APPLICATIONS OF VARIOUS DATA MINING TECHNIQUES USED IN AGRICULTURE SECTOR TO INCREASE PRODUCTIVITY

[1] Srivaramangai. R, [2] Rajendra B.Patil, [3] Vilas T. Mahajan

[1]Assistant Professor, [2] Assistant Professor, [3] Assistant Professor
[1]Department of Information Technology,
[1] University of Mumbai, Vidyanagari, Santacruz-East, Mumbai, India

Abstract :   Absence of mindfulness among ranchers precludes them from utilizing logical strategies in cultivating. To refer to only one occurrence, agriculturists are totally ignorant of those diverse parameters that contain the dirt in totality. Tillers reap their normal products on their ranches as a conventional daily schedule. Subsequently, without knowing the necessities of the dirt, they wind up utilizing pointlessly extensive portions of manures and additionally toxic pesticides on the products. As a result, once in a while the dirt gets tainted and contaminated with these lethal fixings, in this manner rendering it barren. This awful propensity can irreversibly harm both the dirt and nature. As a pre-essential to changing this issue, we can apply information digging for soil and develop the harvest as indicated by the dirt parameters. We can likewise foresee the harvest efficiency dependent on the past information and in like manner products can be cultivated. Data mining is one of the rising patterns of software engineering utilized for different kinds of examination. Information mining has different strategies, for example, grouping, forecast, representation, and so forth. In this paper we have talked about different information mining methods alongside their applications in horticulture and future prospects in the field. This is a rising examination field.

IndexTerms - **Data mining, agricultural application, k-means algorithm, classification, clustering, knowledge discovery, KDD.**

I. INTRODUCTION

Data mining is the procedure of extraction of valuable data and examples from tremendous information. It is additionally called Knowledge Discovery in Database (KDD) process, learning mining from information, information extraction or information/design investigation [2]. The motivation behind data mining utilizing PC supported framework is to separate valuable data from expansive databases or information distribution centers for examination reason and make a few forecasts or to choose the pattern. The following figure1shows KDD process [1]:
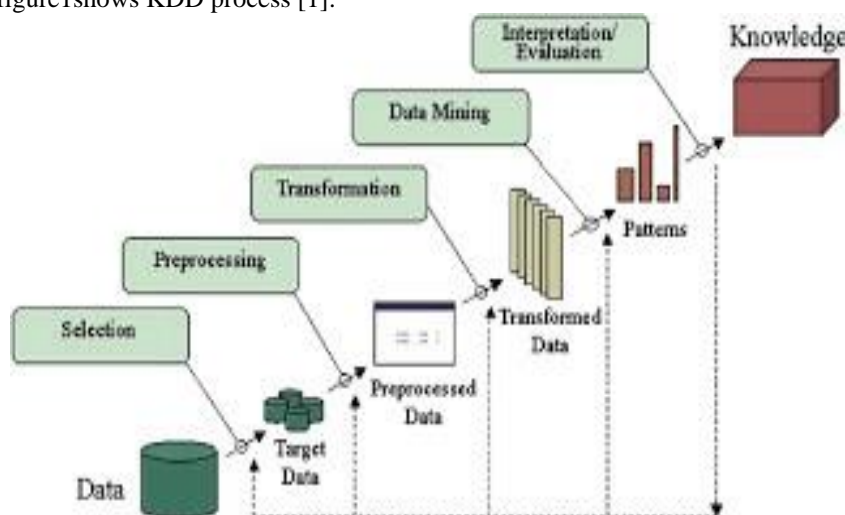

Fig 1. KDD Process

Knowledge Discovery in Database (KDD) is a process that consists of several tasks as shown below:

Clustering: Clustering is the process of partitioning a set of data (or objects) into a set of meaningful sub-classes called clusters. It helps users understand the natural grouping or structure in a data set.

Association rule learning: Association rule learning is a method for discovering the relationship between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness [15]. An example of an association rule would be "If a customer buys a shirt, he is 80% likely to purchase a pair of trousers also."

Data cleansing: It is a process that removes or transforms noisy, unwanted and inconsistent data.

Data integration: In this process, data from multiple heterogeneous data sources may be combined into one.

Pattern evaluation: It is a process that identifies the truly interesting patterns representing knowledge based on some interesting measures.

Data transformation: In data transformation data is transformed or consolidated into forms appropriate for mining.

Data Mining (DM) is one of the core steps of the KDD process. Its goal is to apply data analysis and knowledge discovery algorithms that produce a particular enumeration of patterns (or models) over the data. Data mining model/pattern takes some input data (generally historic data) and returns us some results or predictions based on the given historic data. Various types of models/patterns exist, such as clusters, decision trees, association rules, frequent item sets, sequential patterns, etc.

There are two ways of learning in data mining:

a. Supervised learning:
Supervised learning is one of the data mining algorithms which comprises of a set of examples called training data with known labels. The label data can be nominal values used in the classification task, or numerical values in the case of the regression task. The label data can be nominal values used in the classification task, or numerical values in the case of the regression task. [14]

Example1: If there should be an occurrence of agrarian, information with various kinds of pesticides on various homesteads are recorded with their viability on the yield. For this situation one can know whether they have an explicit sickness (class: yes) or not (class: no). The objective is to build a model for every one of the diverse classes, in order to discover the pesticide and its viability when input data is given.

Example2: In medical data, you have a record of people for which you know whether they have any specific disease (class: yes) or not (class: no). Each disease has its symptoms. The goal is to build a model for each of the different classes according to the symptoms, so as, for any a new patient for which you have no clue about its class, to predict its class.

b. Unsupervised learning
There is no prior knowledge in the instances about the right answer. Based on the instance characteristics, the instances are organized into groups of similar instances.
For example: in a news aggregation site like Google news, the news posts are organized according to their topic. To do so, the similarity between the news posts is employed.

Following are some of the applications of data mining in the field of agriculture:

1. Soil classification
Content of salt greatly affect many processes:
(a) Crop growth (including yield, quality, and economic returns)
(b) Soil physical properties (such as aggregation and water infiltration) •
(c) Sufficiency and toxicity of nutrients

Shweta Taneja, Rashmi Arora, Savneet Kaur [13] UAD WEKA tool and implemented Clustering technique and generated various clusters of the soil database. They selected K-Means algorithm for similarity search. The k-means algorithm follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori.
A survey of data mining techniques applied to agriculture GPS-based technologies (Verheyen et al. 2001), (Meyer et al. 2004) use a k-means approach to classify soils and plants and (Camps-Valls et al. 2003) uses SVMs to classify crops.

2. Checking of apples
There are machines used for scanning fruits. Usually, apples are placed on rollers on which they rotate around themselves and pictures are taken by cameras located above. Some systems have been developed where photos are taken by three cameras to get the details of each region. After getting the images a segmentation process is applied. The result of the segmentation is the division of the image in many related regions, for instance gray color. K-means is used to analyze color images of fruits as they run on conveyor belts[11]. There are systems where X-ray images of apples are analyzed to monitor the presence of water cores, in which a neural network is trained for discriminating between good and bad apples [12].

3. Wine fermentation
Productivity and quality of wine are greatly affected by the problems occurring during the wine fermentation process. The wine

fermentation process can be monitored using data mining techniques. In order to monitor the wine fermentation process, metabolites such as glucose, fructose, organic acids, glycerol and ethanol can be measured, and the data obtained during the entire fermentation process can be analyzed in order to obtain useful information.Fermentation process is predicted with the help of extracted information from the large databases. K-means algorithm has been applied for exploring the data [14-16]. Taste sensors are used to obtain data from the fermentation process to be classified using Artificial Neural Networks (ANNs).

4. Detection of diseases from sounds issued by animals

Various farm animals living on the farms are used as an extra source of income by farmers. Unfortunately, sometimes sick animals become the cause of infections spreading over to the crop. Therefore early detection of creature maladies that live on homesteads can chop down superfluous misfortunes to the rancher and emphatically affect profitability of the ranch. Also, recognizing a disease at its beginning enables the rancher to begin the creature's treatment when the side effects show up. It has been observed that sounds emitted by many animals such as pigs can be analyzed for the detection of diseases [5].

5. Weather/rainfall prediction

KNN and decision trees have applied data mining techniques are applied in agriculture related fields, for instance, for simulating daily rainfall, climate change and other weather variables. The other application of the KNN method in agriculture includes climate forecasting [4]. Various models have been created for weather prediction, rainfall prediction, and climate change predictions that use many data mining techniques as shown in table 1.

Table 1. Comparison Chart of Major Research Works

Authors	Applications	Techniques/ Algorithms	Attributes	Time Period/Dataset /Size/Accuracy	Advantages	Disadvantages
E.G.Petre[07]	Weather Prediction	Decision Tree/ CART	Pressure, clouds quantity, humidity, precipitation, temperature	4 years/48/83%	Good, prediction, accuracy	Data transformations required. Extra computation required.
S. Yeonetal.[08]	Hourly, rainfall, prediction	Decision, Tree/ C4.5, CART	Temperature, wind direction, speed, gust, humidity, pressure	3 years 26280 99%, 93%	High, prediction, accuracy	Small data is left for prediction.
F. Oliya, ABAdeyemo [09]	Weather Prediction and Climate Change Studies	Decision tree, ANN/ C4.5, CART, TLFN	temperature, rainfall, evaporation, wind speed	10 years/36000/82 %	Best network is selected for prediction	Accuracy varies highly with size overtraining dataset
Amarakoon[10]	Climate prediction in Sri Lanka	ANN/KNN	Temperature, humidity, precipitation, wind speed	1 year/365/	Beneficial for dynamic data.	Need to integrate feature selection techniques.

Selection of seed

There is also an additional dimension to the evaluation of the agricultural production data with the aim to increase the production of crop yield by selecting a proper seed based on the characteristics of soil and seed. Therefore, one can improve the accuracy of crop yield prediction and seed panting recommendation. The Western Australian Agricultural Agency, has sought new of data mining in order to improve such predictions and seed variety recommendations. [17]. Further there is more scope to widen the research for solving this problem with respect to individual crop.

7. Conclusion

Agricultural field without a doubt is an imperative area in India. The economy of India is mainly based on agricultural sector. There exist various challenges that are faced by farmers; few of them are reported in this paper. If these difficulties are overcome by research analysts, then that will be a great technological development for this sector. In this paper we have talked about the extent of datamining procedures for understanding the different issues identified with farming field. Additionally this paper will assist the analysts with extending their knowledge and contribution towards this field.

References

[01] Wu X, Kumar V, Quilan JR, Ghosh J, Yang Q, Motoda H, McLanchlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D, Top 10 algorithms in data mining. KnowlInfSyst14 : 1-37, 2008.

[02]Mrs. Bharati M. Ramageri,DATA MINING TECHNIQUES AND APPLICATIONS, Lecturer Modern Institute of Information Technology and Research, Department of Computer Application, Yamunanagar, Nigdi, Pune, Maharashtra, India-411044.

[03] Usama Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth (1995), "From Knowledge Discovery to Data Mining: An Overview," Advances in Knowledge Discovery and Data Mining, U. Fayyad et al. (Eds.), AAAI/MIT Press

[04] Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background. SadeghBafandehImandoust and Mohammad Bolandraftar Department of Economics,Payame Noor University, Tehran, Iran

 [05] Different Techniques Used in Data Mining in Agriculture. Jyotshna Solanki, Prof. (Dr.) Yusuf Mulge, PDM College of Engineering & Technology, Bahadurgarh, Haryana, India

[06] Divya Chauhan, Jawahar Thakur: Data Mining Techniques for Weather Prediction: A Review, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2 Issue: 8, ISSN: 2321-8169

 [07] Elia Georgiana Petre "A Decision Tree for Weather Prediction", Buletinul, Vol. LXI No. 1, 77-82, 2009.

[08] Soo-YeonJi, Sharad Sharma, Byunggu Yu, Dong Hyun Jeong, "Designing a Rule-Based Hourly Rainfall Prediction Model", IEEE IRI 2012, August – 2012.

[09] [17] Kaya, E.; Barutçu, B.; Menteş, S. "A method based on the van der Hoven spectrum for performance evaluation in prediction of wind speed". Turk. J. Earth Science, 22, 1–9, 2013.

[10] A.R.W.M.M.S.C.B. Amarakoon, "Effectiveness of Using Data Mining for Predicting Climate Change in Sri Lanka", 2010.

[11] Leemans V, Destain MF A real time grading method of apples based on features extracted from defects. J Food Eng 61:83–89, 2004.

[12] Shahin MA, Tollner EW, McClendon RW Artificial intelligence classifiers for sorting apples based on water core. J AgricEng Res 79(3):265–274, 2001.

[13] ShwetaTaneja, RashmiArora ,Savneet Kaur Mining of Soil Data Using Unsupervised Learning Technique

[14] S.Neelamegam, Dr.E.Ramaraj, Classification algorithm in Data mining: An Overview, International Journal of P2P Network Trends and Technology (IJPTT) – Volume 4 Issue 8- Sep 2013

[15] Piatetsky-Shapiro, Gregory (1991),Discovery, analysis, and presentation of strong rules, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds.,Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA.

[16] P. Grace Sharon, Application of Data Mining In Agriculture, International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 2, Issue 3, pp: (158-162), Month: July - September 2014.

[17] YunousVaghl Dr. Leisa Armstrong, Dean Diepeveen, Application of a data mining framework for the identification of agricultural production areas in WA, Edith Cowan University Research Online, ECU Publications Pre. 2011