International Journal of Advances in Computer Science and Technology

A Survey on Data Mining Techniques in Agriculture

Ms.Kalpana.R[1],   Dr.Shanthi.N[2] , Dr.Arumugam.S[3]
[1]PG Student , Nandha Engineering College (Autonomous) , Erode, rm.kalpana@gmail.com
[2] Dean, Nandha Engineering College (Autonomous) , Erode, deancse@nandhaengg.org
[3] CEO, Nandha Engineering College (Autonomous) , Erode, principalnec@nandhainstutions.org

ABSTRACT

Data processing in Agriculture is associate rising analysis field. Application of information mining techniques plays an important role in agricultural and environmental connected areas. Survey created on agriculture mistreatment data processing techniques reveals the importance to extract helpful data from dataset. Recently data processing techniques in agriculture required evaluating, storage, monitored and retrieval the resources used. Completely different data processing techniques like K-Means, K-Nearest Neighbor (KNN), Support Vector Machine and Artificial Neural Network square measure terribly recent applications in agriculture. In agriculture, some issues like yield estimation and crop productivity remains to be solved supported obtainable knowledge. This survey aims to search out appropriate data processing models to realize high accuracy and prediction capabilities. It's the opinion that additional techniques and algorithms to be studied connected agricultural issues can provide smart lead to agricultural growth. Finally, mistreatment data processing techniques in agriculture could be a up to date technique to search out the answer over the normal and traditional technique.

Key Words-K-Means (KNN), Support Vector Machine, Artificial Neural Network, Yield Estimation

1. INTRODUCTION

1.1 DATA MINING

Data mining (sometimes known as knowledge or data discovery) is that the method of analyzing knowledge from completely different views and summarizing it into helpful data that may be won't to increase revenue, cuts costs, or both. Data processing computer code is one in every of variety of analytical tools for analyzing knowledge. It permits users to research knowledge from many alternative dimensions or angles, reason it, and summarize the relationships known. Technically, {data mining and data method} is that the process of finding correlations or patterns among dozens of fields in giant relative databases.

Data mining, or data discovery, is that the computer-assisted method of dig through and analyzing monumental sets of information so extracting which means of the info. Data processing tools predict behaviors and future trends, permitting businesses to form proactive, knowledge-driven selections [6]. Data processing tools will answer business queries that historically were too moment overwhelming to resolve. They scour databases for hidden patterns, finding prophetic data that specialists might miss as a result of it lies outside their expectations [2][7].

1.2 ESSENCE OF APPROACH

1.2.1. Basic Facts

Data mining has attracted an excellent attention within the data trade and in society as an entire in recent years, because of wide availableness of giant quantity {of knowledge and of data of data} and therefore the close at hand want for turning such data into helpful information and knowledge. {The information or the data and data} and knowledge gained may be used for application starting from market research, fraud detection, to production management, disaster management and science exploration. Data processing may be viewed as a result of the natural evolution of data technology. The information system trade has witnessed associate organic process path within the development of assorted functionalities: knowledge assortment and information creation, direction (including knowledge storage and retrieval, and information dealing process and advance knowledge analysis. data discovery as a

method consists of associate unvaried sequence of following steps:

1. K*nowledge improvement*, that is, to get rid of noise and inconsistent knowledge.

2. *Knowledge integration*, that is, wherever multiple knowledge sources square measure combined.

3. *Knowledge choice*, that is, wherever knowledge relevant to the analysis task square measure retrieved from the information.

4. *Knowledge transformation*, that is, wherever knowledge square measure reworked or consolidated into forms applicable for mining by playing outline or aggregation operations.

5. *Data processing*, that is, an important method wherever intelligent ways square measure applied so as to extract the info patterns.

6. *Data presentation*, that is, wherever image and data illustration techniques square measure wont to gift the mined data to the user [4][11].

1.3 DATA MINING IN AGRICULTURE

Data mining in agriculture could be a terribly recent analysis topic. It consists within the application of information mining techniques to agriculture. Recent technologies square measure today able to offer plenty of data on agricultural-related activities, which may then be analyzed so as to search out vital data. Carrying out effective and property agriculture has become a crucial issue in recent years.

Agricultural production needs to continue with associate ever-increasing population. A key to the current is that the usage of recent technologies likes GPS (for exactitude agriculture) and data processing techniques to require advantage of the soil's non uniformity. The big amounts of information that square measure today nearly harvested at the side of the crops got to be analyzed and will be wont to their full extent - this is often clearly an information mining task. Data processing permits to extract the foremost vital data from such vast data knowledge and to uncover antecedently unknown patterns that will be relevant to current agricultural issues, thereby serving to farmers and managing organizations to remodel knowledge into business selections[10][19].

1.4 DATA MINING TECHNIQUES IN AGRICULTURE

During this survey we tend to gift a number of the foremost used data processing techniques within the field of agriculture. a number of these techniques, like the k-means, the k nearest neighbor, artificial neural networks and support vector machines, square measure mentioned associated an application in agriculture for every of those techniques is conferred. Data processing in agriculture could be a comparatively novel analysis field[4]. It's our opinion that economical techniques may be developed and tailored for determination complicated agricultural issues mistreatment data processing. At the top of this survey we offer recommendations for future analysis directions in agriculture-related fields.

Data mining techniques may be chiefly divided in 2 groups: classification and agglomeration techniques. Classification techniques square measure designed for classifying unknown samples mistreatment data provided by a collection of classified samples. This set is sometimes remarked as a coaching set, because, in general, it's wont to train the classification technique a way to perform its classification. As an example, neural networks and support vector machines exploit coaching sets for standardization their parameters so as to resolve a selected classification drawback[11]. In alternative words, these 2 classification techniques learn from a coaching set a way to classify unknown samples, i.e., samples whose classification is unknown. Another classification technique, the k nearest neighbor, doesn't have any learning part, as a result of it uses the coaching set on every occasion a classification should be performed. For this reason, the k nearest neighbor is sometimes remarked as a lazy classifier. During this case, agglomeration techniques may be wont to split a collection of unknown samples into clusters. One in every of the foremost used agglomeration techniques is that the k-means technique. Iit's applied to several analysis fields and an outsized range of its variants may be found within the literature[1][7][16].

2. APPLICATIONS

Several data processing techniques employed in agriculture study space. We tend to square measure mentioned the few techniques here. a number of the info mining techniques square measure associated with climatic conditions and forecasts. as an example, the K-Means algorithmic program is employed to perform forecast of the pollution within the

atmosphere, the K Nearest Neighbor (KNN) is applied for simulating daily precipitations and alternative weather variables, and completely different doable changes of the weather situations square measure analyzed mistreatment SVMs. additionally K means that technique is employed to forward the pollution in atmosphere[1][5][9]. Completely different changes of weather square measure analyzed mistreatment SVM. K means that approach is employed to classify the soil and plants. Wine fermentation method monitored mistreatment data processing techniques. By mistreatment Multilayer Perception model of Neural Networks the researchers trained to predict wheat yield by considering detector input and fertilizers as parameters. SVMs for detective work weed and element stress in corn. Data processing techniques square measure typically won't to study soil characteristics. The K-Means approach is employed for classifying soils together with GPS-based technologies [2] [6]. Apples square measure checked mistreatment completely different approaches before causation them to the market, and a neural network is trained for discriminating between smart and unhealthy apples. Apply a supervised biclustering technique to a dataset of wine fermentations with the aim of choosing and discovering the options that square measure to blame for the problematic fermentations and additionally exploit the chosen options for predicting the standard of recent fermentations. Style sensors square measure to get knowledge from the fermentation method to be classified mistreatment ANNs. Similarly, sensors square measure to smell milk, that's classified mistreatment SVMs [5] [12].

3. MATERIAL AND WAYS

Following are unit varied data processing techniques such as:

3.1.Classification

Classification is that the method of finding a model that describes and distinguishes information categories or ideas for the aim of having the ability to use the model to predict category or the category} of objects whose class label is unknown [9].

3.2 Prediction

Prediction has attracted right smart attention given the potential implications of flourishing statement in a very business context. There are unit 2 major kinds of predictions: one will either try and predict some inaccessible information values or unfinished trends, or predict a category label for a few information. The latter is tied to classification. Once a classification model is constructed supported a coaching set, the category label of associate object may be predicted supported the attribute values of the article and therefore the attribute values of the categories [15]. Prediction is but additional typically remarked the forecast of missing numerical values, or increase/ decrease trends in time connected information. The foremost plan is to use an outsized range of past values to think about probable future values [10][11].

3.3 Clustering

Clustering analyses information objects while not consulting a notable category label. The unattended learning technique of bunch may be a helpful methodology for ascertaining trends and patterns in information, once there aren't any pre-defined categories [16].

3.4 Outlier Analysis

Information might contain information objects that don't befit the overall behavior or model of the information. These information objects area unit outliers. Outliers area unit information parts that can't be sorted in a very given category or cluster. Additionally called exceptions or surprises, they're typically vital to spot. Whereas outliers may be thought-about noise and discarded in some applications, they will reveal vital information in different domains, and therefore may be terribly important and their analysis valuable [15]. These area unit techniques that area unit utilized in {data mining or data methoding} within the information of discovery process. However here we have a tendency to used cluster analysis principally [15] [17].

3.5 K-Means algorithmic rule

The k-means methodology is one of or one among in a the foremost standard unattended learning or bunch ways that are applied in a form of fields together with pattern recognition, info retrieval, document extraction and biological science analysis and then forth. The tactic is termed the k-means as a result of

it represents every of k range of clusters by the mean of its weighted points. The goal of the tactic is to classify a given information set through an explicit range of clusters specified some metric relative to the centrist of the clusters is decreased .The k-means algorithms is simple to implement and its time complexness is depend upon its pattern[8][20].

3.6 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[11]. 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 [10].

3.7 Artificial Neural Network

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[3]. A neural network is associate interconnected set of input/output units wherever every association includes a weight related to its [2].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 [10].

3.8 Support Vector Machine

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[13]. In straightforward words given a group of coaching examples, every marked as happiness to at least one of 2 classes, associate SVM coaching algorithmic rule builds a model that predicts whether or not a brand new example falls into one class or the opposite [2].

4. SAMPLE PARAMETER ANALYSIS

SVM models were generated for separating crop kinds of Winter . Major crop varieties known were barley and wheat. Crops like chick pea, and fodder were sorted into different crops [13].Winter crop kind classification accuracy was once more found to be slightly higher (79%) than that of summer (Table 1).

Table 1. Accuracy assessment winter classification of crop varieties

Classified Crops	Wheat	Barley	Woody	Cotton	Sorghum	Maize	Row-wise Total
Wheat	2	14	0	62	16	3	97
Barley	0	0	0	1	2	0	3
Woody	3	6	0	0	0	11	20
Cotton	7	3	12	0	0	11	33
Sorghum	34	30	0	20	32	27	143
Maize	2	0	70	0	0	2	74
Column-wise Total	48	53	82	81	50	54	370
Overall accuracy = 79%							

5. RESULTS AND DISCUSSION

Several data processing techniques utilized in agriculture study space. We have a tendency to area unit mentioned the few techniques here. Additionally one technique known as K suggests that methodology is employed to forward the pollution in atmosphere. Totally different changes of weather area unit analyzed mistreatment SVM. K suggests that approach is employed to classify the soil and plants. Wine fermentation method monitored mistreatment data processing techniques [5] [6].

6. CONCLUSION

Thus we have a tendency to conclude that, there's a growing range of applications of information mining techniques in agriculture and a growing quantity of information that area unit presently obtainable from several resources. This is often comparatively a completely unique analysis field and it's expected to grow within the future. There's plenty of labor to be done on this rising and fascinating analysis field. The multidisciplinary approach of group action engineering science with agriculture can facilitate in forecasting/ managing purpose. Data processing techniques once applied to associate agricultural profile, might improve the verification of valid profile, might improve the verification of valid patterns and profile classification when put next to plain applied mathematics analysis techniques. Data processing and therefore the varied methodologies related to it will scale back the complexness of the information sanctionative farmers to create choices additional simply.

REFERENCES

[1] Elodie Vintrou, Dino Ienco, Agnès Bégué, and Maguelonne Teisseire, **"Data Mining, A Promising Tool for Large-Area Cropland Mapping"**, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 6, NO. 5, OCTOBER 2013

[2] S.S.Baskar, L.Arockiam, V.Arul Kumar, L.Jeyasimman, **"Brief Survey of Application of Data Mining Techniques to Agriculture"**, Agricultural Journal 5(2): 116:118,2010 ISSN:1816-9155

[3] Lei Zhou, Jianjun Wu, Jianhui Zhang, Song Leng, Ming Liu, Jie Zhang, Lin Zhao, Fengying Zhang, and Yu Shi, **"The Integrated Surface Drought Index (ISDI) as an Indicator for Agricultural Drought Monitoring: Theory, Validation, and Application in Mid-Eastern China"**, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 6, NO. 3, JUNE 2013

[4] Phillip McKerrow, *Member, IEEE,* and Neil Harper**, " Plant Acoustic Density Profile Model of CTFM Ultrasonic Sensing"**, IEEE SENSORS JOURNAL, VOL. 1, NO. 4, DECEMBER 2001

[5] Raorane A.A, Kulkarni R.V, "Data Mining: An effective tool for yield estimation in the agricultural sector", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 1, Issue 2, July – August 2012

[6] 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

[7] Latika Sharma, Nitu Mehta, **"Data Mining Techniques: A Tool For Knowledge Management System In Agriculture"**, International Journal Of Scientific & Technology Research Volume 1, Issue 5, June 2012

[8] D. Diepeveen and L. Armstrong**," Identifying key crop performance traits using data mining"**, Iaald Afita Wcca2008 World Conference On Agricultural Information And IT, Vol. 3 No. 1 Jan 2011

[9] A.Vjay Kumar, T.V. Rajini Kanth, **"Estimation of the Influence of Fertilizer Nutrients Consumption on the Wheat Crop yield in India- a Data mining Approach"**, International Journal of Engineering and Advanced Technology (IJEAT) Volume-3, Issue-2, December 2013

[10] 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

[11] D.Rajesh**, "Application of Spatial Data Mining for Agriculture",** International Journal of

Computer Applications (0975 – 8887) Volume 15– No.2, February 2011

[12] Raorane A.A and Kulkarni R.V, **"Review- Role of Data Mining in Agriculture"**, International Journal of Computer Science and Information Technologies, Vol. 4 (2) , 2013, 270 – 272

[13] 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

[14] Han-Wen Hsiao,Meng-Shu Tsai, And Shao-Chiangwang**, "Spatial Data Mining of Colocation Patterns for Decision Support in Agriculture"**, Asian Journal of Health and Information Sciences, Vol. 1, No. 1, pp. 61-72, 2006

[15] Meghali A. Kalyankar Prof. S. J. Alaspurkar, **"Data Mining Technique to Analyse the Metrological Data"**, International Journal of Advanced Research in Computer Science and Software Engineering,   Volume 3, Issue 2, February 2013

[16] Leisa J. Armstrong, Dean Diepeveen, Rowan Maddern**, "The application of data mining techniques to characterize agricultural soil profiles",** International Journal of Computer Science and Technology IJCST

[17] G. Holmes, D. Fletcher and P. Reutemann**, "An application of data mining to fruit and vegetable sample identification using Gas Chromatography-Mass Spectrometry"**, International Environmental Modelling and Software Society (iEMSs) 2012

[18] Chih-Hung Hsu, An-Yuan Chang, Hui-Ming Kuo, **"Data Mining QFD for The Dynamic Forecasting of Life Cycle under Green Supply Chain"**, WSEAS TRANSACTIONS on COMPUTERS Issue 1, Volume 11, January 2012

[19] M.R.Pavan Kumar, K.S.Ranjith, B.Kiran Kumar and G.Mahesh Yadav**, "Analysis of Spatial Data Mining and Global Autocorrelation",** International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 5, May 2012

[20] Prof. Chandrakanth. Biradar, Chatura S Nigudgi, **" An Statistical Based Agriculture Data Analysis"**, International Journal of Emerging Technology and Advanced Engineering Volume 2, Issue 9, September 2012