JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AN ANALYTICAL REVIEW OF TECHNIQUES FOR SOIL QUALITY ANALYSIS

A REVIEW

[1]Aishwarya Bagde, [2]Mona Mulchandani, [3]Gunjan Jewani
[1]Research Scholar, [2]Assistant Professor, [3]Assistant Professor
[1]Department of Computer Science & Engineering,
[1]Jhulelal Institute of Technology, Village Lonara, Off, Koradi Rd, Nagpur, Maharashtra- 441111

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

Despite the fact that data mining is used in numerous monstrous regions today and that numerous off-the-shelf data mining devices, systems, and methods are available as well as effective reach data mining application programming's are reachable, data mining in rural soil datasets is still a relatively new research field. At the moment, the idea and tactics for data mining are being employed to determine the agriculture concerns over a period of many days. Data mining processes in the agricultural industry have been discussed in detail in this work, as well as their connections. Every day, the demand for food is increasing all over the world; as a result, horticulture experts, ranchers, government officials, and scientists are working tirelessly to put up greater effort and apply various agricultural systems for development that are currently occurring. As a result, the data produced in the field of horticultural data has been gradually enhanced over time. As the amount of data grows, it becomes more important to have a natural route for data to be mined and analyzed when it is needed. And it is true that, even now, only a small number of ranchers are genuinely utilizing the modern agricultural strategies, apparatuses, and procedures to improve their production.

Data mining can be used to predict future trends of rural procedures, which can be useful for planning purposes. The methods are beneficial in that they stimulate significant and transferable learning that may be observed by a large number of people. Many diverse ideologies behind data mining tools, the majority of which were developed and are currently in use by commercial enterprises and biological specialists. These tactics are very much geared around the specific learning area in which they are employed.

Using traditional measured inquiry procedures is time-consuming and expensive, therefore it is not recommended. Promising processes for lighting complex soil data sets can be developed and tailored to meet specific needs, with data mining being used to improve the viability and precision of the Classification of large soil data sets [1, 2].

It is the examination of a soil test that determines the substance, synthesis, and distinct properties of a supplement ingredient. In most cases, tests are carried out to quantify fertility and identify any deficiencies that should be addressed [2]. It is provided to the soil testing research institutions with reasonable specialized writing on many aspects of soil testing, including testing methodologies and plans for manure suggestions [4]. In order to maximize yield, agriculturists should choose the amount of compost and ranch yard fertilizer to use at various stages of the crop's development cycle.

A large number of algorithms have been developed throughout the years to extract learning from large amounts of data arranged in complex ways. There are a few different ways to handle this problem, including the following: order, affiliation rule, bunching, and so on. Grouping procedures are intended for the purpose of organizing ambiguous examples based on data provided by a large number of characterized tests.

This set is frequently referred to as a preparation set, due to the fact that it is typically used to instruct the grouping system on how to carry out its order of operations. This can be seen as a directed system where each occurrence has an assigned place in the class, as demonstrated by the estimation of an exceptional objective characteristic or, more specifically, the characteristics of the class in which it occurs. Data mining arrangement schedules make use of a variety of calculations, and the specific calculation that is used might have an impact on the way records are classified and categorized. The calculations of K-Nearest Neighbor and Naive Bayes are discussed in this paper.

As a result, K-Nearest Neighbor [4] does not have a learning stage because it performs groupings using the same preparation set each time they are performed. The underlying assumption behind the k-closest Neighbor computation is that comparable samples can be used to generate a comparative characterization of the data. The parameter K represents the number of comparative realized examples that were used in the process of assigning a characterization to an ambiguous example.

It is acceptable for the Naive Bayes [5] classifier to accept that the proximity (or nonappearance) of a given component of a class is unrelated to the proximity (or nonattendance) of some other element in the class. The precise idea of the likelihood show can influence the preparation of Naive Bayes classifiers, which can be done efficiently in a regulated learning environment. Despite the fact that it has a credulous strategy and most likely obvious suppositions, Naive Bayes performs significantly better in a variety of confusing real-world situations.

Discussion in this paper is primarily focused on the role of data mining in the context of soil investigation in the field of agricultural research and development. Aside from that, it also includes work by a few authors in the field of agriculture data mining that is linked to the work presented here.

II. Literature review

S. S. Bhaskar and colleagues [6] conducted a comparative research of naive bayes, JRip, and J48 soil classification methods for soil classification. They discovered that J48 was the most effective approach. They also employed regression techniques such as linear regression and least squares Median to get their results. They discovered that the least median squares regression technique outperformed the standard linear regression technique when it came to prediction.

When choosing the optimal irrigation pump for a given situation, Ravindra M et al. [7] employ a decision tree. Multiple Linear Regression is used by D Ramesh et al. [8] to forecast rice yield. When S. Veenadhari and colleagues [9] evaluate the influence of meteorological parameters on soybean productivity, they employ the decision tree induction technique. Georg Ruß [10] compares and contrasts four different regression approaches using agricultural data. He discovered that support vector regression can be a superior model for yield prediction when compared to MLP, RBF, and RegTree.

In order to forecast soil fertility class, Jay Gholap [11] makes use of the J48 algorithm. He also employs attribute selection and boosting approaches in order to fine-tune the performance of the J48 algorithm. Suman and colleagues [12] cluster the data using the K-means clustering algorithm. On the soil dataset, clustering is performed, and then the linear regression algorithm is used to classify the clusters. P. Revathi et al. [13] used Naive Bayes, j48, MLP, Random forest, and Random tree to determine the quality of cotton seeds. They discovered that J48 outperformed the other algorithms in terms of cotton seed quality. D Ramesh et al. [14] employ MLP and K- Mean to estimate yields in the East Godavari district of Andhra Pradesh, according to their findings.

III. RESULT And DISSCUSION

Our research team gathered agricultural soil data from the soil testing facility in the Virudhunagar District as part of this project. We have collected 110 data points, each of which comprises attributes such as Village Name, Soil Type or Color, Soil Texture, PH, EC (Electrical Conductivity), Lime Status, and Phosphorous concentration. The soil types Red and Black were predicted by this technique based on the PH and EC values of the soil. The PH value of Black soil was discovered to be larger than 7.7, whereas the PH value of Red soil was discovered to be less than 7.7.

To predict the soil types Red and Black, we used three classification algorithms, including JRip, J48, and Naive Bayes, to make predictions. JRip takes into account the entire property while applying three different classifier techniques. However, the J48 classifier only takes into account the PH and EC values. The tree is constructed using the two attributes listed above. This soil data set demonstrates that the JRip classifier creates rules quickly and efficiently, and that it performs well. When evaluating the accuracy of these three methods, JRip was shown to be the most accurate. The entire dataset is used as the training set in this case.

The weighted average of True Positive Rate of the JRip classifier is 0.982, according to the training data set, which was used in the analysis. When the J48 and Nave Bayes TP Rates are 0.97 and 0.86, respectively, it suggests that the level is low. As a result, the data set was automatically categorised in a higher sense by the JRip classifier. The soil parameters of locations with Red grained soils and sites with Black textured soils were different. It follows that soil with a pH below 7.0 is acidic, whereas soil with a pH above 7.0 is alkaline. The spectral analysis was sensitive enough to capture the differences in soil fertility between the three soil types studied in this study. There are several soil datasets that contain features such as soil type, pH value, and so on. Table 1 presents the results of a comparative comparison of classifiers. When compared to the other algorithms, JRip fared significantly better in classification, and the Kappa Statistic value for JRip method is the closest to 1.00.

Table 1. Comparative Analysis of Classifiers

Evaluation Criteria (Total number of instances 110)	Correctly Classified Instances	Incorrectly Classified Instances	Prediction Accuracy (%)
JRip	106	4	96.60%
J48	103	7	94.03%
Naïve Bayes	95	15	86.36%

Figure 1. Comparison of Algorithm Based on Correctly and Incorrect outcome

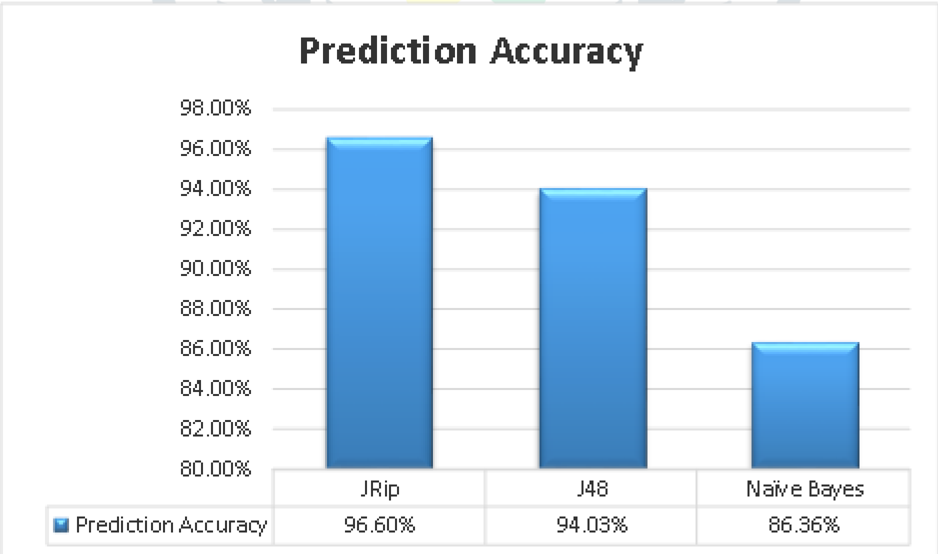

Figure 2. Comparison of Algorithm Based on Prediction Accuracy

IV. CONCLUSIONS

When it comes to developing countries such as India, agribusiness is the most critical sector to understand and master. The application of data innovation in farming has the potential to alter the decision-making environment, and designers can produce better results. The purpose of this review was to explore the role of data mining in agriculture. Because agriculture is a soil-based sector, it is impossible to achieve the necessary yield increases of the principal crops without first guaranteeing that the plants receive an appropriate and balanced supply of nutrients from their environment. The comparative study of three algorithms, such as Nave Bayes, JRip, and J48, is planned in this work. Nave Bayes, Jip, and J48 are the algorithms under consideration. JRip classification algorithm produces a superior result for this dataset and accurately classifies the greatest number of occurrences when compared to the other two algorithms. It is possible to recommend Jip for predicting soil types.

V. 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. Lokesha, 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 – 8887Volume 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