Review of Data Mining Techniques for Soil Quality Analysis

Abha Natu[1], Neha Titarmare[2], Rupal Lokare[3], Shreya Bagde[4], Prof. Mitali Ingle[5]

[1]BE Students, Department of Computer Science & Engineering, Dr. Babasaheb Ambedkar College of Engineering and Research, Nagpur, Maharashtra, India.
[2]BE Students, Department of Computer Science & Engineering, Dr. Babasaheb Ambedkar College of Engineering and Research, Nagpur, Maharashtra, India.
[3]BE Students, Department of Computer Science & Engineering, Dr. Babasaheb Ambedkar College of Engineering and Research, Nagpur, Maharashtra, India.
[4]BE Students, Department of Computer Science & Engineering, Dr. Babasaheb Ambedkar College of Engineering and Research, Nagpur, Maharashtra, India
[5]Assistant Professor, Department of Computer Science & Engineering, Dr. Babasaheb Ambedkar College of Engineering and Research, 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.

Keyword: - *Mining, K-Means, Support vector machines, Artificial neural networks, Agriculture.*

1. INTRODUCTION

In the current days, data mining is utilized in various fields. Presently multi day's data mining idea and strategies used to determine the agriculture issues. In this paper, it has been discussed 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 endeavor 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.

Data mining can be utilized for anticipating the future patterns of rural procedures. Data mining programs comprises of different sets which are created and utilized by business endeavors 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 to choose supplement substance, sythesis and diverse characteristics. Tests are commonly performed to evaluate fruitfulness and show does not have that ought to be relieved [2]. The dirt testing research focuses are outfitted with sensible particular composition on various pieces of soil testing, including testing methodologies and plans of fertilizer recommendations [4]. It urges agriculturists to pick the level of fertilizer.

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 commonly recommended to as a readiness set, in light of the way that, when in doubt, it is used to set up the gathering framework how to play out its request. The request undertaking can be seen as a coordinated framework where every event has a spot with a class, which is appeared by the estimation of a remarkable target trademark or basically the class characteristics. Course of action plans with information mining use a combination of computations and the particular figuring used can impact the way in which records are portrayed. This work talks about K-Nearest Neighbor and Naive Bayes computations.

K-Nearest Neighbor [4] does not have any learning stage, in light of the fact that each time a gathering is performed it uses a planning set. The assumption behind the kclosest neighbor computation is that a relative portrayal is made by similar models. The relative acknowledged models used for appointing out a portrayal to a dark precedent are delineated by the parameter K.

Navie Bayes [5] classifier acknowledge that the proximity (or nonappearance) of a particular segment of a class is detached to the closeness (or nonattendance) of some other component. Dependent upon the definite thought of the probability appear, Naive Bayes classifiers can be arranged viably in a controlled getting the hang of setting. Credulous Bayes work much better in various perplexing conditions.
In this study, the discussion focus on the role of data mining in context of soil investigation in the field of agriculture.

2. LITERATURE REVIEW

S. S. Bhaskar et al. [6] made a relative report for soil characterization of credulous bayes, JRip and J48. They observed J48 to be the best technique. They likewise utilized relapse procedure like straight relapse and least square Median. They discovered least middle squares relapse produce preferable outcomes for expectation over the established direct relapse strategy.

Ravindra M et al. [7] utilizes choice tree in choosing the most appropriate siphon for water system. D Ramesh et al. [8] utilizes Multiple Linear Regression for anticipating rice yield. S. Veenadhari et al. [9] utilizes choice tree acceptance strategy to dissect the impact of climatic parameter on soybean profitability. Georg Ruß [10] assesses four relapse systems on farming information. He discovered help vector relapse can fill in as a superior model for yield forecast among MLP, RBF and RegTree.

Jay Gholap [11] utilizes J48 calculation for foreseeing soil fruitfulness class. Likewise for execution tuning of J48 calculation he utilizes characteristic determination and boosting methods. Suman et al. [12] bunch the information utilizing the K-implies Clustering on soil dataset then the straight relapse is connected to arrange the groups. P. Revathi et al. [13] utilizes Naive bayes, j48, MLP, Random timberland, Random tree, they discovered J48 gives better outcome over alternate calculations for cotton seed quality. D Ramesh et al. [14] utilizes MLP and K-Mean for yield forecast for East Godavari locale of Andhra Pradesh.

3. METHODOLOGY

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.
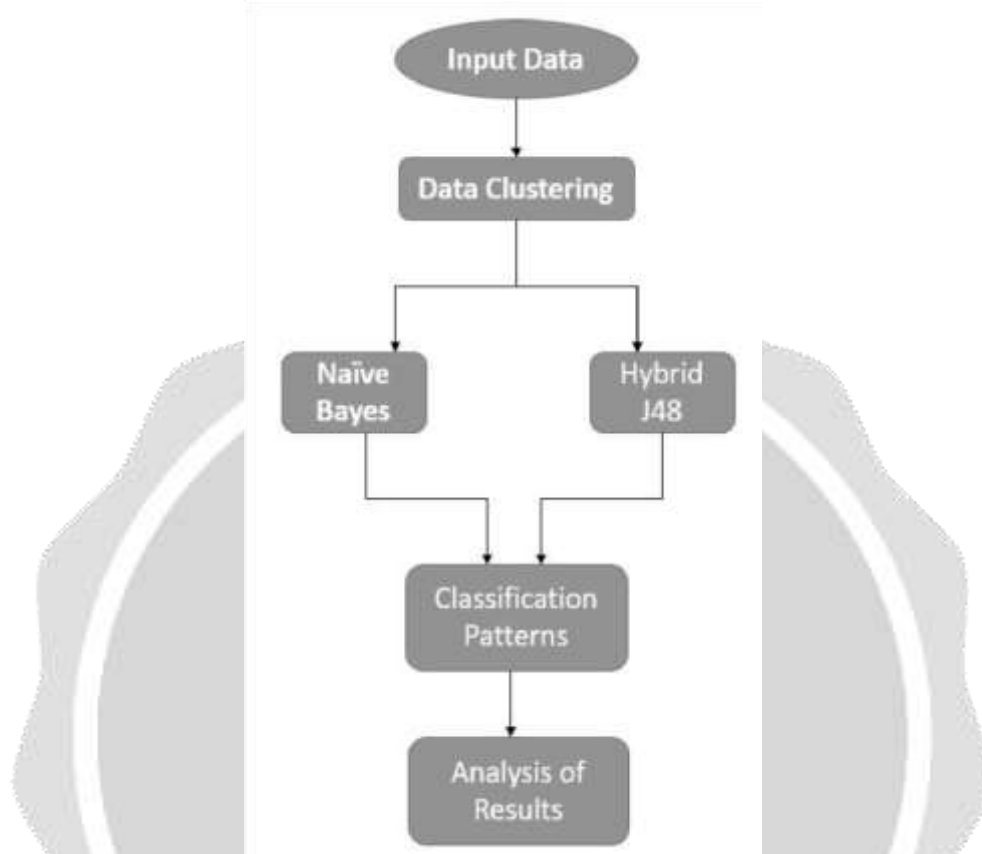

Fig-1: System Architecture

In this model info is gathered from the informational collections and further sent for information cleaning. In information cleaning, information pre-preparing is done in which all the loud information is expelled. Quality determination lessens dataset estimate by expelling unimportant/excess characteristics. It discovers least arrangement of characteristics with the end goal that subsequent likelihood dispersion of information classes is as close as conceivable of unique dissemination. At that point the information which is gathered from informational index is send to the second modular which is information bunching.

In information bunching two calculations are connected Naïve Bayes and Hybrid J48. The outcome is send to grouping model in which arrangement models are produced and the last outcome is send for examination.

4. REFERENCE

[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