

A Review of Data Mining Techniques for Soil Quality Analysis

Snehal Mohod¹, Rajesh Babu², Roshani Talmale³

¹M.Tech Student, Department of Computer Science & Engineering, Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, Maharashtra, India.

^{2,3}Assistant Professor, Department of Computer Science & Engineering, Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, Maharashtra, India.

Abstract— Data Mining is a method which centers on expansive data sets to remove data for expectation and disclosure of concealed patterns. Data Mining is appropriate for different zones like human services, protection, showcasing, retail, correspondence, agriculture. At first, this information extraction was figured and assessed physically utilizing measurable systems. Along these lines, semi-automated data mining systems rose due to the progression in the innovation. Such headway was additionally as a capacity which expands the requests of examination. In such case, semi-mechanized systems have turned out to be wasteful. Consequently, robotized data mining systems were acquainted with blend information productively. A study of the accessible writing on data mining and pattern recognition for soil data mining is displayed in this paper. Data mining in Agricultural soil datasets is a generally novel research field. Proficient strategies can be produced and customized for explaining complex soil datasets utilizing data mining.

Keywords— Data Mining, K-Means, Support vector machines, Artificial neural networks, Agriculture

I. Introduction

In the current days of society, data mining is utilized in monstrous regions and numerous off-the-rack data mining devices, systems, and methods are accessible and effective reach data mining application programming's are reachable, yet data mining in rural soil datasets is a similarly a juvenile research field. Presently multi day's data mining idea and strategies used to determine the agriculture issues. In this paper, it has been talked about how data mining procedures are connected in the agriculture field. All around, every day the prerequisite of nourishment is raising; henceforth the horticultural researchers, ranchers, government, and scientists are tedious to put additional endeavour and utilize various systems in agriculture for development underway. As an impact, the data produced in the field of horticultural data upgraded step by step. As the level of data extends, it requires natural route for these data to be mined and dissected when required. Indeed, even at present, a just a few ranchers are truly utilizing the new strategies, apparatuses and procedures in agriculture for better creation.

Data mining can be utilized for anticipating the future patterns of rural procedures. The procedures are valuable to inspire noteworthy and utilizable learning which can be seen by numerous people. Data mining programs comprises of different philosophies which are dominantly created and utilized by business endeavours and biomedical specialists. These strategies are very much arranged towards their particular learning area. The utilization of standard measurable investigation methods is both tedious and costly. Productive procedures can be created and customized for illuminating complex soil data sets utilizing data mining to enhance the viability and precision of the Classification of huge soil data sets [1].

A soil test is the examination of a soil test to decide supplement substance, synthesis and different qualities. Tests are typically performed to quantify fertility and demonstrate lacks that should be cured [2]. The soil testing research centers are furnished with reasonable specialized writing on different parts of soil testing, including testing strategies and plans of manure proposals [4]. It encourages agriculturists to choose the degree of compost and ranch yard fertilizer to be connected at different phases of the development cycle of the yield.

Over the years numerous algorithms were made to extricate learning from expansive arrangements of data. There are a few unique procedures to approach this issue: order, affiliation rule, bunching, and so on. Grouping methods are intended for arranging obscure examples utilizing data given by a lot of characterized tests.

This set is typically alluded to as a preparation set, in light of the fact that, as a rule, it is utilized to prepare the grouping system how to play out its order. The order undertaking can be viewed as a directed system where each occurrence has a place with a class, which is shown by the estimation of an uncommon objective characteristic or essentially the class qualities. Arrangement schedules with data mining utilize an assortment of calculations and the specific calculation utilized can influence the manner in which records are characterized. This work discusses K-Nearest Neighbor and Naive Bayes calculations.

K-Nearest Neighbor [4] does not have any learning stage, on the grounds that each time a grouping is performed it utilizes a preparation set. The presumption behind the k-closest Neighbor calculation is that a comparative characterization is created by comparable examples. The comparative realized examples utilized for doling out a characterization to an obscure example are depicted by the parameter K.

Naive Bayes [5] classifier accept that the nearness (or nonappearance) of a specific component of a class is disconnected to the nearness (or nonattendance) of some other element. Contingent upon the exact idea of the likelihood show, Naive Bayes classifiers can be prepared effectively in a regulated getting the hang of setting. In spite of its credulous plan and in all likelihood obvious suppositions, Naive Bayes work much better in numerous perplexing genuine circumstances.

In this study, the discussion focuses on the role of data mining in context of soil investigation in the field of agriculture. Furthermore it also presents the related work by a few authors in the domain agriculture data mining.

II. Literature Review

S. S. Bhaskar et al. [6] made a comparative study for soil classification of naïve bayes, JRip and J48. They found J48 to be the best method. They also used regression technique like linear regression and least square Median. They found least median squares regression produce better results for prediction than the classical linear regression technique.

Ravindra M et al. [7] uses decision tree in selecting the best suited pump for irrigation. D Ramesh et al. [8] uses Multiple Linear Regression for predicting rice yield. S. Veenadhari et al. [9] uses decision tree induction technique to analyze the influence of climatic parameter on soybean productivity. Georg Ruß [10] evaluates four regression techniques on agriculture data. He found support vector regression can serve as a better model for yield prediction among MLP, RBF and RegTree.

Jay Gholap [11] uses J48 algorithm for predicting soil fertility class. Also for performance tuning of J48 algorithm he uses attribute selection and boosting techniques. Suman et al. [12] cluster the data using the K-means Clustering on soil dataset then the linear regression is applied to classify the clusters. P. Revathi et al. [13] uses Naive bayes, j48, MLP, Random forest, Random tree, they found J48 gives better result over the other algorithms for cotton seed quality. D Ramesh et al. [14] uses MLP and K- Mean for yield prediction for East Godavari district of Andhra Pradesh.

III. Methodology

The ultimate aim of any technology with respect to agriculture is to make the food production cheap and at the same time to give farmers many immediate and sustainable benefits. There is need to transform huge amount of data that are available in lab and agriculture university into information. This can be possible with data mining. Data mining is a process of discovering previously unknown patterns that are used for strategic decision making.

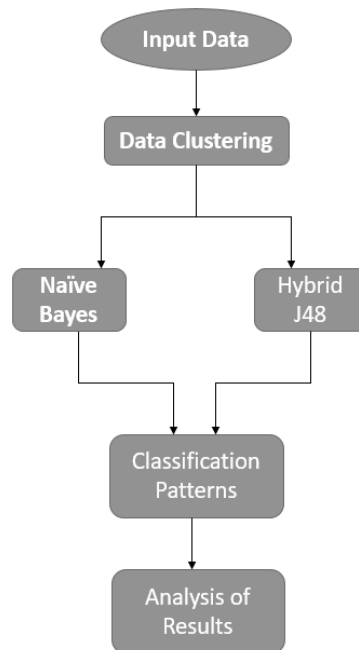


Figure 1: System Architecture

Accuracy of J48 (C4.5) algorithm for predicting soil fertility was highest, hence it was used as a base learner. Now, the aim was to increase its accuracy with the help of some other meta-techniques like attribute selection and boosting with the help of Weka. Thus we propose to implement Hybrid C4.5 with improvised efficiency of classification Result.

Attribute selection reduces dataset size by removing irrelevant/redundant attributes .It finds minimum set of attributes such that resulting probability distribution of data classes is as close as possible of original distribution. Attribute evaluator method – CfsSubsetEval was used, which evaluates the worth subset of attributes by considering the individual predictive ability of each attribute.

IV. Conclusions

Agribusiness is the most extreme essential territory particularly in the progressing nation like India. Utilization of data innovation in farming can change the situation of decision making and designers can yield better. This review examined the job of data mining in agriculture. As agriculture is a soil-based industry, there is no way that required yield increases of the major crops can be attained without ensuring that plants have an adequate and balanced supply of nutrients. The paper also proposes a new method for which uses a hybrid J48 classifier for analysis and predicting the soil behaviour.

References

- [1]. V.K. Singh, R. Piryani, A. Uddin, P. Waila, "Sentiment Analysis of Movie Reviews A new Feature-based Heuristic for Aspect-level Sentiment Classification", 978-1-4673-5090-7/2013 IEEE.
- [2]. V.K. Singh, R. Piryani, A. Uddin, P. Waila, "Sentiment Analysis of Movie Reviews and Blog Posts Evaluating SentiWordNet with different Linguistic Features and Scoring Schemes", 978-1-4673-4529-3/ 2012 IEEE.
- [3]. Peter D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Institute for Information Technology National Research Council of Canada Ottawa, Ontario, Canada, K1A 0R6.
- [4]. Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning".
- [5]. Hai-bing ma, Yi-bing geng, Jun-rui qiu, "Analysis of three methods for web-based opinion mining", Proceedings of the 2011 International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011.
- [6]. S.S.Baskar L.Arockiam S.Charles "Applying Data Mining Techniques on Soil Fertility Prediction" International Journal of Computer Applications Technology and Research Volume 2– Issue 6, 660 - 662, 2013
- [7]. Ravindra M, V. Lokesh, Prasanna Kumara, Alok Ranjan "Study and Analysis of Decision Tree Based Irrigation Methods in Agriculture System" International Journal of Emerging Technology and Advanced Engineering ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 2, Issue 12, December 2012
- [8]. D Ramesh, B Vishnu Vardhan "Region Specific Crop Yield Analysis: A Data Mining Approach" UACEE International Journal of Advances in Computer Science and its Applications – IJCSIA Volume 3 : Issue 2 [ISSN 2250 – 3765] 05 June 2013
- [9]. S. Veenadhari, Dr. Bharat Mishra, Dr.CD Singh "Soybean Productivity Modeling using Decision Tree Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 27– No.7, August 2011
- [10]. Georg Ruß "Data Mining of Agricultural Yield Data: A Comparison of Regression Models"
- [11]. Jay Gholap "Performance Tuning of J48 Algorithm for Soil Fertility" 2012. Asian Journal of Computer Science and Information Technology 2: 8 (2012) 251– 252
- [12]. Suman, Bharat Bhushan Naib "Soil Classification and Fertilizer Recommendation using WEKA" IJCSMS International Journal of Computer Science & Management Studies, Vol. 13, Issue 05, July 2013
- [13]. P. Revathi, Dr. M. Hemalatha "Categorize the Quality of Cotton Seeds Based on the Different Germination of the Cotton Using Machine Knowledge Approach" International Journal of Advanced Science and Technology Vol. 36, November, 2011
- [14]. D Ramesh, B Vishnu Vardhan "Data Mining Techniques and Applications to Agricultural Yield Data" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013