Data Mining Techniques in Soil Quality Analysis- Review

Prof. Mitali Ingle

Assistant Professor, Department of Computer Science & Engineering, Dr. Babasaheb Ambedkar College of Engineering and Research, Nagpur, Maharashtra, India.

Prof.Shraddha Waghmare

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, 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 decide supplement substance, sythesis 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.

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 suggested 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 kclosest neighbor calculation is that a

comparative characterization is created by comparable examples. The comparative realized examples utilized for assigning out a characterization to an obscure example are depicted by the parameter K.

Navie 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. Naive Bayes work much better in numerous puzzling circumstances.

In this study, the discussion focus on the role of data mining in context of soil investigation in the field of agriculture.

II. LITERATURE REVIEW

Shravan Vishwanathan et al., [7] proposed Reviews of ruined tomato is accumulated from one of the databases. At that point on each overview, tokenization is done, channel the tokens by the length. After that stemming is performed and after that oust tokens which are not required for the sentiment analysis.

Increment director is used which differentiate each token and the positive word dictionary and negative word vocabulary. At whatever point given token matches with any of the word dictionaries than token is arranged into that class. After that entire all the occasion at both positive database and negative database. Apply join executive which subtract the positive entire and negative sum and create the class name of review and show it to the customer.

Santanu Modak et al., [8] in this paper considered is done on different approaches for sentiment gathering. With the objective that information is used for future research. Cushy Sets or feathery [9] arrange methodology is used for Opinion Mining or sentiment analysis. In this procedure feathery set is prepared which is used to figure the dimension of positive and negative of sentiment words.

Su, Qi, et al. proposed a shared stronghold approach to deal with the feature-level opinion mining issue [10]. The gathering was done on thing features and opinion words in the meantime and iteratively by interlacing both their substance information and sentiment interface information. They built up the sentiment association set between the two social affairs of data inquiries by recognizing their most grounded n sentiment joins. POS tagger used to recognize sentiment word and thing features. Using sentiment word and thing features they construed alliance guideline to perceive disguised sentiment. Finally, sentiment scoring was done.

This examination presumes that, if Sentiment Analysis is a backslide type issue, we can pick feathery set, which is exceptionally contrasted with different techniques hence. In case we consider Sentiment Analysis is a portrayal type issue, we can pick semi-coordinated learning or controlled machine learning approach. The little dataset is used for getting ready in a semi-managed approach. The classifier is used for managed machine learning approach. Out everything considered, Maximum Entropy Classifier creates as a rule extraordinary result, anyway Support Vector Machine (SVM) convey best result immaculate.

Khin Phyu Shein et al., [11] on the Internet there are piles of substance that opinion or sentiments around a thing, for instance, the study about music, movie, programming, thing, and books, etc. The purpose of sentiment course of action is to remove the feature on which analyst express their inclination or feeling and recognize them as positive, negative or unprejudiced.

In this paper, the proposed presentation is the mix of Support Vector Machine with Natural Language Processing strategies, rationality based on Formal Concept Analysis plan for orchestrating the item reviews are negative, positive or fair-minded. In its proposed model essential spotlight is on feature level sentiment portrayal. the three central parts in this approach are: doling out the POS marks, recognizing spacerelated features and requesting the sentiment words. They use Part Of Speech (POS) tagger to consign.

Kang Wu et al., [12] revolve around sentiment analysis of topical Chinese microblogs. In this paper, most pervasive microblog of China is taken i.e. Sina Weibo. The customer of Weibo forms their messages that contain typically unique sentences, messages length is up to 140 Chinese Microblog contain a couple of sentences, which empower customers to give their understanding. The study shows that Chinese people express their sentiments in an underhanded way. For a gathering of such sentiments, we require more semantics. The proposed model first, analyzed the Chinese Microblogs which express the opinion of the

customer, and analysis of features of the single sentence. Second, to streamline the delayed consequence of sentiment game plan we use sentence relationship.

Asha S Manek et al., [13] proposed a model for recognizing spamming works out, for instance, creating fake reviews around a thing to mislead the customers. This model uses successful Repetitive Pre-processing (SentReP) which is based on focused pre-processing and attempted parameters for characterizing the reviews. To get "once-over of-words" movie reviews are pre-arranged. After that each review experience the going with advances: tokenization, case change, custodian and snowball stemming methodology and a while later stop words are ousted. After pre-processing crossendorsement is performed which include two phases: I) every property weight calculation and ii) by weight select best K properties.

Mostafa Karamibekr et al., [14] Sentiment analysis has done only for the thing, organizations or movies, not for social issues. For government work, it is critical to understand the general opinions with respect to social issues. Accordingly, first, we ought to acknowledge how social issues are interesting in connection to thing and organizations. What is important is that it is definitely not hard to describe features for a thing, anyway not for social issues. In the social domain, activity word expects a basic occupation to express opinions of the customer. In sentiment analysis of social issues first, from each sentence, we assemble the opinions, construct opinion structures, and thereafter, their acquaintances are settled in respects with social issues.

III. 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.

In this model input is collected from the data sets and further sent for data cleaning. In data cleaning, data pre-processing is done in which all the noisy data is removed. 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. Then the data which is collected from data set is send to the second modal which is data clustering.

In data clustering two algorithms are applied Naïve Bayes and Hybrid J48. The result is send to classification model in which classification models are generated and the final result is send for analysis.

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