iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4387

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Review on Analysis of Agriculture Data Using Data Mining Techniques

Ms. Saima Humma[1], Prof. Anil Kumar Mishra[2]
[1, 2]*Department of Computer Science and Engineering, Global Institute of Technology & Management, Maharishi Dayanand University (MDU) Rohtak, Haryana, India*

Abstract: ***In this time of gigantic data, both structured and unstructured, there is a need of systemized management and retrieval of information. Data mining techniques play an imperative part in extraction and analysis of data. The various fields where data mining techniques can be actualized to analyze data are education, engineering, medicine, finance, communication, agriculture, marketing and the rundown goes on. In this paper a survey has been done to outline a various data mining techniques that are implemented to analyze agriculture data. Agriculture is one of the sophisticated fields in the developing countries like India. So as to augment the crop yield, the analysis of agricultural data is a must do exercise. The motivation behind executing data mining techniques on agricultural data is to discover the ideal solutions and parameters that contribute to the overall production of crop and thus to the development of a nation.***

Keywords: Data Mining, Agricultural Data Analysis, Clustering, K-Means, PAM, CLARA and Multiple Linear Regression.

I. INTRODUCTION

Agriculture is the tremendous territory where data has no noteworthiness unless used in a way that it provides solutions to increase the crop yield. Statistics is the most conventional and customary method of agricultural data analysis which consumes a great deal of time, vitality and labor. Data mining techniques using computer software are powerful approach for accomplishment of practical and effective solution for this kind of issue. Retrieval of information through computer systems can help in making effective choices and avoid human error. Environmental conditions, crop type and soil type are the basic data elements that can help to determine the productivity of crops at a particular geographical area. This approach helps farmers and other agricultural analyzers to settle on decisions with respect to the crop productivity in much less time and efforts that were used to be taken traditionally.

There are two sorts of learning methods in data mining one is unsupervised and another is supervised. Clustering is the unsupervised process [1] in which vast collection of abstract objects are assembled into groups or clusters as indicated by a few criteria. The different clustering methods are K-Means, PAM, CLARA, DBSCAN etc. This paper accentuates on the survey of analysis of agricultural data and discovering resolutions to expand the crop yield utilizing various data mining techniques. This process involves mining of colossal measure of data about soil, crop, climatic conditions and so on using data mining techniques and then analyzing the impact of varies quantities and different patterns of these elements on the crop production. The various data mining techniques produce the values that are most prominent and effective values to fulfill the necessary criteria of a parameter to help in maximizing the crop production.

II. RELATED WORK

Since information mining in agriculture is an exceptionally late research subject, not much research has been done on it. It comprises of utilization of information mining methods on agribusiness information. These days, late innovations and recent technologies are implemented on unstructured data to provide a lot of agriculture related information, which is then analyzed and compared in order to devise some useful facts that help to maximize crop yield.

The related research on the proposed work is quickly specified as follows (most recent first).

1) Jharna Majumdar et al., [2] gave a technique where the research was done on agriculture data of different districts of Karnataka. This paper concentrates on the analysis of the agriculture data and finding ideal parameters to expand the crop production utilizing data mining techniques like PAM, CLARA, DBSCAN and Multiple Linear Regression. Mining the immense measure of existing crop, soil and climatic information, and dissecting new, non-experimental data optimizes the production and makes agriculture more resilient to climatic change. As indicated by the investigations of clustering quality metrics, DBSCAN gives the better clustering quality over PAM and CLARA, CLARA gives the better clustering quality over PAM.

©IJRASET (UGC Approved Journal): All Rights are Reserved

2) Niketa Gandhi et al., [3] provided a review of research done on the application of data mining on agriculture decision making. This review has mentioned the application of various data mining techniques and their role in estimating the impact of various climatic conditions and other essential factors on crop production. The conclusion of this review paper states that further investigations can be performed by implementing data mining techniques on complex agricultural parameters.

3) Vinayak A. Bharadi et al., [4] obtained the data for the years from 2005 to 2009 from website of Planning Commission of India. It contains information about plantation, fruits and vegetables of 35 states of India. The dataset contains total 4180 instances having eight attributes viz Year, State, Crop type, Crop name, Area, Production, Rainfall and Temperature. WEKA (Waikato Environment for Knowledge Analysis), an open source system, is used which provides filters for preprocessing of data. In this paper data mining algorithms like K-means, DBSCAN and EM are used to form clusters of data. DBSCAN algorithm gives similar results as base algorithm K-means whereas EM gives more specific production values on given rainfall and temperature range as compared to K-means and DBSCAN.

4) M.C.S.Geetha [5] coordinated the work of various authors in a single place so it becomes valuable for analysts to get information of current situation of data mining strategies and applications in setting to agriculture field. In this paper it has been expressed that how data mining methods like classification, association analysis, clustering and regression can help in examination of the agriculture data and devising ideal parameters to amplify the yield utilizing data mining techniques like K-means, neural networks, fuzzy set, decision tree and Bayesian classification, k-nearest neighbor and Support Vector Machine (SVM).

5) B. Milovi et al., [6] states data mining methods like Classification and Regression, Association rules, Cluster Analysis and so forth can be used to pick up knowledge about agriculture data. The data mining techniques such as Artificial Neural Networks, Condition Tree, Genetic Algorithm, and Nearest Neighbor Method can be actualized on data sets related to agriculture so that agrarian organizations are capable of producing descriptive and predictive information as support to decision making.

6) D Ramesh et al., [7] presents a concise investigation of harvest yield prediction utilizing Density Based Clustering Technique and Multiple Linear Regression (MLR) Technique for the selected district i.e. East Godavari of Andhra Pradesh, India. At first the statistical model Multiple Linear Regression Technique is implemented on existing data. The outcomes so acquired were verified and analyzed utilizing the data mining technique namely Density-Based Clustering Technique

7) Dr P Jaganathan et al., [8] have presented some of the commonly used data mining techniques in the field of agriculture. Some of these techniques, such as the k-means, the k nearest neighbor and bi-clustering are discussed and an application in agriculture for each of these techniques is presented. As per their conclusion proficient techniques can be developed and customized for tackling complex agricultural issues utilizing data mining. They gave a thought of application of k-means to evaluate soil fertility, application of k-nearest neighbor for classification of forest data with remotely detected images and Prediction of tricky wine fermentation.

8) Alberto Gonzalez-Sanchez et al., [9] have done broad investigation on prescient capacity of machine learning techniques, for example, multiple linear regression, regression trees, support vector regression, k-nearest neighbor and artificial neural network for crop yield production.

9) MotiurRahman M et al., [10] had a keen study on climate of Bangladesh and analyzed its impact on the production of rice. Bangladesh has a number of variants in rice that have variant harvesting seasons. So, the parameters like rainfall, wind, humidity, clouds, sunlight and temperature. The average data associated with each parameter was analyzed using clustering method.

10) D. Rajesh [11] gave an overview of application of spatial data mining technique on agriculture data. He implemented the clustering algorithm k-means on data. In this paper the temperature and the rainfall is considered as the initial spatial data. Agriculture meteorology is analyzed for the maximization of crop yield and crop loss too.

III. METHODOLOGIES

A. Clustering

Clustering is an unsupervised learning method of data mining in which abstract objects having similar kind of attributes or characteristics are assembled into groups or clusters (Fig. 1). The distance between any two objects of same cluster is smaller than the distance between objects of distinct clusters. Different techniques are implemented to aggregate the data, reasonable for desired data analysis. These techniques are backed by the algorithms that are applied to different data models. The different clustering techniques are k-means, k-medoids or PAM (Partition Around Medoids), CLARA (Clustering Large Applications), DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

©IJRASET (UGC Approved Journal): All Rights are Reserved

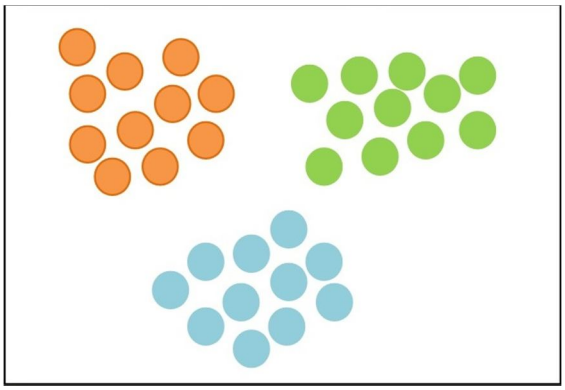


Fig1. Clusters: an Example

Clustering has a broad application spectrum in almost every area like artificial intelligence, medical science, marketing strategies, fraud detection, crime investigation, web search, economy factors and the list goes on. The benefit of clustering over other learning methods is its flexibility to the changes. The various clustering techniques are explained one by one below.

1) *K-Means:* K-means is the technique proposed by Stuart Lloyd in 1957 as a technique for pulse-code modulation at Bell Laboratory [12] in which '*n*' number of data objects are partitioned into '*k*' number of groups according to some mean value with which the elements are compared. Each group has a member called as centroid which becomes the new mean value of that group. This mean value acts as the prototype of the cluster which helps in repeating the procedure until convergence is reached. This algorithm is very quick and can be run multiple times.
2) *PAM:* PAM (Partition Around Medoids) is a partition based clustering technique or algorithm in which data is partitioned into various groups on the basis of centrally located objects called as medoids. Two sets of objects are created one contains selected objects i.e., medoids and another contains unselected objects. The purpose of this technique is to reduce the average dissimilarities of objects with respect to their nearest selected object.
3) *CLARA:* CLARA (Clustering for Large Applications) is a clustering algorithm that was devised by Kaufman and Rousseeuw in 1990.[13] This algorithm was suggested in extension to existing clustering techniques in order to deal with large number of objects. CLARA focuses on sampling of data objects rather than evaluating medoids. It illustrates a small sample from dataset and then implements PAM clustering algorithm to obtain medoids for sampling process. The average dissimilarity is then calculated between each object and the corresponding medoid.
4) *DBSCAN:* DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density based clustering technique and an algorithm which is used for the clustering of data. This algorithm was proposed by Hans-Peter Kriegel, Jörg Sander, Martin Ester, and Xiaowei Xu in 1996.[14] It assembles set of closely packed or dense data values together leaving less dense values aside. In this DBSCAN has two parameters viz. Eps (neighborhood radius value) and minPts (minimum neighbors to consider a point as core point) that are to be determined.

B. Multiple Linear Regression

Multiple Linear Regression (MLR) is a statistical model that models the relationship between a scalar dependent variable and more than one independent variables. When there is only one independent variable then this phenomenon is termed as Simple Linear Regression. Any attribute can be set as dependent variable and all other attributes can be analyzed in its relation. This way we get to know the behavior and values of independent variables with respect to the dependent variable. In MLR technique one attribute is used to predict the other.

Data mining algorithms behave differently with different kind of datasets depending upon the characteristics and the associations of the data. So one needs to determine the behavior of each algorithm on each dataset and find out the optimal resolutions to apply the techniques by comparing the results.

IV. CONCLUSION

Different data mining techniques are actualized on the input data to survey the best execution yielding strategy. The most recent work has been done using data mining techniques such as PAM, CLARA and DBSCAN to obtain the ideal climate prerequisite of

©IJRASET (UGC Approved Journal): All Rights are Reserved

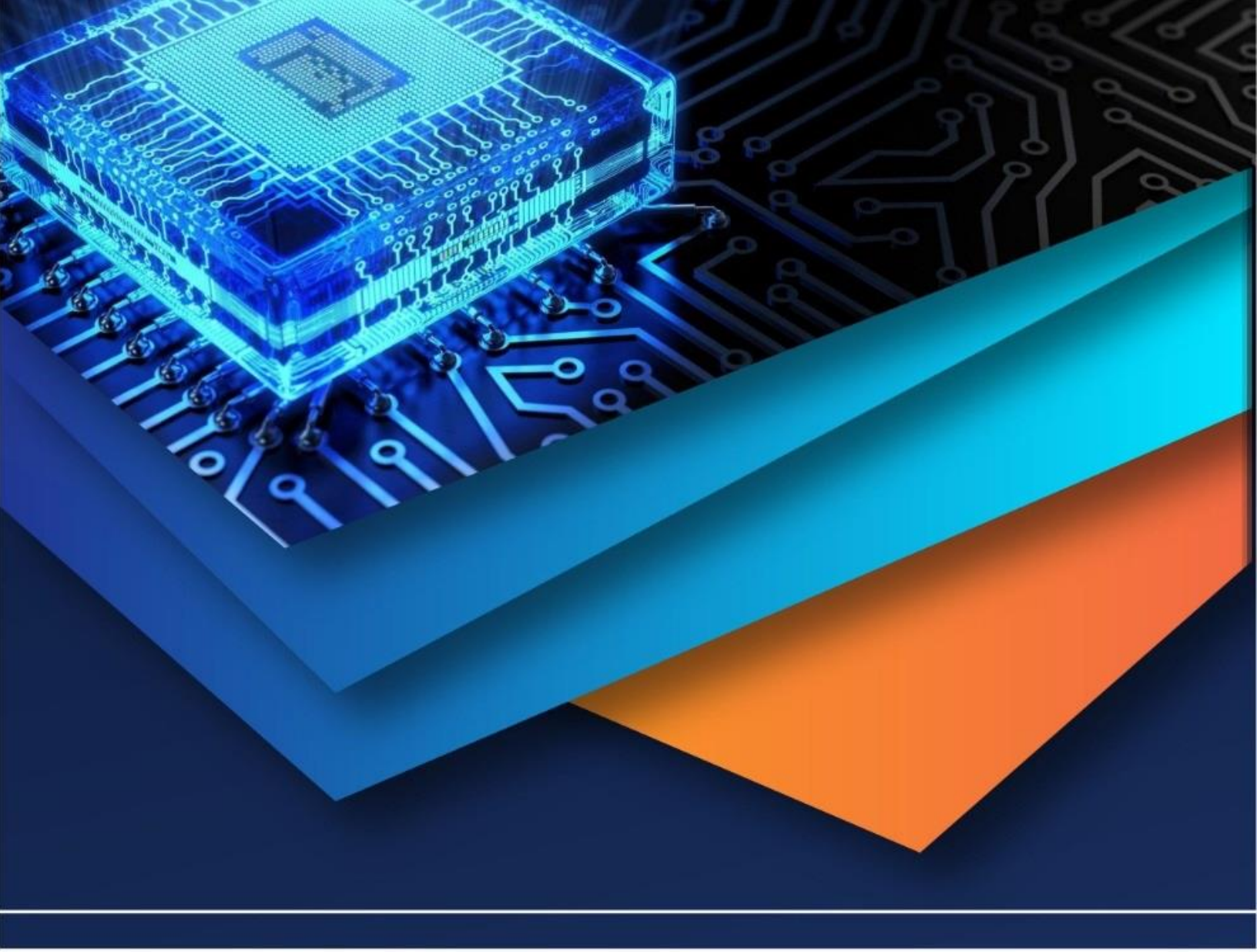
wheat. As the resultant it has been analyzed that DBSCAN gives the optimal clustering quality over PAM and CLARA, CLARA gives the optimal clustering quality over PAM.

The proposed work can likewise be stretched out to analyze some other agriculture related variables for the maximum crop yield. Further investigations are required to be done in order to understand the working of data mining techniques in the analysis of agriculture related data. Clustering techniques are not restricted to be used in particular area like agriculture sector but can be implemented in any area and on any kind of data to illustrate the strategies in which there is the maximum benefit.

V. REFERENCES

[1] Data clustering: a review by Jain A, Murty MN, Flynn PJ. ACM Comput Surv. 1999;31(3):264–323.
[2] Analysis of Agriculture Data Using Data Mining Techniques: Application of Big Data by Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, Department of M.Tech CSE, NMIT, Bangalore 560064 India), Springeropen, DOI 10.1186/s40537-017-0077-4 July, 2017
[3] A review of the application of data mining techniques for decision making in agriculture by Niketa Gandhi (University Dept. of Computer Science, University of Mumbai, Mumbai, Maharashtra, India) and Leisa J. Armstrong (University Dept. of Computer Science, University of Mumbai and School of Science, Edith Cowan University, Perth, Western Australia) IEEE, Accession number: 16854916 PP. May, 2017
[4] Analysis And Prediction In Agricultural Data Using Data Mining Techniques by Vinayak A. Bharadi, Prachi P. Abhyankar, Ravina S. Patil, Sonal S. Patade, Tejaswini U. Nate, Anaya M. Joshi, (Finolex Academy of Management and Technology, India) IJRISE PP. ISSN: 2394-8280 Mar, 2017
[5] A Survey on Data Mining Techniques in Agriculture by M.C.S.Geetha (Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India) IJIRCCE, PP. ISSN:2320-9798 Feb, 2015
[6] Application of Data Mining in Agriculture (Bulgarian Journal of Agricultural Science) by B. Milovi (Agricultural Enterprise "Sava Kovačevic" at Vrbas, 21460 Vrbas, Serbia) & V. Radojevi (University of Novi Sad, Faculty of Agriculture, 21000 Novi Sad, Serbia) www.agrojournal.org, www.researchgate.net, www.academia.edu PP. Jan, 2015
[7] Analysis of Crop Yield Prediction Using Data Mining Techniques by D Ramesh (Associate Professor, Department of CSE, JNTUH College of Engineering, Telangana State, India) & B Vishnu Vardhan (Professor, Department of CSE, JNTUH College of Engineering, Telangana State, India) IJRET pISSN: 2321-7308 Jan, 2015
[8] A Study of Data Mining Techniques to Agriculture by Dr P Jaganathan (Professor & Head, PSNA College of Engineering & Technology, Dindigul, Tamilnadu, India), S.Vinothini (Associate Professor, PSNA College of Engineering & Technology, Dindigul, Tamilnadu, India) & P.Backialakshmi (Assistant Professor, PSNA College of Engineering & Technology, Dindigul, Tamilnadu, India) IJRIT ISSN 2001-5569 April, 2014
[9] Predictive ability of machine learning methods for massive crop yield prediction by Gonzalez-Sanchez Alberto, Frausto-Solis Juan, Ojeda-Bustamante W. Span J Agric Res. 2014;12(2):313–28.
[10] MotiurRahman M, Haq N, Rahman RM. Application of data mining tools for rice yield prediction on clustered regions of Bangladesh. IEEE. 2014;2014:8–13.
[11] Application of Spatial Data Mining for Agriculture by D.Rajesh (AP-SITE, VIT University, Vellore-14) IJCA(0975-8887) Vol.15, No. 2 Feb, 2011
[12] Lloyd, S.P. (1957). "Least square quantization in PCM". Bell Telephone Laboratories Paper Published in journal much later. IEEE transactions on Information Theory. 28 (2): 129-137. DOI:10.1109/TIT.1982.1056489. Retreived 2009-04-15
[13] Kaufman, L., Rousseeuw, P. J., Clustering by Means of Medoids.
[14] Ester, Martin; Kriegel, Hans-Peter; Sander, Jörg; Xu, Xiaowei (1996). Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M., eds. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI PP.226-231. CiteSeerX 10.1.1.121.9220. ISBN 1-57735-004-9.

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)