IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

A Review Technique of Data Mining in the Agronomy field

U. Singh[1*], K. Garg[2]

[1,2]Department of Computer Science and Application, Kurukshetra University, Kurukshetra, Haryana, India

**Corresponding: singhurvashi73@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 18/May/2018, Published: 31/May/2018

***Abstract*—** Agriculture is the main source of income in India. India produces various crops such as jute, wheat, rice, sugarcane, cotton, mustard, pulses and many more. Today India acquires 2nd rank worldwide in farming production. In 16 century, 99% of world crop productions are produced by India, Now, it is only produces 23% of agriculture products. However, there are some critical factors that influence the agriculture such as in efficient use of fertilizers, Chemicals, Heavy Rainfall, Degradation/Increment in temperature, PH value, less components value present in soil. This paper provides a systematic analysis by employing data mining techniques. This paper includes some techniques of Supervised and Unsupervised learning methods. In Agriculture domain, multiple linear regressions (MLR), Support vector machine (SVM), K nearest neighbor (KNN), Density-based spatial clustering applications with noise (DBSCAN) are the most widely used data mining algorithm which aimed to solve the issues of the agriculture up to some extent.

***Keywords*—** *Agriculture, Yield prediction, Data Mining, MLR, KNN, DBSCAN, SVM, Clustering and Classification.*

I. INTRODUCTION

Motivation: The history of farming goes back to Indus Valley Civilization Era. Agriculture and its allied sectors accounted 17.32% GDP in 2017. Slow agriculture growth is an important regard, as 2/3 people in the country depend on agriculture. The present situations are not good as per environmental concern. With advanced technology, the agriculture production is focused toward growth which can be achieved only by high yield. The yield of crops has not increased due to inexact use of fertilizer, less knowledge of soil type, climatic conditions. Earlier crop yield prediction, helps a farmer to know the exact condition of the land.

The process of fetching out the information using the collected data is termed as KDD. The knowledge discovery in databases process has 4 stages described below in fig 1 such as:

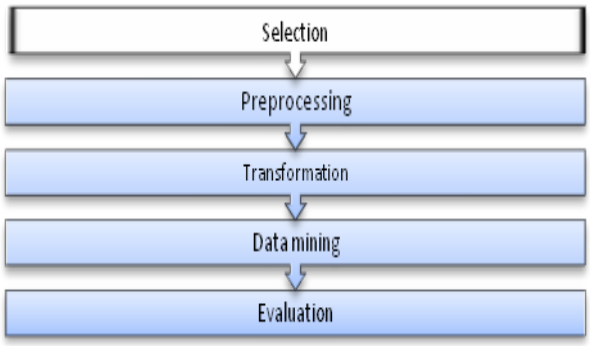

Fig 1: Stages in Data Mining

In Computer Science, the huge amount of raw data stored in a database, and getting the necessary information from that data termed as data mining. It is basically a process of knowledge discovery, analyzing the data from the dataset using statistical, mathematical, pattern recognition technique. In agriculture field, data mining is a relatively novel approach for research. Every day the farmers have to make an innumerable decision in the agriculture sector. An accurate yield prediction for the crop is an essential issue for agriculture planning. This makes farmers able to take a decision and this prediction not only helps out the farmers but also government agencies, to frame policies which can be made according to food chain supplies.

Throughout the year's many algorithms were created, adopted, developed and modified to extract the knowledge from the datasets. The overall goal is to extract the knowledge and transform it into decision making. In this paper Section 2 shows the literature survey, Section 3 describes about various techniques till used in the field of agronomy using data mining, Section 4 includes all the problems which can be addressed later for implementation and last but not least present the conclusion.

II. RELATED WORK

The basic construct of data mining is knowledge elaboration. To figure out the widely used data mining techniques which have been implemented earlier are shown below such as:

A lot of research accounted that development in agriculture field can be achieved by high yield. So, prediction the production of the crop was as a pleasing research area. D Ramesh[1] for estimating outcome crop for the Andhra Pradesh selected region of district East Godavari were used different techniques such as MLR and density-based clustering. The received results were verified, analyzed and compared.

Another area of research in the agriculture is the extraction of hidden knowledge. Using Kharif season of India, Nikita Gandhi [2] presented research that focused to extract knowledge; by applying data mining technique and Statistical methods. The success of rice crop production in a particular region or climatic zone could decide with the variations in climatic factors, precipitation, soil, and hydrology. The exploratory output includes the sensitivity, specificity, accuracy, as the main concern. The outcome showed that best performance was provided by J48 and LADTree classifiers.

The crop recommendation system helps out in selecting a suitable crop for sowing. Monali Paul et al. [3] presented a system, in order to analyzed soil datasets which used the data mining techniques. It used various classification rules such as Naïve Bayes, K-NN methods, and concluded that soil was classified as low, medium and high in form to estimate the outcome of the product using available dataset.

The analysis of the crop was done to obtain better results. Jharna Majumdar et al. [9] showed model on the analysis, to maximize the productivity in which most favourable parameters were found by enforced the techniques of data mining like PAM, CLARA, DBSCAN, and MLR. Compared clustering methods using quality metrics and according to the analyses, CLARA gave the better quality than PAM, DBSCAN gave the better result as compare to PAM and CLARA.

The feature extraction was used for estimating the land area. Pooja G. Mate et al. [16] proposed a task for classifying complex images that were collected through GPS. In this paper feature selection and optimization techniques were observed using SVM.

The climatic factor played a vital role and can also affect the sensitivity, accuracy. Kuljit Kaur[18] introduced the model to predict the essence of temperature and rainfall which exploited mining techniques such as apriori algorithm and concluded that during the vegetative phase with increment in temperature the yield of rice increased where rainfall did not influence the yield.

III. METHODOLOGY

Data mining is the analysis thread of the "knowledge revelation from databases"[25]. The techniques are divided into two board order as Predictive and Descriptive. Now, discussed some of the methods used in different papers for achieving the objective of agriculture field such as:

3.1 Multiple linear regressions: A model that has more than one foreteller attributes. For the analysis, multiple linear regressions were adopted. A study by [1] presented that MLR techniques have one dependent variable known as production and seven independent variables named area of sowing, yield, year, rainfall, and three types of fertilizers. The result of the technique ranged in -between -14% and +13% for the 40-year interval.

3.2 Classification: It is generally mentioned as "supervised learning", has a predefined set of group or pattern founded on that, predicts rate. From [2] know that from the existed data it helps to take put rules and value that could be used for prediction.

3.3 Naïve Bayes: It produced a probabilistic rule that is, when introducing a new data item, the model indicates the probability that the item belongs to each of possible class categories. A naïve Bayes classifier is a simple probabilistic classifier based on applying theorem with strong independent assumptions [4] and demonstrated a comparative study in classification algorithm to predict the accuracy of the soil fertility.

3.4 KNN: KNN can be used for both classification and regression. The sample set is classified based on the "closeness". A study by [5] used IBk and LWL algorithm for KNN.

3.5 Clustering: It is widely called as "unsupervised learning". The data belongs to one another are clustered together and outliers in the different cluster. They have greater intra- cluster similarities and lesser inter cluster similarities [6].

3.6 ANN: The ANN is a neural network, which as sensory layer, association layer, output layer. ANN in agriculture to develop agronomic based model. A study [8] examined the performance of ANN in Nepal to predict the rice crop.

3.7 DBSCAN: The spatial clustering with noise is mostly used the density-based algorithm. It uses the concept of density connectivity, density reachability. The study [9], calculated the eps and min-pts value, where eps value was used to check the temperature, rainfall and soil type.

3.8 SVM: Support vector machine, is capable to arrange samples of data into distinct clusters. SVM is supervised

learning method that it can build a model that fore see either a new example falls into the family or in another. All conclude that SVM could be used for yield prediction. Another study by [11] the goal was to predict the rice yield and produced the best result for Aman, the variety of rice. A study by [21] showed the rice crop prediction for Maharashtra.

IV. POTENTIAL ISSUES:

Agriculture is the keystone of Indian economy. Yet agricultural resources are not good in India, mostly rely on external sources and in the arena, there are various causes of failure some of them are addressed below such as:

4.1 Estimate Outcome: The estimation of the crop outcome is necessary not only for farmers but also for the agribusiness in taking the decision. The study's by [11] showed the data mining techniques such as MLR, SVM were used for estimating the future year yield of the crop and will be dedicated to use suitable approaches for improving the efficiency of proposed techniques.

4.2 Knowledge Driven Decision: The extractions of hidden knowledge are major metrological and agronomic factors. From [6], studies the k- means clustering and various decision tree classifiers were implemented to make a knowledge-driven decision before harvest.

4.3 Crop Recommendation System: These helps the farmers in selecting the best seed for the soil to increase the productivity. [12] have the mining learning models such as a random tree, k nearest neighbor, Naïve Bayes. So that the overall productivity can be increased and the future work is aimed to increase the number of attribute of the datasets.

4.4 Policy Decision: In the agriculture sector, decision making has of great importance for policymakers. From [7], studies the classification techniques and made policies using Apriori algorithm and k nearest neighbor and the work can be extended by implementing another algorithm in order to the best result.

4.5 Accuracy of Soil Fertility: The fertilizer which is necessary for crop growth can sometimes become a reason of fall down because its inexact quantity leads to decrement in the outcome. The [23] showed result based on the classification which can be present in the soil. The proposed work can be stretched to implement other crop factors to increase the crop production.

4.6 Analysis and Prediction: The [8] aimed to analyze agriculture data using data mining techniques such as K-means, DBSCANand to find the useful information which should improve the yield of agriculture. The work can be carried out further by applying advanced mining techniques to larger datasets.

4.7 Feature Selection in Agriculture: It is the process of selecting a subset of relevant attributes for use in the construction of a model. In [16]different feature selection techniques were implemented for classifying satellite imaginary and SVM provided the best result in a classification of remotely sensed data. The upcoming work aimed to implement the feature selection for decision making in the field of agriculture.

4.8 Normalizing the Data: The normalized data is done to remove the noise data, sparse data. In this, the min-max normalization applied on the data set. This kind of work was not implemented much in the field of agriculture so, it can be carried out in the agriculture field for attaining accuracy.

V. CONCLUSION AND FUTURE SCOPE

This review provides the description on the data mining techniques like DBSCAN, MLR in concern of farming. The review sketches the several parameters such as Policy Decision, Recommendation System, and Accuracy of Soil Fertility. From all above, consolidate that much work are to be needed to be done to know how these techniques can be implemented to improve the Decision Making and Productivity. The work can be carried on the arena of agriculture for normalizing the data, Dimension Reduction and Feature Extraction.

REFERENCES

[1] D Ramesh, B Vishnu Vardhan, “Analysis of Crop Yield Prediction Using Data Mining Techniques”, International Journal of Advanced Research in Engineering and Technology.

[2] Nikita Gandhi, Leisa Armstrong, “Applying Data Mining Techniques to Predict Yield of Rice in Humid Subtropical Climatic Zone of India”.

[3] Monali Paul, Santosh K. Vishwakarma and Ashok Verma, “Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach”, IEEE.

[4] Dr. S. Hari Ganesh, Mrs. Jayasudha, “Data Mining Technique to Predict Accuracy of the Soil Fertility”,International Journal of Computer Science and Mobile Computing.

[5] Nikita Gandhi, Leisa Armstrong, “Rice Crop Yield Forecasting of Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques", 2016 3rd International Conference, Computing for Sustainable Global Development, Pg: 357-363.

[6] Anitha, “A Predictive Modeling Approach for Improving Paddy Crop Productivity using Data Mining Techniques”, Turkish Journal of Electrical Engineering and Computer Science 2017, Pg: 4777-4787.

[7] G.Nasrin Fathima, R.Geetha, “Agriculture Crop Pattern Using Data Mining Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, Issue 5,May 2014, Pg: 781-786

[8] Vinayak A. Bhardi, Prachi P. Abhyankar, Ravina S.Patil, Sonal S. Patade, Tejaswani U.Nate, Anaya M.Joshi, “Analysis and

Prediction in Agricultural Data using Data Mining Techniques" IJRISE.
[9] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, "Analysis of Agriculture Data using Data Mining Techniques", Journal of Big data, Springer Open 2017, Pg: 3-15.
[10] Sally Jo Cunningham, Geoffrey Holmes, "Developing Innovative in Agriculture Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering.
[11] Umid Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin, "Rice Yield Prediction Model Using Data Mining Techniques", International Conference on Electrical, Computer and Communication Engineering, February 16-18, 2017, Pg: 321-326.
[12] S. Pudumalar, E.Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, J. Nisha, "Crop Recommendation System for Precision Agriculture", 2016 IEEE EIGHT International Conference on Advanced Computing, Pg: 32-36. [13] Neetu Chahal, Anuradha, "A Study on Agricultural Image Processing along with Classification Model", 2015 IEEE International Advance Computing Conference, , Pg: 942-947.
[14] Pallavi V. Jirapure, Prarthana A. Deshkar, "Qualitative Data Analysis Using Regression Method for Agricultural Data", 2016 IEEE.
[15] Nikhil Sethi, Dr. Kanwal Garg, "Exploiting Data Mining Technique for Rainfall Prediction", International Journal of Computer Science and Information Technologies, Vol. 5(3), 2014, Pg: 3982-3984.
[16] Pooja G. Mate, Kavita R. Singh, Anand Khobragade, "Feature Extraction Algorithm for Estimation of Agriculture Acreage from Remote Sensing Images", 2016 IEEE, Pg: 1-5.
[17] Chowdari K.K, Dr. Girisha R, Dr. K C Gouda, "A Study of Rainfall Over India Using Data Mining", 2015 IEEE Sponsored International Conference on Emerging Research in Electronics, Computer Science and Technology, Pg: 44-47.
[18] Kuljit Kaur, Kanwakpreet Singh Atwal, "Effect of Temperature and Rainfall on Paddy Yield using Data Mining", 2017 Seventh International Conference on Cloud Computing, Data Science and Engineering, Pg: 508-511.
[19] Raorane A. A., Kulkarni R. V., "Data Mining: An Effective Tool for Yield Estimation in the Agriculture Sector", International Journal of Emerging Trends and Technology in Computer Science, July- August 2012, pg.75-79.
[20] P. Hariharan, K.Arulanandham, "Design an Disease Predication Application Using Data Mining Techniques for Effective Query Processing Results", Advances in Computational Sciences and Technology.
[21] N. Gandhi ,L.Armstrong, O.Petkar and A. Tripathy, "Rice Crop Yield Prediction in India Using Support Vector Machine", In IEEE, 2016.
[22] T. Ranjeet and L. Armstrong, "An Artificial Neural Network for Predicting Crop Yield in Nepal", Ninth Conference for Information Technology in Agriculture" ICT's for future Economic and Sustainable Agricultural Systems", Perth, Australia.
[23] S. Jabjone and S. Wannasang, "Decision Support System Using Aritifical Neural Network to Predict Rice Production in Phimai Thailand", IJSCM.
[24] Mehmed Kantardzic, "Data-Mining Concepts", Edition:1, Copyright year: 2011.
[25] Mrs. Bharati, M. Ramageri "Data Mining Techniques and Applications", IJCSE
[26] Fuzail Misarwala, KausarMukadam, and Kiran Bhowmick, "Applications of Data Mining in Fraud Detection", IJCSE.

Authors Profile:

Urvashi Singh pursed Bachelor of Technology from University of Kurukshetra, Kurukshetra in batch of 2012-2016. She is currently pursuing Mtech from Department of Computational Sciences and Applications, Kurukshetra University, Kurukshetra. Her research area is data mining.

Dr. kanwal Garg MCA, Phd. He is assistant professor in Kurukshetra University, Kurukshetra.