



# **A Survey on Data Mining Techniques in Agriculture**

M.C.S.Geetha

Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India.

**ABSTRACT:** Agriculture is the most significant application area particularly in the developing countries like India. Use of information technology in agriculture can change the situation of decision making and farmers can yield in a better way. Data mining plays a crucial role for decision making on several issues related to agriculture field. This paper discussed about the role of data mining in perspective of agriculture field and also confers about several data mining techniques and their related work by several authors in context to agriculture domain. It also discusses on different data mining applications in solving the different agricultural problems. It integrates the work of various authors in one place so it is useful for researchers to get information of current scenario of data mining techniques and applications in context to agriculture field. This paper provides a survey of various data mining techniques used in agriculture which includes Artificial Neural Networks, K - nearest neighbor, Decision tree, Bayesian network, Fuzzy set, Support Vector Machine and K – means[1].

**KEYWORDS:** Agriculture, Data Mining, Artificial Neural Networks, K nearest neighbor, k means, Decision tree, Bayesian network, Support Vector Machine, Fuzzy set.

## **I. INTRODUCTION**

Agriculture is the backbone of the Indian nation. In spite of the fact that large areas in India have been brought under irrigation, only one-third of the cropped part is irrigated. The productivity of agriculture is very low. So as the demand of food is increasing, the researchers, farmers, agricultural scientists and government are trying to put extra effort and techniques for more production. And as a result, the agricultural data increases day by day. As the volume of data increases, it requires involuntary way for these data to be extracted when needed. Still today, a very few farmers are actually using the new methods, tools and technique of farming for better production. Data mining can be used for predicting the future trends of agricultural processes.

Data mining is the process that results in the discovery of new patterns in large data sets. The goal of the data mining process is to extract knowledge from an existing data set and transform it into a human understandable formation for advance use. It is the process of analyzing data from different perspectives and summarizing it into useful information. There is no restriction to the type of data that can be analyzed by data mining.

The data can be analyzed in a relational database, a data warehouse, a web server log or a simple text file. Analysis of data in effective way requires understanding of appropriate techniques of data mining. The intention of this paper is to give details about different data mining techniques in perspective of agriculture domain so researchers can get details about appropriate data mining techniques in context to their work area.

Data mining tasks can be classified into two categories: Descriptive data mining and Predictive data mining. Descriptive data mining tasks characterize the general properties of the data in the database while predictive data mining is used to predict explicit values based on patterns determined from known results. Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest. As far as data mining technique is concern; in the most of cases predictive data mining approach is used. Predictive data mining technique is used to predict future crop, weather forecasting, pesticides and fertilizers to be used, revenue to be generated and so on.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

The yield prediction problem can be solved by employing Data Mining techniques such as K Means, K nearest neighbor (KNN), Artificial Neural Network and support vector machine (SVM). Research paper aims at finding suitable data models that achieve a high precision and a high generality with respect to four parameters namely rainfall, year, production and area of sowing. For this purpose, different types of Data Mining techniques were evaluated on different data sets.[2]

The paper is organized as follows: Chapter 2 discusses the methods of Data Mining. Chapter 3 discusses the applications of data mining techniques used in agriculture domain. Chapter 4 discusses the conclusion.

## II. METHODS

The main techniques for data mining include Association rules, Classification, Clustering and Regression. The different data mining techniques used for solving different agricultural problem has been discussed [3]. The graphical representation of different data mining techniques is shown in figure 1.

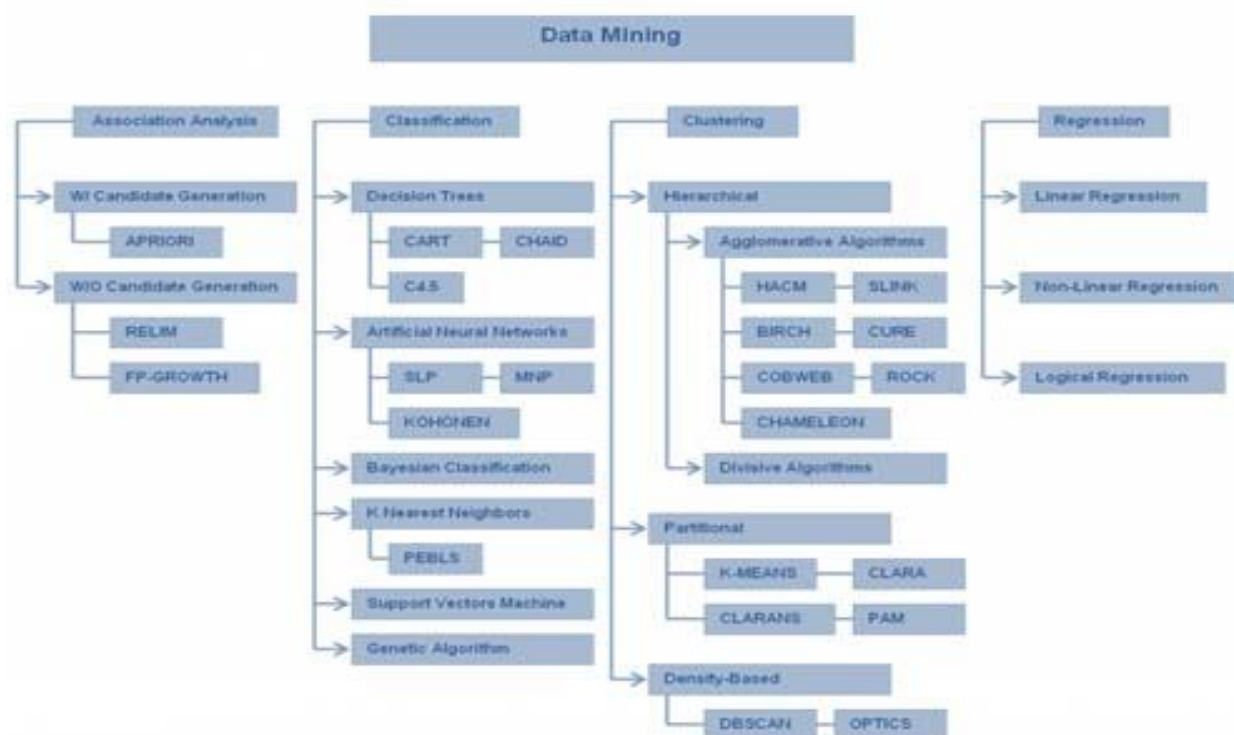


Figure 1: Data Mining Techniques

### 2.1 Association Rule Mining

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. Association rules are used to find out elements that co-occur repeatedly within a dataset consisting of many independent selections of elements (such as purchasing transactions), and to discover rules. The simple problem statement is: Given a set of transactions, where each transaction is a set of literals, an association rule is a phrase of the form  $X \Rightarrow Y$ , where  $X$  and  $Y$  are sets of objects. The instinctive meaning of such a rule is that transactions of the database which contain  $X$  tend to contain  $Y$ . [4] An application of the association rules mining is the market basket analysis, customer segmentation, store layout, catalog design, and telecommunication alarm prediction.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

The different association rule mining algorithm are Apriori Algorithm(AA), Partition, Dynamic Hashing and Pruning(DHP), Dynamic Itemset Counting(DIC), FP Growth(FPG), SEAR, Spear, Eclat & Declat, MaxEclat.[5]

## 2.2 Classification

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. It is a process in which a model learns to predict a class label from a set of training data which can then be used to predict discrete class labels on new samples. To maximize the predictive accuracy obtained by the classification model when classifying examples in the test set unseen during training is one of the major goals of classification algorithm. Data mining classification algorithms can follow three different learning approaches: semi-supervised learning, supervised learning and unsupervised learning. The different classification techniques for discovering knowledge are Rule Based Classifiers, Bayesian Networks(BN), Decision Tree (DT), Nearest Neighbour(NN), Artificial Neural Network(ANN), Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, Genetic Algorithms.[6]

## 2.3 Clustering

In clustering, the focus is on finding a partition of data records into clusters such that the points within each cluster are close to one another. Clustering groups the data instances into subsets in such a manner that similar instances are assembled together, while dissimilar instances belong to diverse groups. Since the aim of clustering is to find out a new set of categories, the latest groups are of interest in themselves, and their assessment is intrinsic. [7] There is no prior knowledge about data. The different clustering methods are Hierarchical Methods(HM), Partitioning Methods (PM), Density-based Methods(DBM), Model-based Clustering Methods(MBCM), Grid-based Methods and Soft-computing Methods [fuzzy, neural network based], Squared Error—Based Clustering (Vector Quantization), network data and Clustering graph [8]

## 2.4 Regression

Regression is learning a function that maps a data item to a real-valued prediction variable. The different applications of regression are predicting the amount of biomass present in a forest, estimating the probability of patient will survive or not on the set of his diagnostic tests, predicting consumer demand for a new product.[9] Here the model is trained to predict a continuous target. Regression tasks are often treated as classification tasks with quantitative class tag. The methods for prediction are Nonlinear Regression(NLR) and Linear Regression (LR).

## III. APPLICATION OF DATA MINING TECHNIQUES IN AGRICULTURE

### 3.1 Neural networks

**Sanjay D.Sawaitul et al.**, focuses the information about weather and are observed and stored. The recorded parameters are used to forecast weather. If there is a change in any one of the recorded parameters like wind speed, wind direction, temperature, rainfall, humidity, then the upcoming climatic condition can be predicted using artificial neural networks, back propagation techniques. The increase in signal range will work in large areas as well[10].

**Somvanshi, V.K. et al.**, deliberate the modeling and prediction of rainfall using artificial neural networks and Box-Jenkins methodology. Other applications of artificial neural networks in hydrology are forecasting daily water hassle and flow forecasting[11].

### 3.2 K-means

**K. Verheyen et al.**, Data Mining is the process of discovering meaningful patterns and trends by shifting through huge amount of data, using pattern detection technologies as well as statistical and mathematical techniques. Data Mining techniques are often used to studied soil characteristics. As an example, the K-Mean approach is used for classifying soils in combination with GPS based techniques[12].



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

**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[13].

### 3.3 Fuzzy set

**I. Jagielska et al.**, describe applications to agricultural related areas. Such as Yield prediction is a very important agricultural problem. Any farmer might be interested in knowing how much yield is expected. In the past, yield prediction was achieved by considering farmer's experience on particular field, crop and climate condition. We have discussed additional information about data like probability in probability theory, grade of membership in fuzzy set theory[14].

**Tellaeche et al.**, detecting weeds in precision agriculture. The paper summarize an automatic computer vision system for the detection and differential spraying of *Avena sterilis*, a toxic weed growing in cereal crops. With such purpose it have been designed a hybrid decision making system based on the Bayesian and Fuzzy k-Means classifiers, where the a priori probability required by the Bayes framework is supplied by the Fuzzy k-Means[15].

### 3.4 Decision tree and Bayesian classification

**Veenadhari, S.** Influence of climatic factors on major kharif and rabi crops production in Bhopal District of Madhya Pradesh State was considered. The findings of the study revealed that the decision tree analysis indicated that the productivity of soybean crop was mostly influenced by comparative humidity followed by temperature and rainfall. The decision tree analysis shows that the productivity of paddy crop was mostly inclined by Rainfall followed by comparative Evaporation and humidity. For Wheat crop, the analysis shows that the productivity is mostly influenced by Temperature followed by relative humidity and rainfall. The result of decision tree were confirmed from Bayesian classification. The rules formed from the decision tree are useful for identifying the conditions intended for high or low crop productivity[16].

**Shalvi D and De Claris N** Bayesian network is a powerful tool and broadly used in agriculture datasets. The model developed for agriculture application based on the Bayesian network learning method. The results show that Bayesian Networks are feasible and efficient. Bayesian approach improves hydro geological site characterization even when using low-resolution resistivity surveys[17].

### 3.5 K-nearest neighbour

**Altannar Chinchulunn et al.**,The k-nearest neighbor classification algorithmic rule may be divided into 2 phases: coaching section and testing section. Bermejo associated Cabestany urged a reconciling learning algorithmic rule to permit fewer information points to be utilized in coaching information set. Several different techniques are projected to scale back procedure burden of k-nearest neighbor algorithms[18].

**B. Rajagopalan and U. Lal** A number of studies have been carried out on the application of data mining techniques for agricultural data sets. For example, the K-Nearest Neighbor is applied for simulating daily precipitations and other weather variables[19].

### 3.6 Support Vector Machine

**S.Veenadhari et al.**,The main plan of Support Vector Machine (SVM) is to classify information samples into 2 disjoint categories. The essential plan behind is classifying the sample information into linearly severable. Support Vector Machine (SVM) area unit a group of connected supervised learning ways used for classification and regression[20].

**Tripathi, S et al.**, The SVM-based DM is applied to future climate predictions from the second generation Coupled Global Climate Model (CGCM2) to obtain future projections of precipitation. The results are then analyzed to assess the crash of climate change on rainfall over India. It is shown that SVMs provide a promising alternative to

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

conventional artificial neural networks for statistical downscaling, and are appropriate for conducting climate impact studies[21]. Table1 shows the data mining methodologies used in agriculture

**Table 1: Data mining methodologies used in agriculture**

Author	Title	Data mining methodologies	Applications
Sanjay D. Sawaitul, Prof. K.P. Wagh, Dr. P.N. Chatur	Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach	Neural Networks	Focuses on weather forecasts
V. K. Somvanshi, et al.,	Modeling and prediction of rainfall using artificial neural network and ARIMA techniques	Neural Networks	Prediction of rainfall
I. Jagielska, C. Mathehews, T. Whitfort	An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems	K-means	Classifying soil in combination with GPS
Tellaeche, A., BurgosArtizzu, X. P., Pajares, G., & Ribeiro, A.	A vision-based hybrid classifier for weeds detection in precision agriculture through the Bayesian and Fuzzy k-Means paradigms	K-means	Wine fermentation problem
K. Verheyen, D. Adriaens, M. Hermy, and S. Deckers	High resolution continuous soil classification using morphological soil profile descriptions	Fuzzy set	Yield Prediction in agriculture
Urtubia, A., Pérez-Correa, J. R., Soto, A., & Pszczolkowski, P.	Using data mining techniques to predict industrial wine problem fermentations.	Fuzzy set	Detecting weeds in precision agriculture
S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh	Data mining Techniques for Predicting Crop Productivity	Decision Tree Analysis	Influence of climatic factors on major kharif and rabi crops production
Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S.	Downscaling of precipitation for climate change scenarios: a support vector machine approach	Bayesian network	Developed the model for agriculture purpose based on the Bayesian network learning method
Veenadhari, S.	Crop productivity mapping based on decision tree and Bayesian classification	K-nearest Neighbor	Scale back procedure burden of k-nearest neighbor algorithms
Shalvi D and De Claris N,	Unsupervised neural network approach to medical data mining techniques	K-nearest Neighbor	Simulating daily precipitations and other weather conditions
Altannar Chinchulunn, Petros Xanthopoulos, Vera Tomaino, P.M.Pardalos	Data Mining Techniques in Agricultural and Environmental Sciences	Support Vector Machine	Classifying the sample information into linearly severable
B. Rajagopalan and U. Lal	A K-nearest neighbor simulator for daily precipitation and other weather variable	Support Vector Machine	Appropriate for conducting climate impact studies





# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

## IV. CONCLUSION

Agriculture is the most significant application area particularly in the developing countries like India. Use of information technology in agriculture can change the situation of decision making and farmers can yield in better way. Data mining plays a crucial role for decision making on several issues related to agriculture field. It discusses about the role of data mining in the agriculture field and their related work by several authors in context to agriculture domain. It also discusses on different data mining applications in solving the different agricultural problems. This paper integrates the work of various authors in one place so it is useful for researchers to get information of current scenario of data mining techniques and applications in context to agriculture field.

## REFERENCES

1. Yethiraj N G , "Applying Data Mining Techniques In The Field Of Agriculture And Allied Sciences", International Journal of Business Intelligents ISSN: 2278-2400, Vol 01, Issue 02, December 2012.
2. Ramesh D, Vishnu Vardhan B., "Data Mining Techniques and Applications to Agricultural Yield Data", IJARCCCE, Vol. 2, Issue 9, September 2013.
3. Mucherino, A., Papajorgji, P., & Pardalos, P. (2009), "Data mining in agriculture" (Vol. 34), Springer.
4. Srikant, R V Q & Agrawal, R (1997, August), "Mining Association Rules with Item Constraints. In KDD" (Vol. 97, pp. 67-73).
5. Zaki, M J (1999), "Parallel and distributed association mining: A survey". IEEE concurrency, 7(4), 14-25.
6. Beniwal, S. & Arora, J. (2012), "Classification and feature selection techniques in data mining", International Journal of Engineering Research & Technology (IJERT), 1(6).
7. Lior Rokach, Oded Maimon, "Clustering Methods", Chap-15
8. Xu, R & Wunsch, D (2005), "Survey of clustering algorithms", Neural Networks, IEEE Transactions on, 16(3), 645-678.
9. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996), "From data mining to knowledge discovery in databases". AI magazine, 17(3), 37.
10. Sanjay D. Sawaitul, Prof. K.P. Wagh, Dr. P.N. Chatur, "Classification and Prediction of Future Weather by using Back Propagation Algorithm- An Approach", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, Issue 1, January 2012, pp. 110-113.
11. K. SOMVANSHI, ET AL., "MODELING AND PREDICTION OF RAINFALL USING ARTIFICIAL NEURAL NETWORK AND ARIMA TECHNIQUES", J. IND. GEOPHYS. UNION, VOL. 10, NO. 2, PP. 141-151, 2006.
12. K. Verheyen, D. Adriaens, M. Hermy, and S. Deckers, "High resolution continuous soil classification using morphological soil profile descriptions", *Geoderma*, vol. 101, pp. 31-48, 2001.
13. 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.
14. I. Jagielska, C. Mattheews, T. Whitfort, "An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems", *Neurocomputing*, Vol. 24, pp. 37-54, 1999.
15. Tellaeche, A., BurgosArtizzu, X. P., Pajares, G., & Ribeiro, A. (2007), "A vision-based hybrid classifier for weeds detection in precision agriculture through the Bayesian and Fuzzy k-Means paradigms", In *Innovations in Hybrid Intelligent Systems* (pp. 72-79). Springer Berlin Heidelberg.
16. Veenadhari, S. 2007, "Crop productivity mapping based on decision tree and Bayesian classification". Unpublished M.Tech Thesis submitted to Makhnallal Chaturvedi National University of Journalism and Communication, Bhopal.
17. Shalvi D and De Claris N., "Unsupervised neural network approach to medical data mining techniques", in *Proceedings of IEEE International Joint Conference on Neural Networks*, (Alaska), pp. 171-176, May 1998.
18. Altannar Chinchulunn, Petros Xanthopoulos, Vera Tomaino, P.M.Pardalos, "Data Mining Techniques in Agricultural and Environmental Sciences", *International Journal of Agricultural and Environmental Information Systems*, 1(1), 26-40, January-June 2010.
19. B. Rajagopalan and U. Lal, "A K-nearest neighbor simulator for daily precipitation and other weather variable", *Water Resources*, vol. 35, pp. 3089-3101, 1999.
20. 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 IJCST* Vol. 2, Issue 1, March 2011.
21. Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. (2006), "Downscaling of precipitation for climate change scenarios: a support vector machine approach", *Journal of Hydrology*, 330(3), 621-640.

## BIOGRAPHY

**M.C.S.Geetha** is an Assistant Professor in the Department of Computer Applications, Kumaraguru College of Technology, Coimbatore. She received Master of Computer Applications (MCA) degree in 2004 from P.S.G.R.Krishnnammal College for Women, Coimbatore, India. She received M.Phil (Computer Science) in 2006 from Bharathiyar University. Her research interest is Data Mining.