

An Impression of Various Data Mining Techniques in the Field of Agriculture

K. Priya¹ A. Anbarasi²

¹Ph.D. Research Scholar ²Assistant Professor

^{1,2}Department of Computer Science

¹Bharathiyar University, Tamil Nadu, India ²AVP College of Arts & Science, Tamil Nadu, India

Abstract— India is a nation where agriculture and agriculture-related trades are the most important sources of living for the general population. Agriculture is a major source of economy of the country. Utilization of data innovation in agriculture can change the circumstance of decision making and agriculturists can yield in a better way. Agriculture Data Mining (ADM) is an interdisciplinary ingenious research area that handles the development of methods to explore data arising in agriculture fields. A computational approach used by ADM is to examine farming data in order to study cultivation questions. As a result, it provides intrinsic knowledge of farming and crop growing process for effective agriculture planning. Data mining participates a vital part in decision making on numerous issues related to the agriculture pasture. This paper discussed the role of data mining in perception of agriculture field and as well confers regarding some data mining techniques and their related work by several authors in circumstance to agriculture domain. It also discusses various data mining applications in solving many agricultural problems.

Key words: Agriculture, Classification, Clustering, Data Mining, Regression

I. INTRODUCTION

Agriculture is the backbone of the Indian economic system. In India, the greater parts of farmers are not receiving predictable crop return appropriate to numerous reasons. Farming harvest principally trusts on climate conditions. The rainfall situation, in addition, influences the rice farming [1]. In this background, the farmers unavoidably require judicious guidance to envisage future gather efficiency plus investigation is to be completed in organizing to help farmers make the best use of the crop manufacture in their crops. Agriculture research is rapidly rising, due to the progression of technologies as well as upcoming challenges. It has been established to be a leading role in improving the largely growth rate of any country [2].

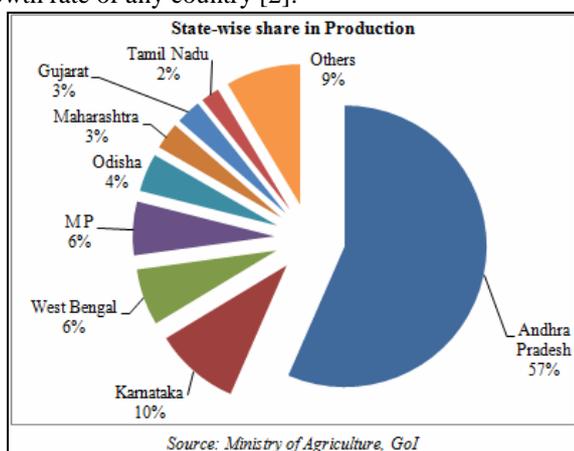


Fig. 1: Major Crop Producing States

To develop the growth rate of agriculture, researchers of this domain utilized different data mining techniques to resolve agriculture related difficulties. Currently, the minority farmers are using the different methods, tools plus technique of farming for fine production. Data mining can be used for forecasting the prospect values of agricultural processes. Some of the agricultural facts in India are listed in the following table [3].

S. No	Products	States
1	The largest producer of rice in India	West Bengal, Uttar Pradesh, Andhra Pradesh and Punjab
2	The largest Wheat producing state of India	Uttar Pradesh, Punjab, Haryana and Madhya Pradesh
3	India is the largest producer of Cotton and groundnut	Gujarat, Maharashtra, Telangana and Karnataka
4	The largest producer of Sugarcane in India	Uttar Pradesh, Maharashtra, Karnataka and Tamilnadu
5	The largest producer of fresh Vegetables (ginger, okra and second of potatoes, onions, cauliflowers, brinjal and Cabbages) in India	West Bengal, Uttar Pradesh, Bihar and Madhya Pradesh
6	The largest producer of Jute in India	West Bengal, Bihar, Assam and Andhra Pradesh.
7	The largest producer of Tea in India,	Assam, West Bengal, Himachal Pradesh and other states of North India.
8	The largest producer of Coffee in India	Karnataka, Kerala, Tamil Nadu and Andhra Pradesh
9	The largest pulse producing state in India	Madhya Pradesh, Uttar Pradesh, Maharashtra and Rajasthan
10	The largest producer of Rubber in India	Kerala, Tamil Nadu, North East state of Tripura and Karnataka

Table 1: Largest Agricultural Producing States of India

Information technology can be used to avert the risk associated with agriculture and it can also be used to predict the crop yield more accurately prior to harvest. Yield prediction needs different kinds of data gathered from different sources like meteorological data, agri-meteorological, soil data, remotely sensed data, agricultural statistics etc [4]. To handle such huge data, the best option is Data Mining. Use of Data Mining in agriculture can change the scenario of decision making and farmers can yield in a better way. This paper covers some very recent applications of data mining techniques in the field of agriculture. Data mining is a method by which one can extract the knowledge from the huge bulk of data.

II. LITERATURE REVIEW

There are number of researches which have been carried out on the application of data mining techniques for agricultural data sets. The agricultural information system will affiliate farmer for data availability on a customer farmer portal with crops details and its users and researches can get online information about the crops.

S.Veenadhari, et al., [5] observed the research studies on application of data mining techniques in the field of agriculture.

A portion of the methods, for example, ANN, ID3, the k- means, the k-NN and SVM connected in the field of agriculture were introduced. Data mining in the application in agriculture is a moderately new methodology for anticipating or foreseeing of a rural harvest or creature the executives. This article investigates the uses of data mining strategies in the field of agriculture and allied sciences. The supply chain operation of companies engaged in industries that use agricultural produce as raw material is important for Historical crop yield information. Animal feed, seed, chemical, poultry, fertilizer pesticides, seed, paper and many other industries use agricultural products as intergradient in their production processes. An accurate estimate of crop size and risk helps these companies in planning supply chain decision like production scheduling. Business such as seed, fertilizer, agrochemical and agricultural machinery industries plan production and marketing activities based on crop production estimates.

Vrushali Bhuyar focuses on different classifier algorithms used on soil dataset to predict fertility rate. Study shows that among the classifier [6] J48 classifier perform better to predict fertility index.

Dr. D. Ashok Kumar and N. Kannathasan focuses on different data mining techniques we can use in agriculture. Their research survey recommending that a comparison of different data mining techniques could produce an efficient algorithm for soil classification for multiple classes. The benefits of a greater understanding of soils could improve productivity in farming, maintain biodiversity, reduce reliance on fertilizers and create a better integrated soil management system for [7] both the private and public sectors.

S.Gokila et al. presented an approach based on clustering and classification rule in support of climatology to mine weather data. The aim of this paper is to provide a review report on various popular data mining techniques

applied on weather data set in support of weather prediction and climate analyzers. In this research work, clustering techniques applied on the climate data helps to produce similar patterns of climate with the consideration of spatial nature. The classification techniques are used to relate the attributes of weather data to predict the future climate [8].

Dingsheng, Yaming, Nan and Yufeng [9] proposed the annual average exciting rainfall prediction model based on BP network combined with stepwise discriminant method and use Bayesian statistical method to further improve the network's generalization ability and model prediction accuracy, but the overall performance of the technique can be further improved, such as to further improve the correct ratio on discriminant analysis. The experimental results validate the method and the prediction accuracy is satisfactory.

Urtubia et al., The prediction of wine fermentation problems can be performed by using a k-means approach. Knowing in advance that the wine fermentation process could get jammed or be slow can help the enologist to correct it and ensure a good fermentation process [10].

Markl, et al, introduces the idea of using Multidimensional hierarchical clustering (MHC) of OLAP to speed up the aggregation queries and improving the performance and cost measures in data warehousing. An encoding scheme has been defined for hierarchical dimensions that enable clustering of data with respect to multiple hierarchical dimensions [11].

III. NEED OF DATA MINING IN AGRICULTURE

Data Mining is defined simply as mining of data from a data warehouse or some number of databases. It can also be considered as a new evolution in information technology. Data Mining played a very important role in the field of agriculture. Agriculture is purely dependent on some features like soil, climate, crop yield, pest types, pesticides etc. By using data mining, many issues related to the soil like soil classification and soil prediction and for crop yield also the techniques are used and analysis is done. This data mining is used efficiently in the field of agriculture and always been on the same side for those who rely on these techniques [12]. Many decisions like which crop to be cultivated in which season, which soil to be considered for a particular crop etc. are taken by only considering the results obtained using data mining techniques. Data mining techniques are frequently used to study soil properties, crop yield properties, climatic change conditions etc.

Agriculture is rich in data but poor in information, meaning that agricultural centers collect and store many facts, but rarely translate them into meaningful information. Agriculture is an information-intensive manufacturing from an important point of vision. Many factors such as soil, fertilizer, temperature, precipitation, sunray, etc. affect harvest, so that information about them has to be carefully investigated by expert persons in deciding agricultural activities. Many researchers designed a data warehouse to assist the experts and to help improve agricultural technologies. Although some statistical tools are already applied to Agricultural data for getting knowledge, it has been observed that data mining tools and techniques are not much applied for agricultural data analysis. Clustering of large data

sets like soil datasets remains a computationally intensive and challenging task [13].

A. Data Mining

Data Mining is the process of extracting useful and important information from large sets of data. It is the process of discovering meaningful new correlation, patterns, and trends by sifting through a large amount of data using Pattern recognition technology as well as statistical and mathematical techniques. Data mining and knowledge discovery in database (KDD) are concerned with extracting patterns and models of interest from huge databases. KDD says that “knowledge” is the necessary end product of given data-driven discovery. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. The patterns, associations, or relationships among all this data can provide information [14]. This section describes an overview of Data Mining techniques applied to agricultural and their applications to agricultural related areas. Data mining techniques are mainly divided into four groups [15]:

- Classification

Classification techniques are intended to categorize unknown samples using information provided by a set of classified samples. This set is usually referred to as the training set it is used to train the classification technique to perform its classification.

- Clustering

In case a training set is not there, then there is no knowledge about the data to categorize. In such cases, the clustering technique can be used to split a set of unknown samples into clusters.

- Association Rules

Association rule mining technique is one of the most efficient techniques of data mining to search unseen or desired pattern among the vast amount of data. In this method, the focus is on finding relationships between the different items in a transactional database.

- Regression

Regression is learning a function that maps a data item to a real-valued prediction variable.

The graphical representation of different data mining techniques is shown in the following figure,

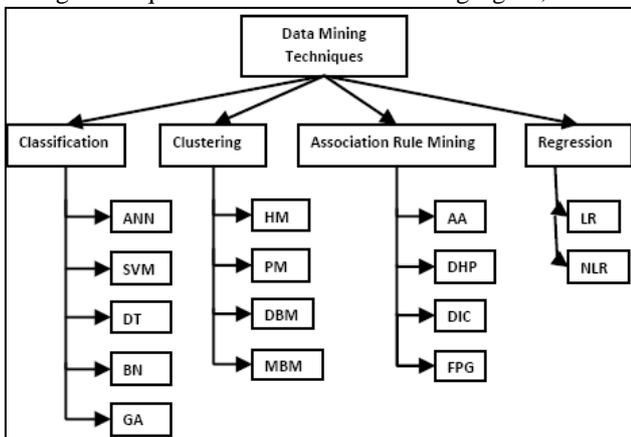


Fig. 1: Various Data Mining Techniques

1) General Methodology

For the agriculture system, better planning and decisions can be chalked out to achieve the conclusion

a) Input

Most of the research papers that were studied have considered some climatic parameters like temperature, humidity, rainfall. Some agronomical parameters like soil, nutrient contents, and pesticides etc. The values of these variables have been taken as input.

b) Preprocessing (Noise Removal)

For the successful application of data mining, a huge set of the dataset is required. The data which is acquired from various resources are sometimes in raw form. It may contain some incomplete, redundant, inconsistent data. Therefore in this step, such redundant data should be filtered. Data should be normalized.

c) Feature Extraction (Attribute Selection)

This step aims at identifying and using the most relevant attribute from the dataset. Through this process, irrelevant and redundant information is removed for the application of classifiers [16].

B. Agriculture

Agriculture is that the cultivation of animals, plants, fungi, and different life forms for food, fiber, and biofuel, healthful and different merchandise accustomed to sustain and enhance human life. Agriculture was the key growth within the rise of stationary human civilization, whereby farming of domesticated species created food surpluses that nurtured the event of civilization. The study of agriculture is understood as agricultural science. The history of agriculture dates back thousands of years, and its development has been driven and outlined by greatly completely different climates, cultures, and technologies. Within the civilized world, industrial agriculture supported large-scale monoculture farming has become the dominant agricultural methodology.

In the agriculture sector, achieving maximum crop yield at minimum cost is a goal of production. The process of taking a decision is complex as there are several factors affecting the entire farming process. Recent data mining technologies use different units that are able to offer lots of data on agricultural related activities, which might be analyzed so as to seek out necessary data [17].

IV. APPLICATION OF DATA MINING TECHNOLOGY IN AGRICULTURE

Agriculture is a complex giant system. India is a vast country with multifarious soil types and abundant crop varieties. Conversely, the pests as well as disease arise regularly, and the relationship among fertilizer, water, density, and climate needs to be studied. Agricultural information data have the properties of a large number, various dimensions, dynamic degree, incompleteness and uncertainty [18]. The following table covers some applications of data mining techniques in the field of agriculture.

S.NO.	Author	Methodology (Data mining)	Application
1		Artificial Neural	To predict flowering and physiological

	D.A. Elizondo et al.,	Networks Models	maturity for soybean
2	Glaucia M. Bressan et al.,	Bayesian network	To infer the risk of weed infestation in a corn-crop
3	Shalvi D et al.,	K-nearest Neighbor	Simulating daily precipitations and other weather conditions
4	I. Jagielska et al.,	K-means Clustering algorithm	Classifying soil in combination with GPS
5	MA. Shahin et al.,	Neural network and fuzzy logic classifiers	Sorting Apples based on Water core
6	Umid Kumar Dey et al.,	Modified Nonlinear Regression	Rice Yield Prediction
7	Milos Ilic et al.,	Artificial Neural Networks	Early fruit diseases detection
8	Umair Ayub et al.,	Ensemble Classification (Neural Networks)	Predicting Crop Diseases
9	HaiYue Jin et al.,	Decision-making tree algorithm	Oral cavity disease diagnosis
10	Dai Xiaopeng et al.,	Case-Based Reasoning (CBR)	Predicting middle or short term occurrence of rice disease and pest

Table 2: Application of Data Mining Technology in Agriculture

V. CONCLUSION

This paper presents a quick update with respect to the state-of-the-art in the field of data mining and agriculture. There are a growing number of applications of data mining techniques in agriculture as well as a lot of data that are presently accessible from many resources. This is comparatively a novel research field and it is expected to mature in the future. There is a lot of work to be done on this rising and interesting research field. The multidisciplinary approach of integrating computer science with agriculture will help in forecasting/managing agricultural crops effectively and to increase the income of the farmer. This paper is an attempt to provide an overview of some previous researches and studies of data mining and data mining technique in agriculture. This paper concludes that the agricultural mining technique becomes highly active research area in data mining research.

VI. REFERENCE

- [1] Yogesh Gandge, Sandhya "A Study on Various Data Mining Techniques for Crop Yield Prediction" 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECOT)
- [2] N. G. Yethiraj, "Applying Data Mining Techniques in the field of agriculture and allied sciences" International Journal of Business Intelligentsvol. 1, no. 2, December 2012.
- [3] <http://www.walkthroughindia.com/offbeat/10-largest-agricultural-producing-states-of-india/>
- [4] Raoranne and R. V. Kulkarni, "Data Mining: An effective tool for estimation in the agricultural sector," International Journal of Emerging Trends and Technology in Computer Science, vol. 1, no. 2, pp. 75-79, July-August 2012.
- [5] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, "Data mining Techniques for Predicting Crop Productivity – A review article", International Journal of Computer Science and technology, march 2011.
- [6] VrushaliBhuyar. Comparative Analysis of Classification C2Techniques on Soil Data to Predict Fertility Rate for Aurangabad District. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com, Volume 3, Issue 2, March – April 2014 ISSN 2278-6856.
- [7] Dr. D. Ashok Kumar, N. Kannathasan. A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining. IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, May 2011,ISSN: 1694-0814.
- [8] S.Gokila, K.Anand Kumar, A.Bharathi, "Clustering and classification in support of climatology to mine weather data-a review", International conference computing and intelligence systems, ISSN: 2278-2397, Volume: 04, Pages: 1336-1340, March 2015.
- [9] Dingsheng Wan, Yaming Wang, Nan Gu and Yufeng Yu, "A Novel Approach to Extreme Rainfall Prediction Based on Data Mining", 2nd International Conference on Computer Science and Network Technology, 2012.
- [10] Urtubia, A., Pérez-Correa, J. R., Soto, A., & Pszczolkowski, P. (2007, "Using data mining techniques to predict industrial wine problem fermentations", Food Control, 18(12), 1512-1517.
- [11] Markl et al (1999): Improving OLAP Performance by Multidimensional Hierarchical Clustering. Bayerisches Forschungszentrum für Wissensbasierte Systeme Orleansstraße 34, 81667 München, Germany.
- [12] N.Neelaveni, Ms. S. Rajeswari "DATA MINING IN AGRICULTURE- A Survey" International Journal of Modern Computer Science (IJMCS) ISSN: 2320-7868 (Online), Volume 4, Issue 4, August, 2016.
- [13] Jin HaiYue, Song Kai "IBLE Algorithm in agricultural disease diagnosis" 2010,IEEE, Third International Conference on Intelligent Networks and Intelligent Systems

- [14] Umair Ayub, Syed Atif Moqurrab "Predicting Crop Diseases Using Data Mining Approaches: Classification" 978-1-5386-5482-8/18/2018 IEEE
- [15] M.C.S.Geetha "A Survey on Data Mining Techniques in Agriculture" International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 2, February 2015.
- [16] R. Sujatha, Dr. P.Isakki, A Study on Crop Yield Forecasting Using Classification Techniques, 978-1-4673-8437-7/16/\$31.00 ©2016 IEEE.
- [17] Dai Xiaopeng, Li Donghui "Research on Rice Pests and Diseases Warning Based on CBR" 2013, IEEE International Conference on Computational and Information Sciences.
- [18] George RuB, "Data Mining of Agricultural Yield Data: A comparison of Regression models.

