IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 23-25
www.iosrjournals.org

A Survey on Agriculture Analysis for Crop Yield Prediction Using Data Mining Techniques

Dr.A.Senthil Kumar, P.Arun
Assistant Professor, Dept of CS, Sankara College of Science and Commerce
M.Phil., Part-Time Research Scholar, Sankara College Of Science And Commerce

Abstract: *Now-a-day's agriculture is one of the most important field in the emerging real world and it is the main occupation and backbone of our country. Agriculture is in poor condition since before comparing previous years. The main reasons for this is without a well formed pattern about farming and proper guidance to the farmers. Due to these problems, farming affects the yield of crop and unawareness about the crop cultivation methodologies. And also season to cultivate the crop and choosing which soil is the best to cultivate the particular crop based on the weather condition and also when to harvest the crop for the best yield.*
If the farmer is aware about the crop cultivation methodologies and harvesting it will more helpful for the people in the real world and also to maximize the crop productivity. Data mining is the process of finding new template from large data sets, this technology which is employed in inferring useful knowledge that can be put to use from a vast amount of data, various data mining techniques such as classification, prediction, clustering and outlier analysis can be used for the purpose. Climate is one of the meteorological data that is rich by important knowledge.
This paper presents a brief comparative study of various papers that deal with various techniques used to figure out the crop yield. Different data mining techniques that are in use for the crop yield estimation are K-Means, K-Nearest neighbor (KNN), Artificial Neural Networks(ANN). Recently data processing techniques in agriculture required decide, storage, monitored and retrieval the resources used. This survey aims to search out appropriate data processing models to realize high accuracy and prediction capabilities.
Keywords: *Crop yield, Data mining, K-Means, K-Nearest Neighbor(KNN), Artificial Neural Networks(ANN).*

I. Introduction

Agriculture is the backbone of the Indian economy. 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. Data mining can be used for predicting the future trends of agricultural processes.

Data mining is the process to discover interesting knowledge from large amounts of data. 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. We can analyze data contained 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 technique 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.

II. Data Mining Techniques

In this paper we have used the data mining techniques include Classification and Clustering to predict the exact pattern for the best crop yield prediction. With the help of these data mining techniques farmers can grow the crop yield.

2.1 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: supervised learning, unsupervised learning, or semi-supervised 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.

2.2 K-Nearest Neighbor

The k-Nearest neighborhood methodology is wide used adopted thanks to its potency. The key plan of the algorithmic rule is to categorize a brand new sample within the most frequent class of its nearest neighbors within the coaching set. This is often the foremost selection formula on the category labels of its neighbors. 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.

2.3 Artificial Neural Network(ANN)

Artificial neural network is one of the new data mining techniques that are based on biological neural processes of human brain. According to this technique once the neural network is trained it can predict the crop yield in similar patterns even if the past data include some errors. Even if the data is complex, multivariate, nonlinear this network gives the accurate results and also without any of underlying principles the relationship between them the output is extracted. Artificial Neural Networks (ANN) networks during which every node represents a somatic cell and every link represents the method 2 somatic cell act. Every somatic cell performs straightforward tasks, whereas the network representing of the work of all its neurons is ready to perform the additional complicated tasks. A neural network is associate interconnected set of input/output units wherever every association includes a weight related to its. The process of classification by ANN may be broadly speaking outlined as follows:

- Run a sample from the coaching set, by giving its attribute values as input.
- The summation of weights and activation functions area unit applied at every
- node of hidden and output layers, till the output is generated.
- Compare output with the expected output from coaching set.
- If output doesn't match, return layer to layer and modify area unit weights and
- biases of nodes.
- Run consecutive sample and method a similar.
- Eventually the weights can coverage and method stops.

2.4 Clustering

The process of grouping a set of physical or abstract object into classes of similar objects is called clustering. Clustering is unsupervised technique used to group similar instances on the basis of feature. It has no labels required. Clustering does not require training data. Each cluster that is form can be viewed as a class of object, form which rule can be derived The different clustering methods are Hierarchical Methods(HM), Partitioning Methods (PM), Density-based Methods(DBM), Model-based Cluster Methods(MBCM), Grid-based Methods and Soft-computing Methods [fuzzy, neural network based], Squared Error-Based Clustering (Vector Quantization), network data and Clustering graph.

2.4.1 K-Means Approach

The most important clustering technique is K-Means clustering. This technique is used to classify the data which have no previous knowledge about the data or the training set. The parameter K denotes the amount of clusters required to partition the data.The idea of this clustering technique is, given K number of clusters we can define K centers, one for each cluster based on all samples belonging to a cluster. These centers must be placed far away from each other and then associate each sample to the cluster that has the closest centroid. When no samples are left, the process of finding new K centers and assigning samples to the clusters that has the closest centroid is iteratively carried out until no longer the samples can change their clusters. In the research article ,

the researcher states that using this K-Means approach the Government could help the agricultural firms to increase one of their production practices such as acquiring new farmers by framing new profitable agriculture schemes based on crop yield prediction and eventually campaigning to different groups of farmers about a particular scheme based on the result of the grouping of various crops depending on common features.

III. Conclusion

Agriculture is the most important application area particularly in the developing countries like India. Use of information technology in agriculture can change the scenario of decision making and farmers can yield in better way. For decision making on several issues related to agriculture field; data mining plays a vital role. In this paper we have discussed about the role of data mining in perspective of agriculture field. We have also discussed several data mining techniques and their related work by several authors in context to agriculture domain. This paper also focuses 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]. G. Nasrin Fathima, R.Geetha, "Agriculture Crop Pattern Using Data Mining Techniques" , "International Journal of Advanced Research in Computer Science and Software Engineering , Volume 4, Issue 5, May 2014 ISSN: 2277 128X".

[2]. Hetal Patel, Dharmendra Patel, "A Brief survey of Data Mining Techniques Applied to Agricultural Data", "International Journal of Computer Applications (0975 – 8887)Volume 95– No. 9, June 2014".

[3]. Ms.Kalpana,Dr.Shanthi,Dr.Arumugam, "A Survey on Data Mining Techniques in Agriculture", "International Journal of Advances in Computer Science and Technology, 3(8), August 2014, 426 – 431 426".

[4]. Ramesh A, Vijay. S. Rajpurohit, " A survey on Data Mining Techniques for Crop Yield Prediction", "International Journal of Advance Research in Volume 2, Issue 9, September 2014".

[5]. Surabhi Chouhan ,Divakar Singh, Anju Singh, "A Survey and Analysis of Various Agricultural Crops Classification Techniques", "International Journal of Computer Applications (0975 – 8887) Volume 136 – No.11, February 2016 25".

[6]. Perpetua Noronha, Divya .J, Shruthi .B.S, "Comparative Study of Data Mining Techniques in Crop Yield Prediction ", " International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Special Issue 2, October 2016 ".

[7]. Dhivya BH,, Manjula R, Siva Bharathi S, Madhumathi R, "A Survey on Crop Yield Prediction based on Agricultural Data", "International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization), Vol. 6, Issue 3, March 2017".

Biography

Dr. A.SENTHIL KUMAR MCA., M.Phil., MBA.,Ph.D., working as a Asst.Prof in Computer Science Department for Sankara college of Science and Commerce. He has teaching experience of 15 years in teaching at UG and PG levels. He has published papers in more than 10 International Journals and 6 National Journals. He has Presented more than 20 papers in International and National Conferences. His area of interest is Data Mining and Knowledge Engineering.

P.Arun, MCA., has Working and development experience of 2 years in IT. His area of interest is Data mining and Networking.