankelkar, et al.,[7] surveyed and says that data selection is the data relevant to the analysis is decided and retrieved from the various data locations. Data preprocessing is the process of data cleaning and data integration is done. Data cleaning is also known as data cleansing; in this phase noise data and irrelevant data are removed from the collected data. Data integration is multiple data sources, often heterogeneous, are combined in a common source. In Data transformation the selected data is transformed into forms appropriate for the mining procedure. Data Mining is the crucial step in which clever techniques are applied to extract potentially useful patterns. The decision is made about the data mining technique to be used. Interpretation and Evaluation is interesting patterns representing knowledge are identified based on given measures. The discovered knowledge is visually presented to the user. [8]

## III. COMPARATIVE STUDY OF SOIL CLASSIFICATION

Soil classification was measured serious to study due to depending upon the fertility class of the soil domain knowledge experts determines which crops should be taken on that particular soil and which fertilizers should be used for the same. The following section describes Naive Bayes, J48, JRip algorithms briefly.

### A. Naive Bayes

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes theorem with strong (naive) independence assumptions. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. An advantage of the naive Bayes classifier is that it

only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.

Table 2: Contrast of classifiers

| Classifier                       | NB Tree | JRip   |
|----------------------------------|---------|--------|
| Correctly Classified Instances   | 822     | 1898   |
| Incorrectly Classified Instances | 1265    | 201    |
| Accuracy (%)                     | 38.65%  | 90.24% |
| Mean Absolute Error              | 0.321   | 0.0423 |

**B. JRip**

This algorithm implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by William W. Cohen as an optimized version of IREP. In this paper, three classification techniques (naïve Bayes, J48 (C4.5) and JRip) in data mining were evaluated and compared on basis of time, accuracy, Error Rate, True Positive Rate and False Positive Rate. Tenfold cross-validation was used in the experiment. Our studies showed that J48 (C4.5) model turned out to be the best classifier for soil samples.

**IV. PREDICTION OF UNTESTED ATTRIBUTES**

Algorithms similar to Linear Regression, Least Median Square, Simple Regression different attributes were predicted. Due to these outcome the values of Phosphorous attribute was found to be most truthfully predicted and it depends on slightest number of attributes. When all attributes are numeric, linear regression is a natural and simple technique to consider for numeric prediction, but it suffers from disadvantage of linearity. If data exhibits non-linear dependency, it may not give good results. In this case, least median square technique is used. Median regression techniques incur high computational cost which often makes them infeasible for practical problems. Several regression tests were carried out using WEKA data mining tool to predict untested numeric attributes. Linear-Regression test for predicting phosphor gave the best and accurate results. These predictions can be used to find out phosphor content without taking traditional chemical tests in soil testing labs, and this will eventually save a lot of time. Statistical results of these tests are given in Table3. There was very limited variations amongst the predicted values of phosphor attribute. Though the Least Median of Squares algorithms is known to produce better results, we observed that the accuracy of linear regression was relatively equivalent to that of least median of squares algorithm.

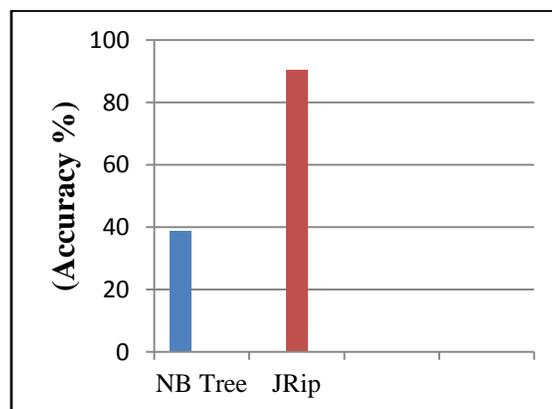


Figure 1: Comparison among NB, JRip and J48 algorithms

The observation of above reading shows the Relative Absolute Error is nearly identical for both the prediction algorithm. Also Least Median Square regression gives better numeric predictions but the time taken to build the model is 67 times that of Linear Regression, hence computational cost used by Linear Regression is much lower than that of least Median Square.

## V. CONCLUSION

In this paper, we have suggested an analysis of the soil data using different algorithms and prediction technique. In spite the fact that the least median squares regression is known to produce better results than the classical linear regression technique, from the given set of attributes, the most accurately predicted attribute was “P” (Phosphorous content of the soil) and which was determined using the Linear Regression technique in lesser time as compared to Least Median Squares Regression. In this paper we have demonstrated a comparative study of various classification algorithms i.e. Naïve Bayes, J48 (C4.5), JRip with the help of data mining tool WEKA. J48 is very simple classifier to make a decision tree, but it gave the best result in the experiment. In future, we contrive to build Fertilizer Recommendation System which can be utilized effectively by the Soil Testing Laboratories. As per the soil sample given to lab for testing and cropping pattern the system will recommend suitable fertilizer.

## References

- [1] Cunningham, S. J., and Holmes, G. (1999). Developing innovative applications in agriculture using data mining. In the Proceedings of the Southeast Asia Regional Computer Confederation Conference, 1999.
- [2] A Thesis by D.Basavaraju, Characterisation and classification of soils in Chandragiri mandal of Chittoor district, Andhra Pradesh. Jiawei Han and Micheline Kamber, Data Mining concepts and Techniques, Elsevier..
- [3] Minaei-Bidgoli, B., Punch, W. Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System. Genetic and Evolutionary Computation, Part II. 2003. pp.2252–2263. Quinlan, J.R. C4.5: Programs for Machine Learning. Morgan Kaufman. 1993.
- [4] V. Ramesh and K. Ramr Classification of agricultural land soils: A data mining approach” International Journal on Computer Science and Engineering (IJCSSE) ISSN : 0975-3397 Vol. 3 No. 1 Jan 2011 379
- [5] Rainfall variability analysis and its impact on crop productivity Indian agriculture research journal 2002 29,33.,8) SPRS Archives XXXVI-8/W48 Workshop proceedings: Remote sensing support to crop yield forecast and area estimates GENERALIZED SOFTWARE TOOLS FOR CROP AREA ESTIMATES AND YIELD FORECAST by Roberto Benedetti A, Remo Catenaro A, Federica Piersimoni B
- [6] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, “Data mining Techniques for Predicting Crop Productivity – A review article”, International Journal of Computer Science and technology, march 2011.
- [7] D.Rajesh ,International Journal of Computer Applications ,Volume 15, February 2011.
- [8] International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 2, February 2012) 275 “Survey on Data Mining “ VibhaMaduskar and Prof. yashovardhankelkar
- [9] Isbell, R. F. (1996). The Australian Soil Classification. Australian soil and land survey handbook. (Vol. 4). Collingwood, Victoria, Australia: CSIRO Publishing.
- [10] Ibrahim, R. S. (1999). Data Mining of Machine Learning Performance Data. Unpublished Master of Applied Science (Information Technology), Publisher; RMIT University Press.
- [11] Soil Survey Staff 1998, Keys to soil taxonomy. Eight Edition, Natural Resource Conservation Services, USDA, Blacksburg, Virginia. Soil Survey Staff 1951, Soil Survey Manual. US Department of Agricultural Hand book No. 18.
- [12] P. Bhargavi, and Dr. S. Jyothi , Soil Classification Using GA Tree. International journal of computer science & information Technology (IJCSIT) Vol.2, No.5, October 2010.
- [13] P. Bhargavi, and Dr. S. Jyothi , Soil Classification by Generating Fuzzy rules. International Journal on Computer Science and Engineering, Vol. 02, No. 08, 2010, 2571-2576.
- [14] P. Bhargavi, and Dr. S. Jyothi , Fuzzy C-Means Classifier for Soil Data. International Journal of Computer Applications (0975 – 8887), Volume