

Agriculture Decision Support Using Data Mining Algorithms-A Survey

Krishna Janardh

M.Phil Computer Science, Nehru College of Arts and Science Coimbatore, India.

Dr. D Vimal kumar

Assistant professor Nehru College of Arts and Science Coimbatore, India.

Abstract – Data mining is the backbone of almost all applications in the recent trend. Every activity can be handled and analyzed by data mining techniques. Data mining have the ability to handle many uncertain and heterogeneous data. This paper gives a brief survey on the data mining techniques and approaches used in farming activities. This simply poses different types of clustering techniques and classification techniques. The clustering techniques such as K-Means, Fuzzy, KNN, and classification techniques such as Bayesian, Artificial Neural network, SVM and Decision Tree based approaches are compared in this paper.

Index Terms – Data mining, decision tree, classification, clustering.

1. INTRODUCTION

Data mining is the method of retrieving useful information and patterns from massive dataset [1]. Especially in agriculture the dataset size are huge and dynamic in nature and hard to predict based on statistics. In the context of agriculture field, many data mining techniques are proposed. The most popular and well used data mining techniques in the field of agriculture are: statistics based analysis, prediction and classification. With the enormous amount of agricultural data collected from various sensors and components, files and other sources, it is gradually more important that the technique is necessary to be developed with more powerful features for analysis, interpretation and decision process. This process refers to the nontrivial extraction of hidden, previously unknown and potentially useful information from data in databases. Agriculture Data mining is the extraction of unknown and hidden predictive information from hefty and dynamic databases [2]. It is considered as a powerful and new technology with great potential to focus on the most important information in the agricultural data sources. This paper surveys about the various techniques and tools available in data mining to manage agricultural activities.

2. DATA MINING TECHNIQUES

Clustering: This model is used to divide the database into subsets, or Clusters, based on a set of attributes. For example, in the process of understanding its customer base, an organization may attempt to divide the known population to

discover clusters of potential customers based on attributes never before used for this kind of analysis (for example, the type of plants, the total crop values per year, and so on). Clusters can be created either statistically or by using artificial intelligence methods. Clusters can be analyzed automatically by a program or by using visualization techniques.

Classification: In data mining, classification is the supervised learning process, which is used to classify database records into a number of predefined classes based on training samples. In agriculture dataset, a soil details or the crop details can be classified based on its features. It contains several iterations and feature classes. Based on the training sample this classifies the agricultural records.

Prediction: This model predicts the value for a specific attribute of the agricultural data set. For example, given a predictive model of harvest level data, predict the likelihood that a specific crop is profitable or the current weather can affect the plant or not. Prediction may also be used to validate a discovered hypothesis. This type of prediction analysis is mostly carried out in the weather analysis in the agricultural activities.

A. Applications of data mining in agriculture

There are several applications areas are studied in the literature under agriculture using data mining. The followings are the applications of data mining in various process of agriculture.

i. Weather analysis and agriculture activity prediction:

Weather forecasting is the process of collecting data on special conditions, which able to detect the humidity, temperature, wind direction, wind speed and rainfall etc. These types of Weather attributes are collected from meteorological satellites and radars. These data's will be stored and compared in future for different decisions in agriculture. To make an accurate forecast statistical data should be compared. This data's are dynamic and unique to location to location.

In [3] authors discussed about weather prediction techniques and finally proposed a new forecasting algorithm using back propagation algorithm, in general back propagation algorithm

is a supervised learning method, which falls under classification techniques. For this algorithm, users need to provide necessary training samples with class labels. This contains two phases one is propagation and another one is weight update process. The algorithm repeats the above two steps until the satisfactory result found.

In [4] authors proposed a new data mining technique to predict the temperature and its changes in seasonal aspects. This also shows how the seasonal changes affect the yield of crops.

In [5] author surveyed and analyzed Influence of climatic factors on major kharif and rabi crops production in Bhopal District of Madhya Pradesh State. This paper utilized decision tree concept to find the climatic factors and its productivity result. The result of this study shows that the decision tree analysis specified that the production and crop of soybean was mostly influenced by the detected weather condition, especially the humidity, temperature and rainfall has been detected sequentially. In this paper, the result concludes that the productivity of paddy crop was mostly depends on the Rainfall followed by proportional Fading and humidity values. And the paper concludes the Wheat crop productivity is mostly influenced by Temperature followed by relative humidity and rainfall. In the outcome of the decision tree based analysis, the rules are created are useful for identifying the conditions intended for high or low crop productivity.

ii. Crop patterns using data mining:

In literature, data mining classification technique to improve predictive accuracy. This approach segments the databases into several identical groups and then the data of each group can be analyzed, investigated and classified. This paper is unsuccessful because the proposed classification algorithms require that the classes be defined based on data attributes values and parameters.

iii. Soil classification:

Data Mining is important to discover the soil fertility, crop prediction and soil type detection. A few researches made on the soil related crop prediction and classification. This helps to detect the soil features and quality with necessary suggestion and crop analysis.

The paper [6] utilized Classification algorithms involve finding rules that partition the data into disjoint clusters, which is a set of classification rules. This helps to classify the future values from the historical records. This paper performed a comparative study on classification algorithms such as Naive Bayesian classifier and J48 decision tree classifier etc.

In paper [7] authors used Bayesian network, it is a powerful tool. And this algorithm broadly used in agriculture datasets. Bayesian approach improves hydro geological site characterization even when using low resolution resistivity data logs from the soil related agricultural data.

3. CONCLUSION

The agricultural activity analysis plays an important role in current trend. Detection and analysis of agricultural activity is the most important issue in real time scenario, because the lack of training samples and sufficient data's make these processes much complicated.

There are several different methods to active crop and yield detection. This survey presents a various techniques of the data mining approach to solve the agricultural data analysis problem. From the analysis we discover an optimal and effective algorithm and technique to handle agricultural datasets.

REFERENCES

- [1]. Thamaraiselvi, G., and A. Kaliammal. "Data mining: concepts and techniques." *SRELS Journal of Information Management* 41, no. 4 (2004): 339-348.
- [2]. Mucherino, Antonio, Petraq Papajorgji, and Panos M. Pardalos. *Data mining in agriculture*. Vol. 34. Springer Science & Business Media, 2009.
- [3]. Badhiye, S. S., P. N. Chatur, and B. V. Wakode. "Temperature and humidity data analysis for future value prediction using clustering technique: an approach." *International Journal of Emerging Technology and Advanced Engineering* 2, no. 1 (2012): 88-91..
- [4]. Jan, Zahoor, Muhammad Abrar, Shariq Bashir, and Anwar M. Mirza. "Seasonal to Inter-annual Climate Prediction Using Data Mining KNN Technique." In *IMTIC*, pp. 40-51. 2008.
- [5]. Geetha, M. "A survey on data mining techniques in Agriculture." *International journal of innovative research in computer and communication engineering* 3, no. 2 (2015): 887-892.
- [6]. Ramesh, V., and K. Ramar. "Classification of agricultural land soils: a data mining approach." *Agricultural Journal* 6, no. 3 (2011): 82-86.
- [7]. Bhargavi, P., and S. Jyothi. "Applying naive bayes data mining technique for classification of agricultural land soils." *International journal of computer science and network security* 9, no. 8 (2009): 117-122.