International Research Journal of Engineering and Management Studies (IRJEMS)
Volume: 03 Issue: 03 | March -2019 e-ISSN: 2395-0126

Data Mining Techniques Applied to Agricultural Data

Prof. K.R.Sarode[1], Dr. P.P.Chaudhari[2]

GCOE Aurgabad[1],
HOD GP Jalgaon[2].

Abstract - *Agriculture is one of the major revenue producing sectors of India and a source of survival. Agriculture contributes nearly 17% to total GDP of India and 10% of the total exports which helps in increasing foreign exchange. Over 70 % of the rural households depend on agriculture. Crop yield depends on multiple different factors such as rainfall, climate changes, soil type etc. With the advent of data mining, crop yield can be predicted by deriving useful insights from these agricultural data that aids farmers to decide on the crop they would like to plant for the forthcoming year leading to maximum profit. This paper presents a survey on the various data mining techniques used for crop yield prediction.*

Key Words: Agriculture, Data Mining, Crop yield .

1.INTRODUCTION

Agriculture is the backbone of Indian economy. Agriculture majorly contributes to the exports of India, directly improving foreign currency exchange. In India, majority of the farmers do not get expected yield due to several reasons. The agricultural yield primarily depends on environmental factors such as rainfall, temperature and geographical topology of the particular region. These factors along with some other influence the crop cultivation. In this context farmers require timely advices to predict the crop productivity and to predict this, an intensive analysis should be made in order to achieve desired results accurately. Yield is an important agricultural issue. Large amount of data can be gathered from Indian agriculture sector. Knowledge acquired from data is highly useful for many purposes. Data mining is a field in Information Technology that deals with finding unknown and hidden patterns from the available data. Applying data mining techniques in agricultural field to predict useful crop productivity related information is a noble work [1]. Data Mining is the process of extracting useful and important information from large sets of data. Data mining in agriculture field is a relatively novel research field. Yield prediction is a very important agricultural problem. Any farmer is interested in knowing how much yield he is about to expect. In the past, yield prediction was performed by considering farmer's experience on particular field and crop. In any of Data Mining procedures the training data is to be collected from historical data and the gathered data is used in terms of training which has to be exploited to learn how to classify future yield predictions [3].

2. DATA MINING TECHNIQUES

Data mining techniques are mainly divided in two groups, classification and clustering techniques. Classification techniques are designed for classifying unknown samples using information provided by a set of classified samples. This set is usually referred to as a training set as it is used to train the classification technique how to perform its classification. Generally, Neural Networks and Support Vector Machines, these two classification techniques learn from training set how to classify unknown samples [1]. Another classification technique, K- Nearest Neighbor, does not have any learning phase, because it uses the training set every time a classification must be performed. A training set is known, and it is used to classify samples of unknown classification. The basic assumption in the K-Nearest Neighbor algorithm is that similar samples should have similar classification. The parameter K shows the number of similar known samples used for assigning a classification to an unknown sample. The K-Nearest Neighbor uses the information in the training set, but it does not extract any rule for classifying the other [1].

In the event a training set not available, there is no previous knowledge about the data to classify. In this case, clustering techniques can be used to split a set of unknown samples into clusters. One of the most used clustering

technique is the K-Means algorithm . Given a set of data with unknown classification, the aim is to find a partition of the set in which similar data are grouped in the same cluster. The parameter K plays an important role as it specifies the number of clusters in which the data must be partitioned. The idea behind the K-Means algorithm is, given a certain partition of the data in K clusters, the centers of the clusters can be computed as the means of all samples belonging to clusters. The center of the cluster can be considered as the representative of the cluster, because the center is quite close to all samples in the cluster, and therefore it is similar to all of them. There are some disadvantages in using K-Means method. One of the disadvantages could be the choice of the parameter K. Another issue that needs attention is the computational cost of the algorithm. There are other Data Mining techniques statistical based techniques, such as Principle Component Analysis (PCA), Regression Model and Biclustering Techniques. Artificial neural network (ANN) is based on the human brain's biological neural processes. ANN learns to recognize the patterns or relationships in the data by observing a large number of input and output examples. Once the neural network has been trained, it can predict by detecting similar patterns in future data. They include the ability to learn and generalize from examples to produce meaningful solutions to problems even when input data contain errors or are incomplete, and to adapt solutions over time to compensate for changing circumstances and to process information rapidly. Furthermore, a system may be nonlinear and multivariate, and the variables involved may have complex interrelationships. ANNs are capable of adapting their complexity, and their accuracy increases as more and more input data are made available to them. They are capable of extracting the relationship between the input and output of a process without the any knowledge of the underlying principles.

The recent increased interest and use of neural models stems primarily from its nonlinear models that can be trained to map past and future values of the input-output relationship. This adds analytical value, since it can extract relationships between governing the data that was not obvious using other analytical tools. ANNs are also used because of its capability to recognize pattern and the speed of its techniques to accurately solve complex processes in many applications. They help to characterize relationships via a nonlinear, nonparametric inference technique; this is very rare and has many uses in a host of disciplines. The network offer the added advantage of being able to establish a 'training' phase, where example inputs are presented and the networks learn to extract the relevant information from these patterns. With this, the network can generalize results and lead to logical and other unforeseen conclusions through the model [11].

3. APPLICATION OF DATA MINING TECHNIQUES IN AGRICULTURE

There are number of studies which have been carried out on the application of data mining techniques for agricultural data sets. Naïve Bayes Data Mining Technique is used to classify soils that analyze large soil profile experimental datasets.[5] Decision tree algorithm in data mining is used for predicting soil fertility. [6] By using clustering techniques (Based on Partitioning Algorithms and Hierarchical Algorithms) author examines the current usage and details of agriculture land vanished in the past seven years. The overall aim of the research was to determine the land utilization for agriculture and non-agriculture areas for the past ten years.[7] D Ramesh [8] used k-means approach to estimate the crop yield analysis. Some data mining methodology which are used in agricultural domain are reviewed by author Vamanan, R, & Ramar, K [4] .The k-means algorithm is used for soil classifications using GPS-based technologies. [9] , The k-nearest algorithm is used in simulating daily precipitations and other weather variables [10] and Estimating soil water parameters and Climate forecasting [11].

4. CONCLUSIONS

With the improvement of data mining technologies, especially those without any premises or humans subjective, data mining can be applied in many areas. In this paper some data mining techniques were adopted in order to estimate crop yield analysis with existing data and their use in data mining. This paper presents new research

possibilities for the application of modern classification methodologies to the problem of yield prediction. There is a growing number of applications of data mining techniques in agriculture and a growing amount of data that are currently available from many resources.

REFERENCES

[1] Jiawei Han, Micheline K, Jian Pie, "Data Mining Concepts and Techniques", Morgan Kaufmann, ASIN B0058NBJ2M.

[2] Vinayak A. Bharadi, Prachi P. Abhyankar , Ravina S. Patil, Sonal S. Patade ,Tejaswini U. Nate, Anaya M. Joshi, "Analysis And Prediction In Agricultural Data Using Data Mining Techniques".

[3] D Ramesh, B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.

[4] D Ramesh, B Vishnu Vardhan, (2013). "Data Mining Techniques and Applications to Agricultural Yield", Data. International Journal of Advanced Research in Computer and Communication Engineering 2(9).

[5] Bhargavi, P, & Jyothi, S. (2009). "Applying Naive Bayes data mining technique for classification of agricultural land soils", International journal of computer science and network security, 9(8), 117-122.

[6] Jay Gholap. (2012). "Performance tuning of j48 algorithm for prediction of soil fertility",. Asian Journal of Computer Science And Information Technology 2: 8 (2012) 251– 252.

[7] Megala, S., & Hemalatha, M. (2011). "A Novel Datamining Approach to Determine the Vanished Agricultural Land in Tamilnadu",. International Journal of Computer Applications,.

[8] V. Ramesh and K. Ramar, 2011. "Classification of Agricultural Land Soils", A Data Mining Approach. Agricultural Journal, 6: 82-86.

[9] Verheyen, K., Adriaens, D., Hermy, M., Deckers, S. (2001). "High-resolution continuous soil classification using morphological soil profile descriptions", Geoderma 101(3), 31-48.

[10] Rajagopalan, B., Lall, U. (1999). "A k–nearest-neighbor simulator for daily precipitation and other weather variables",. WATER RESOURCES RESEARCH,35(10), 3089-3101.

[11] Mucherino, A., Papajorgji, P., & Pardalos, P. (2009) Data mining in agriculture (Vol. 34). Springer