AN ANALYSIS ON CROP YIELD PREDICTION USING DATA MINING TECHNIQUES

S. Kavitha[1], G. Sangeetha[2]

ABSTRACT

In day to day life the requirement of food is increasing at rapid rate and hence the farmers, government and researchers are using several techniques in agriculture for the improvement in production. Plants are usually affected by many pests and diseases. In the process of resolving agricultural issues the concepts of data mining play a fundamental part. Research in agriculture is increasing due to development of technologies and forthcoming challenges [1]. In improving the general growth of a country the plant disease detection has an important place. Diseases in plants and production loss can be predicted with the help of data mining approaches like classification. Future trends in agricultural processes can be forecast with Data mining techniques. Generally damages are examined by using classifiers namely SVM, K-Nearest Neighbor, Decision Tree, Random Forest, Naive Bayes and so on. [13].

Keywords **:** ***Agriculture, Classification.***

I. INTRODUCTION

Food safety is measured by calculating crop yield. The backbone of the economy is agriculture. This crop yield basically depends on many conditions like climate, biology, geography and economy.

Data mining is the collection of techniques which are used to find some meaningful patterns, information and knowledge from the existing data [3]. To maximize productivity the given data should be examined thoroughly to get significant results. As predicted farmers are not getting good yield due to various reasons [2]. The amount of data is also massive in case of agriculture.

Data mining techniques give better results to resolve the issues in predicting challenges, identifying diseases and making timely decisions to improve production. In the process of extracting the data patterns using the various intelligent methods, data mining plays a major role.

Data mining process, also known as KDD, basically follows seven steps as in Figure 1 [7].

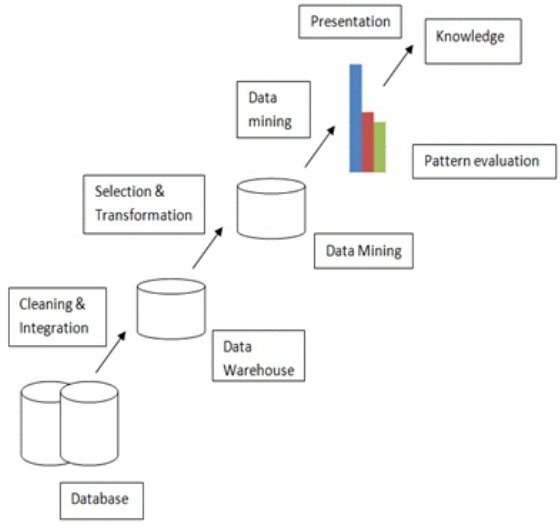

Fig 1 : Data mining process

II DATA MINING TECHNIQUES

Data Mining is stated as the process of discovering patterns or knowledge by exploring a huge set of data and selecting the useful data for calculating future

[1] *Assistant Professor, Department of BCA, PSGR Krishnammal College for Women*

[2] *Assistant Professor, Department of BCA, PSGR Krishnammal College for Women*

development. The available data need to be converted to meaningful information, which will be later used for many applications.

There are broadly two functions involved in mining the data.

The first one is a descriptive model and the next is predictive model. The method of finding the patterns in data and then discussing the property or behavior of those data is the function of descriptive model. The practice of predicting a representation or a model explains the classes of the data and identifies, with the help of this model, the objects of unknown class label [4].

Classification and Clustering are the two major techniques in data mining.

The process of Classification follows two steps:

1) Training, 2) Testing.

Classification

A classification model is built by training steps based on the data collected for generating classification rules. The test data estimate the classification rules and the accuracy of the classification model built is based on these rules. It is termed as training set because it is specifically used only to train the classification technique.

If there is no training set then it intimates that no previous knowledge about those data exists for classification [6].

Clustering

It does not have any predefined classes. A vast database is separated into a number of small groups based on similarities of the data present and they are called clusters. Objects in the same clusters are similar to one another compared to other clusters.

The clustering techniques are followed in those cases to split the samples which are not known into specific groups [6].

III METHODS

A. Classification Techniques

i) Artificial neural network

The back-propagation neural network (BPNN) is generally used for calculating the yield of crops because its structure is simple [7]. The nodes' weight connects the neurons of the adjacent layers. There are three layers in BPNN and two weights in those layers.

The weight keeps on changing till the accurateness of the model is obtained based on the difference between the actual values and the output.

After frequent revisions the weights are determined, and based on this new data are predicted. The count of neurons in the hidden layer is usually predicted by means of trial and error method. Every network has three layers [refer fig.4].

i) Input layer: It has the raw information which is provided to the network.

ii) Hidden layer: mapping of input information to the output class is done here.

iii) Output layer: It gives the value of target class as predicted by the trained neural net [8].

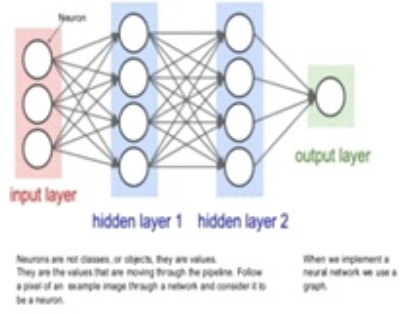

Fig. 2: Artificial Neural Network

ii) Linear Regression

Linear Regression tries to fit a straight line to a dataset and tries to reduce errors between the points and the fitted line. On the basis of this measure, potency of the relationship between one dependent variable and independent variables is determined. The relationship

between a scalar variable and another variable is proved by linear regression methods. By fitting the linear equation it can be carried out for the observer.

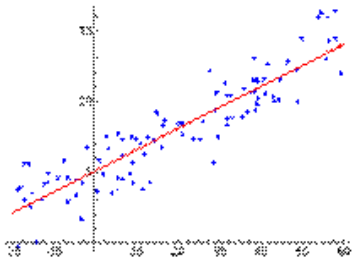

Fig. 3: Linear Regression

iii) k-Nearest Neighbor (KNN)

Here the process starts by comparing the test examples and the training examples which are almost identical. It stores all cases and classifies new cases on the basis of similarity measures. K will be a positive integer here. The training examples are saved in n-dimensional space because the samples indicate points in n dimensional space. The simple and basic k-NN algorithm is applied here.

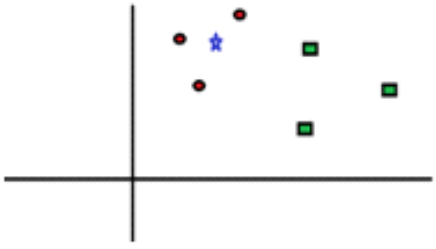

Fig. 4: Classification in K-NN

B. Clustering Techniques

i) K-means clustering

The basic principle of K-means clustering algorithm is that it works on the basis of centroids and generates a non-hierarchical group of like points. The centroids as kept as small as possible by first It finds out the k number of centroids and then assigns each and every data point to the nearby cluster.

The *means* here indicates the averaging of data. The process starts with the randomly selected centroids and then cyclic calculations are performed to upgrade the positions of the centroids.

The centroids will be stabilized if there is no change in their values because the clustering has been successful [9].

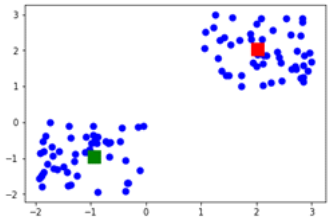

Fig. 5: K-means Clustering

iv) Density-based Clustering

The main functionality here is that for each and every point in the cluster the contiguous cluster must contain minimum number of points.

It means that the density in that region should attain a minimum value. It is assumed that the clusters are in regular shapes. On the basis of specific density functions the objects are grouped. The number of objects in given area of data objects is denoted as its density.

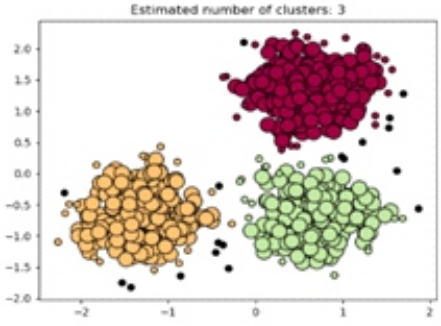

Fig. 5 : Density-based Clustering

S.No	Article Title	Concept/ Objective	Methodology used	Summary
1	Association Rule s Algorithm based on Clustering in Corn Yield [14].	To study the relation between soil and corn yield.	K-means, Hierarchical , PAM, Association Rules	K-means proves to be the best
2	Crop Yield Prediction using Data Mining Approach.	Prediction of the yield of the crops	Naive Bayes	**Crop prediction system** 1.Crop database 2. Input data 3. Preprocessing 4. Feature selection 5.Classification algorithm (naive Bayes) 6. Result
3	Crop Disease Prediction Using Data Mining Method [15].	Predicting the loss due to grass grub insect dataset.	SVM, KNN, NB Ensemble models.	Better performance in Naive Bayes.
4	Prediction of Plant Disease from Weather Forecasting [10].	Predicting the disease in orange plant.	k-means clustering , neural network	Data collection & Segmentation 1. clustering 2. neural network 3.Spatial Grey Level 4.Dependence Matrix 5.Pre-processing 6.Feature extract
5	Plant disease detection and its solution using image classification.	Leaf disease prediction	1.K-means clustering algorithm 2.Image processing algorithm	Steps 1.image acquisition, 2.image preprocessing, 3.image segmentation, 4.feature extraction classification
6	Survey of data mining Techniques Applied to application of Agriculture [11].	To select most applicable method for prospect	Classification Techniques & Clustering Techniques	Steps 1. Linear Regression 2. Regression Tree 3.Artificial neural network 4. KNN 5.SVM
7	Seasonal Crops Disease Prediction and classification Using Deep Convolution Encoder Network	Detecting crop diseases using the encoder networks.	Back propagation learning algorithm	1.CNN and auto encoders 2. Deep learning neural network

Table 1: A Comparative Study of techniques used in data mining

V. CONCLUSION

A challenging issue for farmers is predicting crop yield. From this research work it is concluded that there is still need for examination in the field of agricultural to get better accuracy. Particularly in developing countries agriculture is the most important application area. With the help of information technology and its role in agriculture there can be predictable changes in decision making and hence the farmers can get better yield.

Different data mining applications are focused on solving different problems in the agricultural domain. On the available data the Multiple Linear Regression technique is applied and then the results are verified using Data Mining techniques. Based on the predictions made by the research farmers can cultivate different types of crops in different areas and can increase their productivity. In future more indicators need to be included in order to examine the variability in crop yields at different locations.

VI. REFERENCES

[1] Patel, Hetal and D. Patel, "A brief survey of data mining techniques applied to agricultural data", IJCA, 2014.

[2] Ami Mistry, Vinita Shah, "Brief Survey of data mining Techniques Applied to applications of Agriculture", IJARCCE, 2016.

[3] P. Berkhin, Kogan, Jacob; Nicholas, Charles "A Survey of Clustering Data Mining Techniques"; Marc (Eds) Grouping Multidimensional Data, SpringerPress (2006) 25-72

[4] D Ramesh, B Vishnu Vardhan, "Analysis Of Crop Yield Prediction Using Data Mining Techniques", IJRET, 2015.

[5] B. Milovic and V. Radojevic, "Application of data mining inagriculture", Bulgarian Journal of Agricultural Science, 2017.

[6] Prof. Rakesh Shirsath, Neha Khadke, "Agriculture decision support system using data mining", 2017 International Conference on I2C2.

[7] S.Pudumalar, E.Ramanujam, "Crop Recommendation System for Precision Agriculture", IEEE Eighth International Conference on Advanced Computing (ICoAC), 2016.

[8] Mucherino, Petraq Papajorgji , P. M. Pardalos, "A survey of data mining techniques applied to agriculture", Springer, 2009.

[9] D Ramesh, B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data", IJARCCE, 2013

[10] Yunous Vagh,Jitian Xiao,"A data mining perspective of dual effect of Rainfall and Temperature on Wheat Yield", ECU, 2012.

[11] Patel, Hetal and D. Patel, "A brief survey of data mining techniques applied to agricultural data", IJCA, 2014.

[12] D Ramesh, B Vishnu Vardhan, "Crop Yield Prediction Using Weight Based Clustering Technique", IJCEA, 2015.

[13] Umair Ayub, Syed Atif Moqurrab, "Predicting Crop Diseases Using Data Mining Approaches: Classification", IEEE, 2018.

[14] Deepti Dighe, Harshada Joshi, " Survey of Crop Recommendation Systems", IJRET, 2018.

[15] Ms. D.S.Mokashi, Ms. P.M.Ghodke, "Crop Disease Prediction Using Data Mining Method", IJIRMPS, 2018