

Data Mining In Agriculture

Sandeep Chauhan
Research Scholar, MCA
Thakur Institute of Management Studies, Career
Development and Research, Mumbai (MS), India
sandeep11012@gmail.com

Suraj Chaurasia
Research Scholar, MCA
Thakur Institute of Management Studies, Career
Development and Research, Mumbai (MS), India
surajkumar3433@gmail.com

Abstract— Agriculture in today's life isn't like as our forefather done. The robust environmental condition changes due to several reasons like global warming cause difficulty to grasp weather conditions. therefore the farmers unable to grasp that crop to pick out by that the production can improve. By understanding soil and climate conditions by using these data mining system farmers are going to be able to take right crop at right place which can improve yields. therefore it's straightforward for farmers to come to a decision that crop to require in unpredictable climate conditions.

Keywords—Data mining, agriculture, Techniques, K-Means, K-Nearest Neighbor.

I. INTRODUCTION

Agriculture field is that the backbone of any country. Agriculture supplies the food and raw materials to the individuals in country. It is the only financial gain supply of the many peoples. Peoples who belongs with agriculture field faces several issues like decreasing production due to unsuitable environmental condition changes, flood, dearth and many alternative natural reasons and rarely factors. they're unable to do agriculture attributable to this reasons. we will use information Technology (IT) to beat this issues. In today's life Information Technology is used in each field worldwide. The Data Mining could be a part of IT that we will use to solve agriculture problems mentioned above. the fundamental plan of the data Mining is that it generates helpful information by extracting from giant datasets. To be additional correct, it's a method of extracting helpful information from great amount of data. it's the observe of automatically looking giant stores of data to find associations and trends that transcend easy analysis. Data mining will answer queries that can't be addressed through simple question and reportage techniques.

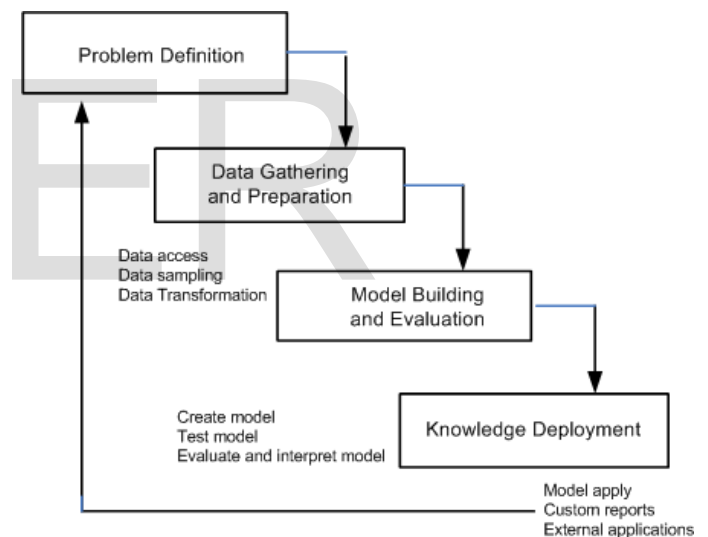
Data mining this method regarding results in the invention of recent patterns in giant data sets. the most goal of the data mining process is to extract information from the previous data set .This method of examine data from the various outlooks and describing into helpful information. No restriction to the kind of information which will be examined by the data mining. the main focus of this paper is to offer information regarding totally different data processing techniques in outlook of agriculture domain.

Data mining is a modern information analysis technique. it's big selection of applications within the field of agriculture. during this study, applications of the data mining techniques within the area of agriculture and its allied areas are studied. completely different techniques of data mining are used in this field. Though,

there are many techniques available within the data processing, few methodologies such as K-means approach, K- nearest neighbor, Bi-clustering are widespread presently depends on the character of the data

II. THE DATA MINING PROCESS

The below shown figure illustrates the phases, and also the repetitive nature, of a data mining prediction. the method flow shows that a data mining prediction doesn't stop once a particular resolution is deployed. The results of data mining activate new business queries, that successively will be used to develop a lot of focused models.



A. Problem definition

In problem definition stage we are focusing the matter that that we are going to solve.

B. Data gathering & preparation

This is a second step for data mining process; here we are assembling the information associated with problem and make getting to analysis the problem.

C. Model building & evaluation

Model building & evaluation is that the third step for data mining process; here we done the works like model creation, take a look at model and evaluating the model.

D. Knowledge deployment

Knowledge deployment is that the last for data mining process; during this stage we are progressing to apply the created model for external process and collecting the feedback from the customers. Finally getting ready the reports for any development.

III. DATA MINING TECHNIQUES

Data Mining techniques are primarily divided in 2 groups, classification and clustering techniques [1].

A. Classification Techniques

Classification techniques are designed for classifying unknown samples using data provided by a collection of classified samples. This set is sometimes spoken as a coaching set because it is employed to train the classification technique the way to perform its classification. Generally, Neural Networks[2,3,4] and Support Vector Machines[5], these 2 classification techniques learn from coaching set the way to classify unknown samples.

The k nearest neighbor (k-NN)[6] is a technique for classification, K-Nearest Neighbor, does not have any learning section, as a result of it uses the training set anytime a classification should be performed. A training set is thought, and it's accustomed to classify samples of unknown classification. The essential assumption within the K-Nearest Neighbor rule is that similar samples ought to have similar classification. The parameter K shows the number of comparable fanned samples used for assignment a classification to Associate in Nursing unknown sample. The K-Nearest Neighbor uses the knowledge within the coaching set, however it will not extract any rule for classifying the opposite.

B. Clustering Techniques

In the event a coaching set not offered, there's no previous knowledge concerning the information to classify. during this case, clustering techniques[7] are often wont to split a group of unknown samples into clusters. one amongst the foremost used cluster technique is the K-Means rule. The k-means is a data mining technique for clustering. Given a group of information with unknown classification, the aim is to search out a partition of the set within which similar information square measure sorted within the same cluster. The parameter K plays a very important role because it specifies the number of clusters within which the information should be divided. The idea behind the K-Means rule is, given a particular partition of the information in K clusters, the centers of the clusters can be computed because the suggests that of all samples happiness to a clusters. the middle of the cluster are often thought-about because the representative of the cluster, as a result of the middle is kind of shut to all samples within the cluster, and so it's kind of like all of them. There square measure some disadvantages in victimisation K-Means method. one amongst the disadvantages may well be the selection of the parameter K. Another issue that wants attention is that the computational value of the rule. There square measure alternative information Mining techniques applied mathematics based mostly techniques, such as Principle element Analysis(PCA), Regression Model and Biclustering Techniques[8,9] have some applications in agriculture or agricultural - connected fields.

IV. APPLICATIONS

Agriculture is a advanced large system. india is large country with multifaceted soil sorts and plentiful crop

varieties. However, the pests and unwellness occur frequently, and therefore the relationship among fertiliser, water, density and climate has to be studied. Agricultural information knowledge have the properties of large range, numerous dimensions, dynamic degree, wholeness and uncertainty.

A. Application:

- *K-nearest neighbour to weather and forecasts*

There are many applications of data Mining techniques in the field of agriculture. a number of the data mining techniques are associated with weather and forecasts. for instance, the K-Means algorithm is used to perform forecast of the pollution within the atmosphere[10], the K Nearest Neighbor(KNN) is applied for simulating daily precipitations and different weather variables[11], and totally different possible modification of the weather eventualities are analyzed victimization SVMs[12].

- *K-nearest neighbour to check sound recognition*

Data Mining techniques are applied to check sound recognition issues. for example, Fagerlund S[13] uses SVMs to classify the sound of birds and alternative completely different sounds. Holmgren et al.[14] uses a K-Nearest Neighbor approach to gauge forest inventories and to estimate forest variables for analyzing satellite imagery. Das kc et al.[15] uses ANNs to classify eggs as fertility and Patel VC et al.[16] uses pc Vision to acknowledge cracks in eggs. Du C-J et al. [17] uses SVMs to classify pizza pie sauce unfold and Karimi Y et al.[18] uses SVMs for detecting weed and nitrogen stress in corn.

- *K-means to evaluate soil fertility*

Weighted K-means clustering algorithm are often used to judge the soil fertility[19]. The algorithm uses AHP to induce the load of soil nutrient attributes. Then combined with K-means bunch algorithm. Finally through the operational potency and accuracy to work out the optimum classification, that can improve the clustering algorithm of intelligent. The algorithm and therefore the ancient K-means clustering algorithm are utilized in the comparison, tests showed that the weighted K-means clustering algorithm has a better accuracy, operational potency, considerably above the un-weighted clustering algorithm; Comprehensive analysis of the changes in soil nutrients once precision fertilization that used rule. The soil fertility status has a considerably improvement once years of continuous precision fertilizing. The results show that the improved clustering algorithm may be a sensible technique to comprehensive analysis of soil fertility.

- *K-nearest neighbour for Classification of forest data with remotely sensed images*

With the advancement of satellite remote sensing technologies[20] forest and agricultural land observation has become very convenient. These sensors turn out probably helpful data in huge amount on daily basis. This amount of data conjointly presents a challenge of data interpretation and classification to the researchers. Researchers are wide victimisation techniques like K-means, K-nearest neighbor and artificial neural network for classification remotely detected pictures. Reese et. al used K-nearest neighbour algorithm for estimation of

forest parameters like wood volume, age and biomass. They used satellite data, digital mapped data and forest inventory data for classification.

V. PROPOSED SYSTEM ARCHITECTURE

The objective of System for Agriculture Recommendation System using data mining is to enhance the assembly of crops moreover as improve the economic conditions of farmers. At identical time using the available resources optimally and efficiently in today's critical situation of natural resources like lack of availability of sufficient water, electricity and degrading quality of fertile land.

The system can extracting databases for crop choice once farmer needed and sends that data first off to the user's computer and then the message to farmer. The system can mining the suitable crop for explicit farm of a farmer and generates results in user compatible format.

VI. SMS ALERTS TO FARMER

The system can send SMS alerts to the farmer regarding the crops appropriate for farm supported the mining result. It's going to contain one or additional crop for selection however the priority is given to only that crop that is additional appropriate. Therefore he can sow or plant these crop without worrying regarding production of crops in field or environmental condition changes.

VII. CONCLUSION

By understanding soil and climate conditions by using these data mining system farmers are going to be able to take right crop at right place which can improve yields.

REFERENCES

- [1] A. Mucherino, P. Papajorgji, P.M. Pardalos, A Survey of Data Mining Techniques Applied to Agriculture, *Operational Research: An International Journal* 9(2), 121–140, 2009
- [2] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Estimation of neural network parameters for wheat yield prediction. In Max Bramer, editor, *Artificial Intelligence in Theory and Practice II*, volume 276 of IFIP International Federation for Information Processing, 109–118. Springer, July 2008
- [3] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Optimizing wheat yield prediction using different topologies of neural networks. In Jos é Luis Verdegay, Manuel Ojeda-Aciego, and Luis Magdalena, editors, *Proceedings of IPMU-08*, 576–582. University of M álag a, June 2008.
- [4] Georg Ruß, Rudolf Kruse, Peter Wagner, and Martin Schneider. Data Mining with neural networks for wheat yield prediction. In Petra Pernert, editor, *Advances in Data Mining (Proc ICDM 2008)*, 47–56, Berlin, Heidelberg, July 2008. Springer Verlag.
- [5] M. Kovacevic, B. Bajat, B. Gajic, Soil Type Classification and Estimation of Soil Properties using Support Vector Machines, *Geoderma* 154(3–4), 340–347, 2010
- [6] Cover TM, Hart PE, Nearest Neighbor pattern classification. *IEEE Trans Info Theory* 13(1) : 21-27, 1967
- [7] J. Hartigan, *Clustering Algorithms*, John Wiles & Sons, New York, 1975.
- [8] A. Mucherino, A. Urtubia, Consistent Biclustering and Applications to Agriculture, *IbaI Conference Proceedings, Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop "Data Mining in Agriculture" (DMA10)*, Berlin, Germany, 105-113, 2010.
- [9] A. Mucherino, S. Cafieri, A New Heuristic for Feature Selection by Consistent Biclustering, *arXiv e-print, arXiv:1003.3279v1*, March 2010
- [10] Jorquera H, Perez R, Cipriano A, Acuna G Short term forecasting of air pollution episodes. In: Zannetti P (eds) *Environmental modeling 4*. WIT Press, UK, 2001.
- [11] Rajagopalan B, Lall U, A K-Nearest Neighbor simulator for daily precipitation and other weather variables. *Wat Res Res* 35(10) : 3089–3101, 1999.
- [12] Tripathi S, Srinivas VV, Nanjundiah RS Downscaling of precipitation for climate change scenarios: a Support Vector Machine approach. *J Hydrol* 330:621–640, 2006.
- [13] Fagerlund S Bird species recognition using Support Vector Machines. *EURASIP J Adv Signal Processing*, Article ID 38637, p 8, 2007.
- [14] Holmgren P, Thureson T Satellite remote sensing for forestry planning: a review. *Scand J For Res* 13(1):90–110, 1998.
- [15] Das KC, Evans MD Detecting fertility of hatching eggs using machine vision II: Neural Network classifiers. *Trans ASAE* 35(6):2035–2041, 1992.
- [16] Patel VC, McClendon RW, Goodrum JW Crack detection in eggs using computer vision and neural networks. *Artif Intell Appl* 8(2):21–31, 1994
- [17] Du C-J, Sun D-W Pizza sauce spread classification using colour vision and support vector machines. *J Food Eng* 66:137–145, 2005
- [18] Karimi Y, Prasher SO, Patel RM, Kim SH Application of support vector machine technology for Weed and nitrogen stress detection in corn. *Comput Electronics Agricult* 51:99–109, 2006.
- [19] Albrecht, W.A., Smith, G.E. *Biological Assays of Soil Fertility*. Soil. Sci. Soc. Amer., Proc. 1942;6:252
- [20] Keil, M., Schardt, M., Schurek, A. and Winter, R., 1990. Forest mapping using satellite imagery. the regensburg map sheet 1:200.000 as example. *ISPRS Journal of Photogrammetry and Remote Sensing* (45), pp. 33–46.