

SURVEY ON DATA MINING TECHNIQUES FOR CROPPREDICTION**L. Abirami* & Dr. R. Rathipriya****

* Research Scholar, Department of Computer Science, Periyar University, Salem, Tamilnadu

** Assistant Professor, Department of Computer Science, Periyar University, Salem, Tamilnadu



Cite This Article: L. Abirami & Dr. R. Rathipriya, "Survey on Data Mining Techniques for Crop Prediction", International Journal of Computational Research and Development, Special Issue, January, Page Number 43-46, 2017.

Abstract:

This paper mainly focuses on various crop yield prediction methods using data mining techniques. Agriculture system is very complex because it deals with large data from different factors such as geographical, biology, climate, and economy. Data mining is the emerging research field in agricultural crop yield analysis. Data mining is the extraction of hidden information from large datasets. The yield prediction is the main problem in agricultural field and that is solved by data mining techniques. The main aim of this work is to find the appropriate data models that attain high accurate and high generality of crop yield prediction.

Key Words: Data Mining, K-Means, K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), Yield Prediction & Regression Analysis.

1. Introduction:

Crop yield prediction is the most challenging problem in agriculture field. Every farmer is interested in knowing how much yield he is about to expect. In past, the yield prediction is estimated by considering farmers experience (Ramesh A. medar, 2014). Raw data, which is gained from history of crop yield, is required for the crop prediction (Perpetua Noronha, 2016). Data mining is the process of extracting data from raw and large dataset of information. Data mining functionalities are used to specify the kinds of patterns to be found in data mining tasks. In general such task can be classified into two categories "Descriptive" and "Predictive". Descriptive mining tasks characterize the properties of the data in a target dataset. Predictive mining tasks perform induction on the current data in order to make predictions. In data mining there are two categories:

- ✓ Clustering
- ✓ Classification

Clustering is the process of grouping a set of data objects into multiple groups or clusters so that objects within a cluster in have high similarity but very dissimilar to objects in other clusters. It is a dynamic field of research in data mining and it is related to unsupervised learning in machine learning. The clustering techniques are:

- ✓ K-Means Clustering.
- ✓ Density Based Clustering.

Classification is the process of finding a model that describes and distinguishes data classes or concepts. The models are derived based on the analysis of a set of training data. It is used to predict the class label of objects for which the class label is unknown. The classification techniques are:

- ✓ K-Nearest Neighbor (KNN),
- ✓ Artificial Neural Network (ANN),
- ✓ Support Vector Machine (SVM).

2. Data Mining Techniques:

2.1 K-Means Algorithm: K-means algorithm is the most widely used clustering technique. This algorithm can able to classify the data without the prior knowledge about the data or training set. The parameter 'k' indicated the number of clusters in which the data need to be partitioned. The k-means algorithm first chooses k, the number of clusters we want to find in the data. Then the centers of those k-clusters, called centroids, are initialized. Then recalculate each centroid location as the mean (center) of all the points assigned to its cluster. We then iterate these steps until the centroids stop moving or equivalently until the points stop switching clusters (Rajshekhhar Borate, 2016).

2.2 Density Based Method: In Density based method the clusters objects based on the concept of density. It grows clusters either according to the density of neighbor of objective e.g. DBSCAN or according to the density function e.g. DENCLUE. The primary idea of Density based clustering technique is that for each point of a cluster, the neighborhood of a given unit distance contains at least a minimum number of points. The density in the neighborhood should reach some threshold (D. Ramesh, 2015). It is used to filter out noise or outliers and it discovers the clusters of arbitrary shape. The Density based method consider exclusive clusters only and does not fuzzy clusters.

2.3 K-Nearest Neighbor (KNN): The KNN algorithm is a method for classifying objects based on the closest or most similar training sample in the feature space (Minakshi Sharma, 2013). The main idea of KNN is to base estimation on a fixed number of observations, say k, which are closest to the desired output (Miss Mayanka 2013). The KNN algorithm steps are shown in below figure (Ramesh A. medar, 2014).

2.4. Artificial Neural Network (ANN): Artificial neural network (ANN) structure which is parallel system and it is based on human brains biological neural process. It is used to solve complex problems. An ANN, which is also called a multilayer perceptron, work as information processed by the human brain. ANN model can produce more consistent and accurate crop yield prediction. There are different neural network designs where by information is processed in different manners. The most used

type of ANN today is nonlinear feed forward and by far the most popular feed forward type is back propagation (Siti khairunniza Bejo, 2014).

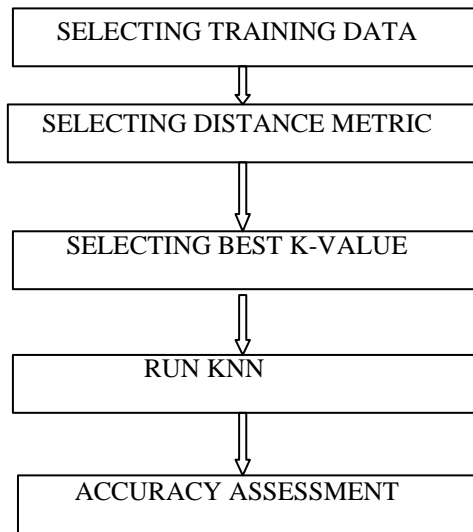


Figure 1: KNN Algorithm Steps

2.5 Support Vector Machine (SVM): Support vector machine is an algorithm for the classification of both linear and nonlinear data. A SVM is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize patterns used for classification and regression analysis (Perpetua Noronha, 2016). SVM is less vulnerable to over fitting problem and it uses a hypotheses space of linear functions in a higher dimensional feature space (Ami Mistry, 2016).

2.6 Regression: Regression uses existing values to forecast what other values will be. It uses standard statistical techniques such as linear regression and multiple linear regressions.

2.6.1 Linear Regression: A Linear regression model is computed that evaluates the relationships among the variables. The conditions associated with model include Linearity, Nearly normal residual and constant variability (V.Sellam, 2016).

2.6.2 Multiple Linear Regression: A regression model that involves more than one predictor variable is called multiple regression models. MLR is the method that is used to model the linear relationship between a dependent variable and one or more independent variables (D.Ramesh, 2013). MLR is a predictive analysis based on least squares and it is probably the most widely used method in climatology. The three major uses of MLR analysis is in casual regression, forecasting an effect, trend forecasting (Perpetua Noronha, 2016).

Table 1: Various Machine Learning Techniques

Reference	Statistical Model	Number of Variables
Warner and Misra(1996)	LR	12
Drummond et al.(2003)	SMLR,PPR	11
Paul et al.(2004)	LRM	11
Jiang et al.(2004)	MLR	5
Kaul et al.(2004)	MLR	20
Yong et al.(2005)	MLR	6
Ji et al.(2007)	MLR	60
Li et al.(2007)	MLR	15
Sing et al.(2008)	MLR	5
Mehnatkesh et al.	MLR	57
Ayoubi et al.(2011)	PCA	14
Zaefizadeh,M.et al.(2011)	MLR	5
Laxmi et al.(2011)	SMLP	5

Table1 list the various works in the literature for crop prediction using machine learning techniques like LR = Linear Regression, SMLR=Stepwise Multiple Linear Regression, PPR=Projection Pursuit Regression, LRM=Logistic Regression Model, MLR=Multiple Linear Regression, PCA=Principal Component Analysis.

3. Literature Review:

Ramesh A. Medar et, al. presented different data mining techniques for crop yield prediction. It reviews the research studies on application of data mining in the field of agriculture. Some of the technique are artificial neural network (ANN), k-means, KNN, SVM are presented. Raorane et, al. proposes that several changes in the weather can be analyzed by support vector machine and k-means method is used to forward the pollution in atmosphere. D. Ramesh et. Al, finds the region of specific crop yield analysis and it is processed by implementing both multiple linear regression techniques density based clustering technique. These models were experimented in respect of all districts of Andhra Pradesh. S. Veenadhari et al, observed the research studies on application of data mining techniques in the yield of agriculture. Some of the

International Journal of Computational Research and Development

Impact Factor 4.775, Special Issue, January - 2017

International Conference on Smart Approaches in Computer Science Research Arena

On 5th January 2017 Organized By

Department of Computer Science, Sri Sarada College for Women (Autonomous), Salem, Tamilnadu

techniques ID3 alg, k-means, k-nearest neighbor, ANN, SVM are presented this article explores the application of DM techniques in the field of agriculture. These techniques help in predicting and forecasting the crops effectively.

G. Nasrin fathima et al, The authors express that government should encourage the use of data mining technique to increase the yield and support the farmers to get reliable and timely information on crop area, crop production and land use. B. Vishnu Vardhan et al, explains different data mining techniques. This paper compares the yield prediction based on rainfall between MLR and k-means algorithm. Ami Mistry et al, briefly explained about cluster and classification technique that is applied to agriculture. V. Sellam et al, used regression analysis to determine and analyze the relationship between the response variable and explanatory variable. The variable considered for this analysis are Annual Rainfall (AR), Area under cultivation (AUC), Food price Index (FPI), Production, yield, area. The value $r^2 = 0.7$ is obtained by implementing the regression analysis. The r^2 value clearly states the Annual Rainfall (AR), AUC, FPI have an average of 70% influence in the crop yield. Milovic et al, explains the application of data mining in agriculture. By the use of data mining technique acquired knowledge can be used to make successful, decisions which will advance the success of agricultural organization on the market. Raju Prasad paswan et al, used neural network and Regression models to predict the crop production. It gives a summary of some of the study of comparison of ANNs and statistical models.

Table 2: Tabulates the various data mining methods used for prediction with different set of parameters

S.No	Author Name	Data Mining Methods	Prediction	Parameters	Accuracy
1	D.Ramesh, B.VishnuVardhan	Multiple Linear Regression, Density Based Clustering.	Crop Yield Prediction	Year, Rainfall, Area, Yield, Fertilizer, Production.	82%
2	V.Sellam, E.Poovammal	Linear Regression	Rice Prediction	Area Under Cultivation, Annual Rainfall, Food Price Index.	70%
3	Rajshekhar borate, Rahul Ombale	Regression, Kmeans	Yield Prediction	Rainfall, Temperature, Humidity, Area.	95%
4	Minakshi Sharma, Suresh Kumar Sharma	K-Nearest Neighbor	Humidity Prediction	Minimum temperature, Maximum Temperature, Low humidity.	82%
5	Tritha R.Ranjeet, Leisa J.Armstrong	Artificial Neural Network	Crop Yield Prediction	Climate, Soil, Fertilizer, Sowing Area.	95%
6	Sagar S.Badhiye, Nilesh U.Sambhe	K-Nearest Neighbor	Humidity, Temperature Prediction	Temperature, Dew point, Humidity, Sea level press, Visibility, Wind	85%
7	K.Kandiannan, R.Karthikeyan	Multiple Regression	Rice prediction	Area, Temperature, Sunshine, Rainfall, Solar Radiation	94%
8	D.Ramesh, B.VishnuVardhan	Kmeans, Multiple Linear Regression	Yield prediction	Year, Rainfall, Area, Production	98%
9	M.Sayedur Rahman, Md.Monimul huq	Multiple Regression, Ridge Regression	Rice Prediction	Maximum Temperature, Minimum Temperature, Average Temperature, Diurnal Temperature, Total Rainfall, Cumulative Rainfall, Net Solar Radiation, Water Balance index, Time, Dummy variable	98%

4. Applications in Agriculture:

This study shows that the use of data mining techniques and machine learning techniques as a great impact in agriculture. The data mining can be used primarily for:

- ✓ Crop/Yield Prediction.
- ✓ Irrigation Estimation.
- ✓ Water Consumption
- ✓ Manure Prediction
- ✓ Plant Nutrient.
- ✓ Nitrate Loading.

5. Conclusion:

The collection of data on agriculture field has increased every day with the advent of new computing technology. Because of the enormous of data, it is tough to analyze with help with traditional computing techniques. Hence, this study was designed to list the data mining and machine learning techniques available in the literature for smart agriculture practices. It has seen that these techniques have achieved a high accuracy and a high generality regarding yield prediction, irrigation usage and manure usage.

6. References:

1. Ramesh A. Medar and Vijay. S. Rajpurohit (2014), "A survey on Data Mining Techniques of crop prediction", International journal of advance research in computer science and management studies, 2321-7782.
2. Raorane A.A and Kulkarni (2012) R.V, "Data Mining: An efficient tool for yield estimation in the agricultural sector", International journal of emerging trends and technology in computer science, 2278-6856.
3. D. Ramesh, B. Vishnu vardhan (2015), "Analysis of crop yield prediction using Data mining techniques", International Journal of Research in Engineering and Technology, 2321-7308.
4. S.Veenadhari, Dr.Bharat Misra, Dr.CDsingh (2011), "Data Mining Techniques for Predicting Crop Productivity-A review article", IJCST, Vol.2, 2229-4333.
5. G.Nashrin Fathima and R.Geetha(2014), "Agriculture Crop Pattern using Data Mining Techniques", International journal of Advance Research in Computer Science and Software Engineering, Vol.4, 2277128X.
6. D.Ramesh,B.Vishnu vardhan (2013), "Data Mining Techniques and Application to Agricultural Yield data", IJARCCCE, Vol.2, 2319-5940.
7. Ami Mistry, Vinita shah (2016),"Brief Survey of Data Mining Technique Application of Agriculture", IJARCCCE, Vol.5, 2278-1021.
8. V.Sellam and E.Poovammal (2016), "Prediction of Crop Yield Using Regression Analysis", Indian Journal of Science and Technology, Vol.9, 0974-6846.
9. B.Milovic and V.Radojevic(2015), "Application of Data Mining in Agricultural Science", Bulgarian Journal of Agricultural Science, 26-34.
10. Raju Prasad paswan, Shahin Ara begum (2013), "Regression and Neural Networks Models for Prediction of Crop production", International Journal of Scientific and Engineering Research, Vol.4 , 2229-5518.
11. Perpetua Noronha, Divya.J,Shruthu.B.S (2016), "Comparative Study of Data Mining Techniquein Crop Yield Prediction", IJARCCCE, Vol5, 2278-1021.
12. Jiawei han,Micheline Kamber, Jian Pei, "Data mining Concepts and Techniques, Third edition, ISBN: 978-0-12-381479-1.
13. Rajshekhar borate, Rahul Ombale,Sagar Ahire (2016), " Data Mining Techniques to Predict Annual for major crops", IJSRD, Vol 4, 2321-0613.
14. Siti khairunniza Bejo, Samihah Mustaffha (2014), "Application of Artificial Neural Network on Predicting Crop Yield: A Review", Journal of food science and engineering, 1-9.
15. Minakshi Sharma and Suresh Kumar Sharma(2013), "Generalized K-nearest neighbor Algorithm-A predicting Tool", IJARCSSE.
16. Miss.Mayanka, B.Khuman (2013), "Classification of Remote Sensing data using KNN method", Journal of Information, Knowledge and research in electronics and communication engineering, Vol.2, 0975-6779.
17. M.Sayedur Rahman, Md. Monimul huq (2005), "Statistical Analysis of Crop Weather Regression Model for Forecasting production impact of Aus Rice in Bangladesh", International Journal of Statistical Sciences, Vol.4, pp 57-77.
18. Tritha R.Ranjeet, Leisa J.Armstrong (2004), "An Artificial Neural Network for Predicting Crop yields in Nepal", ECU publication post .
19. Sagar S.Badhiye, Nilesh U.Sambhe (2013), "KNN techniques for analysis and Prediction of temperature and humidity data", International Journal of Computer Applications, vol.61.
20. K.Kandiannan, R.Karthikeyan (2002), "A Crop Weather Model for Predicting Rice Yield using Empirical Statistical Technique", Journal of Agronomy and crop science, pp 59-62.