

Data Mining in Agriculture-A Novel Approach

Yash V. Bagal
Computer Science

Vidyalankar Institute of Technology
Mumbai, India

Shivam V. Pednekar
Computer Science

Vidyalankar Institute of Technology
Mumbai, India

Ashutosh R. Pandey
Computer Science

Vidyalankar Institute of Technology
Mumbai, India

Tanmay B. Dhamdhare
Computer Science

Vidyalankar Institute of Technology
Mumbai, India

Abstract—Data mining is an approach through which in an synchronized manner we can find a workable solution that will be beneficial to increase the growth. The Farmers in agriculture sectors face a lot of issues and difficulties due to the improper understanding and implementation of the activities to enhance their growth and productivity. A large amount of data is available for analyses and scrutiny, however those related to agriculture sector is in a small quantity. Hence segregation and processing of the same from the sources has to be done with proper methodology. Places having multiple grain growth and different soil structure makes it complex to have a perfect estimation of the crops yield both in quantity and quality. Creating a close link between the customer expectation and the producing capabilities of the agriculture sector can be win-win situation at both ends, this can be achieved with capturing data segment wise and in a structured manner. Thus, the customer will be able to fulfil his requirement as per his wish, rather than being satisfied by what is being offered to him. The application of such techniques enables us to predict and make analysis of various problems and helps farmers to make difficult farming decisions based on the conditions, soil fertility, crop duration, disease and other important factors that can result in poor yield production.

Agrarian economy can get a boost and can up their financials by making use of such data mining techniques and they can become self-reliant with their needs.

Keywords— Data mining, data analysis, agriculture, agriculture sector

I. INTRODUCTION

Agriculture the science or practice of cultivation, including progressive improvement of the quality of soil for developing harvests to give food, fleece and other items. Due to increasing urbanization and industrialization, the land under cultivation is decreasing drastically over the years, also agriculture sector is being greatly affected due to population control and climate change. Till date, only a few number of farmers are utilizing new procedures, apparatuses and methods of cultivation for better yield. Information plays a prominent role in agriculture sector

For this, resource utilization should be done in such a way that maximum production and productivity can be achieved. Agricultural enterprises has the ability to collect and generate large amount of data, which extracts the required data using automation. By using data mining, beneficial knowledge and patterns of data can be retrieved. Here data mining comes into picture, which can be used to study and predict the future traits of agriculture. Data mining is a procedure of separating

valuable and significant data from huge sets of data. In our paper, we have explained an outline of application of data mining in agrarian field.

Agriculture is dependent on various factors such as cultivation, irrigation, rainfall harvesting, fertilizers, climate, soil, pesticide, weeds and other factors.[1] Companies in the agricultural sector use data mining to have an estimation of production in order to plan and implement supply chain strategies. Yield prediction is very important in agriculture. The enormous amount of information procured out of these procedures have unexplored potential for improving the effectiveness of the related sectors. Also Data mining discloses hidden information which agricultural management uses for improving its decisions. Data mining is also known as Knowledge Discovery in Data (KDD) and is divided in 2 categories which is Descriptive and Predictive data mining[2].

In Predictive data mining there are values to predict the future whereas in Descriptive mining tasks describe the characteristics of the data in a target data set. But Predictive analysis has a greater application than Descriptive analysis.

Methods in data mining

- Clustering
- Classification
- Association Rule mining
- Regression
- Market Size

II. APPLICATIONS

A. Segregation of fruits and vegetables based on water level content

On a normal basis, fruits and vegetables are classified based on their size and color into different categories which determines its cost. However, these are external factors and they don't really contribute to determine the quality of fruits and vegetables. It mainly depends on the water level content in the fruit, i.e. the amount of water present in the fruit to its weight. Water level in a particular fruit or vegetable can affect its course of life to a certain extent, also abnormal water content can deteriorate the quality of neighboring fruits and vegetables in packaged boxes. Data mining can help us resolve this problem where images of fruits and vegetables are captured at the packaging line, these images are further processed to generate a good guess of the quality of product.

Moreover, records of variety of specimens help to generate a more accurate prediction of the quality of fruits and vegetables. [3]

These images can be fed to VGG 19 model, it is a 19 layer deep Convolutional layer used for large scale image recognition. The image of 224 X 224 RGB image is fed to VGG19 model and at the output layer we use 'Softmax' as an activation function that would give quality rating for the input image in the range of 10 output labels. The training of the model is required with the images of fruits taken with labels provided to them as the rating out 10 by a human.

B. Using data mining to maximize yield depending on the quality of the soil

Cultivation of a particular crop on any land which doesn't meet the minimalistic requirements of the crop would generate yield of lower quality and less revenue for the farmer. Determination of quality of the soil is pre-requisite in agriculture. This gives an analysis of proportions of nutrients and minerals present in the soil. Quality of the soil depends on factors like alkalinity, salinity, moisture contents etc. Data mining is used to study various natures of soil. Soil data analysts suggest the type of crop to be grown and harvested depending upon the fertility of soil which would generate maximum yield. Data mining provides a large set of data for different varieties of soil which can help to predict several traits for cultivation depending on the season and climatic conditions. Implementation of data mining techniques with a wider range of statistical and analytical data and improvises accuracy in extracting information and can also automate results for generic cases [4]. Data mining can also be used to study cross cultivation where in different crops which can be grown simultaneously which would bring in more revenue than single crop cultivation utilizing resources to the best possible extent without affecting the fertility of soil. The scope of data mining is large and its scope can be seen on the soil analysis as follows

- a) Crops can be adopted by sensing and detecting soil which can be done by Artificial Neural Networks.
- b) Previously unknown patterns of soil can be discovered
- c) Traits and behavior of soils can be predicted on the basis of climate conditions and ingredients.
- d) Testing of soil fertility can be done by statistical methods[5].

C. Using data mining to achieve accurateness in agriculture using data mining techniques

Excessive use of pesticides hampers the overall agricultural productivity and in order to tackle this issue there is a need to minimize the use of pesticides in agriculture. Data mining can be used to design automated systems to detect weeds growing in fields [6]. This uses image processing mechanism and it primarily depended on aspect ratios, shapes and surface area. Later on, images of area under cultivation are being processed to find weed patches using specific algorithms [7]. Color density in the images is used to represent the density of

growth of crops in a particular area where the irregular growth of crops was represented by a different color.

D. Analyzing performance of chicken by neural network models

Neural network is a set of artificial neurons which is functional replica of biological neurons interconnected to form different layers of neurons. The first neural layer is said to be as input layer and the last layer as output layer, and the layers between them is called as hidden layers. The neural networks containing multiple hidden layers is called as Deep Neural Network. This setup of neurons can be used in many prediction and classification problems by training the neural network with previous records.

With slight hyper-parameter tuning these type of trained networks can achieve human level accuracy or more but eventually less than 100% accuracy. This analysis of artificial neural networks on feeding Efficiency and weight gain from set of data proposed that concentration of dietary protein is more significant than threonine concentration. A study was proposed that diet containing 0.73% threonine and 18.69% protein may lead to producing weight gain, while on the other hand the efficiency can be achieved from the above data with 0.2 or 0.3 percent standard deviation.

E. Optimizing the use of pesticides by data mining

Excessive use of pesticides can harm the farmer in multiple ways . In agriculture, crop yield forecast is a very important problem. Agriculture researchers conducted a recent study which have shown that in order to maximize the crops yield, the pesticides are over used which is extremely dangerous for the environment. Also, excessive pesticide usage may lead to immunity in pests, which ultimately makes them more harmful to crops and less susceptible to reduction. As a result the overuse of pesticides is creating a health hazard and is imposing a financial burden on farmers and their family. With the help of clustering, one of the method of data mining which can cluster the features by giving interesting patterns of farmer practices and thus provide useful information which will highlight the detrimental effect of excessive pesticide use [9].

Advanced recent concept of spatial correlation has thoroughly impacted the yield prediction in a positive way. For development of forecasting and forewarning models of plant diseases the Artificial Neural Networks plays a very important role. Independent Component Analysis is used for the extraction of independent sources as it is a signal processing technique. It is a numerical and statistical technique for finding hidden characteristics that subtend signals, measurements and various sources[1]. A generative model is proposed for the obtained or observed data..It helps to identify patterns of weather data for optimization of pesticides usage method like integration of agricultural data is more often used as it includes pest scouting, pest usage, and meteorological recording.

REFERENCES

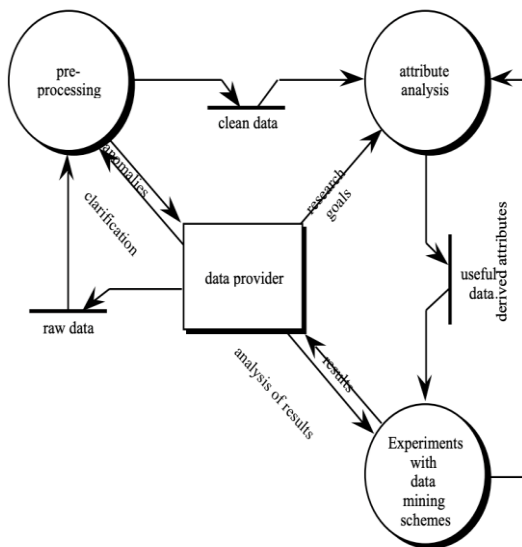


Figure 1. Process model for a machine learning application (data flow diagram)

III. CONCLUSION

Agricultural sector organizations try every day to search information in huge databases for decision making. The condition of decision making can be changed by the use of information technology by which the farmers can yield in much better way. In the agricultural field which is a very dominant and important field, data mining plays a crucial role. Usually the scenario is that the solution for their problems is far within their reach. In this case Data mining proves to be effective for making decisions on various problems associated to agriculture field. Data mining, through better data analysis and management, can assist related organizations to achieve greater benefits. Data mining also provides user oriented access to find hidden pattern in data. In this paper, we have discussed various problems faced in agriculture sector and how data mining solves the problem. The work done by several authors and the role of data mining in this sector is discussed. Agricultural institutions use data mining applications for different areas, such as prediction of problem, disease detection, optimizing the pesticide and so on for making optimal decisions. Hence we can say that Data mining has become a boon to agriculture sector.

- [1] Kaushik Bhagawati, Amit Sen, Kshitiz Kumar Shukla, Rupankar, Bhagawati, Application and scope of data mining in agriculture, International Journal of Advanced Engineering Research and Science, 3(7) 2016, 66-69
- [2] R. S. Kodeeshwari, K. Tamil Ilakkiya, Different Types of Data Mining Techniques Used in Agriculture - A Survey International Journal of Advanced Engineering Research and Science (IJAERS), ISSN: 2349-6495(P) | 2456-1908(O)
- [3] Namita Mirjankar, Smitha Hiremath, Application of Data Mining In Agriculture Field, International Journal of Computer Engineering and Applications, iCCSTAR-2016, Special Issue, May 2016.
- [4] 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 6.
- [5] Ramesh Babu Palepu, Rajesh Reddy Muley, An analysis of agricultural soils by using Data Mining techniques, 2017 IJESC Volume 7 Issue no 10
- [6] Alberto Tellaechel, Xavier-P. Burgos Artizzu, Gonzalo Pajares, and Angela Ribeiro, A Vision-Based Hybrid Classifier for Weeds Detection in Precision Agriculture Through the Bayesian and Fuzzy k-Means Paradigms, Advances in Soft Computing, vol 44
- [7] Yang, C.C., Prasher, S.O. Landry, J.A. and Ramaswamy, H.S.: Development of an Image Processing System and a Fuzzy Algorithm for Site-specific Herbicide Applications. Precision Agriculture, 4 (2003) 5–18.
- [8] Mythili. R, Pradeep Raj. D: Survey of Data Mining Techniques and Applications of New Techniques in Agriculture IOSR Journal of Engineering, (2018) PP 31-35.
- [9] Shalin Paulson St. Joseph's College : A Survey on Data Mining Techniques in Agriculture International Journal of Engineering Research & Technology, 3(30)-2015