Survey: Data Mining Techniques Used In Agricultural Field

T.Pavithra[1], Dr.A.Dhanalakshmi[2]
[1]Dept of Computer Science
[2]Assistant Professor, Dept of Computer Science
[1, 2] Gobi Arts & Science College, Gobichettipalayam, Gobi.

Abstract- *This paper shows that the different crop yield prediction methods with the use of data mining techniques. The agricultural system is very complex since it deals with large data situation which comes from a number of factors. Different data mining techniques are in use, such as k-means, k-nearest neighbor (KNN), Artificial Neural Network (ANN) and Support Vector Machine (SVM) for very recent application of data mining techniques in agriculture field. Data mining is an important tool for extracting invisible information from large data. The paper explores the chance out of applying different types of mining algorithms for extract the important information.*

Keywords- Agriculture, Yield Prediction, agricultural productivity, Classification, Clustering, k-means, k-nearest neighbor, Artificial Neural Network and Support Vector Machine.

I. INTRODUCTION

Agriculture is one of the back bones of our country. Everywhere the agricultural productivity is very low. Farmers, agricultural scientist, researchers are trying to improve the food productivity. Because agriculture is the main need for human being and animals. The real income of the country is yield of the agriculture. This yield is measured based on the site and production of each crop in the country. To get a better economic result they must study the agriculture data base with information interconnected to crop, season, surrender, region, etc.

Analysis of huge agriculture data base is done through data mining techniques. Agricultural specialist systems are being used extensively nearly in every walk of life. Different tools have been implemented for assessment, justifying, improve and alter the previous agricultural specialist systems thus making them more useful in their deliberate reason. Large amounts of agricultural information are made available by various government organizations, for agricultural planning. Data mining tools can be employed to forecast trends about agricultural return and yield.

II. LITERATURE SURVEY

Sally Jo Cunningham et.al discusses the procedure model for examining data and contemplates the WEKA (Waikato Environment for Knowledge Analysis). The domain model 'scholarly' by the data mining algorithm can then be willingly absorb into a software application. In this work, they concentrate on machine learning techniques for persuade empire models or survey datasets.

Saiyyad Mohmmad Ali et.al researcher survey is used to distribute with study of agriculture sector to supply successful path to develop making and powerful use of resources which in turn develop the economy. Machine Learning algorithms could be used to forecast the proper crop so that it will conduct to less loss in insert and develop the profits. To deal with agriculture more algorithms be used to address a particular district successfully.

Sneha N et.al they focused on the survey of rice yield production and summarize the data mining techniques are Chameleon, Random Forest, regression Techniques to derived the optimal parameters required by rice. Also discusses data mining techniques like regression and clustering which is a procedure model for surveying data and report the carry that SPSS (Statistical Package for the Social Sciences) supply for this model. Cluster analysis or clustering are used in the task to allocate a set of objects into type so that the objects in the cluster is more equal to every other to those in later clusters.

Hooman Fetanat et.al used to differentiate and survey the soil data with the use of DM techniques are GATree (Genetic Algorithm), Fuzzy Classification rules and Fuzzy C-means algorithm for classifying soil texture in cultivation soil statistics. This research is used to increase pattern recollection and survey of huge soil side view exploratory datasets. Soil is very important for agriculture and it influences fecundity, drainage, water belongings capacity, aeration, tillage and power of soils.

D. Sabareeswaran et.al plant illness detection techniques, soil wetness prediction methods and crop

extension monitoring techniques are deliberate and differentiate for analyzing the production. It is clear that the crop extension should be increase by plant illness prediction, extension monitoring techniques. The difficulty in plant illness prediction and extension monitoring structure for supplement crop production are developed by optimization techniques in data mining.

V. Sellam et.al Regression Analysis (RA) is used to examine the environmental element and their infliction on crop surrender. RA is a multivariate examination technique which inspect the factors classification them into explanatory and reaction variables and assist to obtain a resolution. Linear Regression (LR) is used to start connection between explanatory variables (Annual Rainfall (AR), Area under Cultivation (AUC), Food Price Index (FPI)) and the crop surrender as response shifting. R2 value distinctly shows that yield is largely dependent on AR. AUC and FPI are the other two components influencing the crop capitulate.

III. DATA MINING TECHNIQUES

CLUSTERING

A clustering is a collection of data elements that are highly similar to one another within the same cluster, but weakly similar from the data elements in other clusters. In [6], Data mining in agriculture is comparatively new investigation field and the use of cluster survey has almost been just begun in this region. Some clustering techniques are hierarchical agglomerative clustering, fuzzy clustering, hierarchical clustering and kohonen self-organizing characteristic plan are discussed because these are the widely used data mining methods in the meadow of agriculture and allied science.

CLASSIFICATION

Classification is very difficult data mining techniques, which compel to collect different attributes together into discernable categories. It recognition of new design with before defined goals; These tasks are predictive and require the construction of model in the point of forecast a goal, or dependent, variable from the set of describe, or individualistic variables. Classification is the process of discovering the purpose for classifying data in one of some classes. For job of classification the aim variable normally has a small integer of separate values, while with jobs of regression the aim variable is continual.

REGRESSION

Regression, used primarily as a form of arrangement and modeling, is used to identify the likelihood of a definite variable, given the existence of extra variables. The survey is used to examine and control the connection between response variable and explanatory variable. The variables contemplate for analyses in this investigation effort are Annual Rainfall (AR), Area under Cultivation (AUC), Food Price Index (FPI). Crop yield is a dependent changeable which depends on all these ecological elements.

PREDICTION

It is one of the most valuable data mining techniques; to project the different types of data are used in the future. In many places, just recognizing and understanding historical trends is enough to chart a somewhat correct forecast of what will happen in the future. For example, one might enquiry buyer solvency histories and finished buy to imagine whether they will be a credit chance in the later.

PREDICTING YIELD PRODUCTION

In [8], Yield prediction is a very main agricultural difficulty. Some agriculturist would like to know how considerably they can anticipate the yield. Attempts to resolve this difficulty data back to the hour when initial agriculturist got gain by working on varies soils. Since years, yield predictions have been executed by considering farmer's incident on specific fields and crops. However, this understanding can also be obtained by exploiting statistics given by present technologies, such as GPS. A multitude of sensor data can nowadays be comparatively comfortably calm, so that farmers do not only congregate crops but also growing quantity of data.

Examine that sensor data are obtainable for some time back to the over, where the corresponding yield performance have been noted. Fully this statistics form a instruction set of data which can be utilize to absorb how to classify later yield performance, previously new sensor data are obtainable. Moreover, to better quality of the forecast in sequence, the idea of spatial autocorrelation has further recently been contemplated.

Consequently, the spatial connection between data records should be taken into report. Classical sampling methods do not remove spatial neighborhoods of data documentation into report. The method that follows can be seen as a spatial cross-validation technique. In common, when considering the k-fold cross-validation technique, the indigenous dataset can be split in three parts: a training set, a validation set and a test set. Setting k equal to 10 or 20 is

normally considered to be suitable to detach bias. The regression model is instructed on the training set till the prediction mistakes on the validation set begin to increase.

Formerly this happens, the training procedure is concluded and the mistake on the test set is announced for this fold. In spatial data, required to spatial autocorrelation, nearly identical data evidence may conclusion up in training, validation and test sets. In quintessence, the model above fits the training data and returns an overoptimistic (biased) approximation of the forecast mistake. This can be attaining by modifying the sampling policy for spatial data. Formerly this issue is accommodated, the cross-validation system can pursue in the customary way. A spatial clustering strategy can be engage to subdivide the green into spatially disjoin clusters or sector.

The clustering algorithm can then is hurry on the data records' spatial map; utilize the data records' longitude and latitude. Depending on the clustering algorithm framework, this consequence in a tessellation map which does not appraise some of the attributes, but only the spatial Neighborhood between data information. In analogy to the non-spatial regression treatment of these data records, a spatially aware cross-validation regression difficulty can therefore be handled apply the k resulting zones of the clustering algorithm as an insert for k-fold cross-validation. These assure that the training set has only a little quantity of spatial autocorrelation with the test set.

Worth models can be used straightforwardly, without need changes to the models themselves. Initial computational trial can be found, which show that it is literally important to closely contemplate spatial relationships inherent in the data sets in this sort of data mining difficulty. This effort demonstrates that, if spatial autocorrelation survive, standard regression models should be adjusted to the spatial case.

IV. CONCLUSION

This topic is a trending concept and has a very much helpful towards the agriculture and data mining plays the main role where clustering and k-means, k-nearest neighbor, Artificial Neural Network, Support Vector Machine algorithms are extensively being used. To develop the income of the farmland, classification and prediction techniques of data mining are used. For agricultural and environmental fields knowledge discovery techniques are comparatively new. This paper concludes that data mining techniques become very active research in agricultural developing process.

REFERENCES

[1] Sally Jo Cunningham and Geoffrey Holmes "Developing innovative applications in agriculture using data mining".
[2] Saiyyad Mohmmad Ali, Muzffar Ali "A Survey on Agriculture Crop Recommender System".
[3] Sneha N, Dr. Jharna Majumdar "Big Data Application in Agriculture to Maximize the Rice Yield Crop Production using Data Mining Techniques".
[4] Hooman Fetanat, Leila Mortazavifarr ,Narsis Zarshenas "The analysis of agricultural data with regression data mining technique".
[5] P. Bhargavi, Dr. S. Jyothi "Soil Classification Using Data Mining Techniques:A Comparative Study".
[6] Tiwari, Dr. Bharat Misra "Application of Cluster Analysis in Agriculture–A Review Article" international Journal of Computer Applications (0975 –8887)Volume 36–No.4, December 2011-Mamta.
[7] Namita Mirjankar, Smitha Hiremath "Application of Data Mining In Agriculture Field" International Journal of Computer Engineering and Applications,iCCSTAR-2016,Special Issue,May2016.
[8] B. Milovi and V. Radojevic "Application of Data mining in Agriculture" Bulgarian Journal of Agricultural Scienve,21(No 1)2015,26-34 Agricultural Academy.
[9] D. Sabareeswaran, A. Edwin Robert "A survey on data mining techniques in agriculture".
[10] V. Sellam and E. Poovamma "Prediction of Crop Yield using Regression Analysis".