A review of different techniques utilized for-casting crop yield

Priya Bhardwaj*, Mrityunjay Singh

Dept. of Computer Science Engineering, SRM University, Ghaziabad, India- 201204
Corresponding author E-mail: priya18091993@gmail.com

Abstract

The farming structures the establishment of Indian economy. The harvest creation mainly depends on atmospheric conditions such as climate change, rain, soil etc., that impacts on yield improvement. The most of existing algorithms for crop yield prediction utilizes the existing data mining (DM) techniques for forecasting. This paper exhibits an overview on some of the existing techniques mostly used for crop yield prediction.

Keywords: Agriculture, Crop Analysis, Forecasting, Yield Prediction, Data Mining.

1. Introduction

Farming structures the establishment of Indian economy. Vast territories of rural land aren't accomplishing satisfactory yield generation thanks to each environmental condition and financial difficulties. The harvest creation on a really basic level depends upon atmospheric condition, that impacts on yield improvement. The lack of water, nonattendance of high yielding assortments and poor innovation exchange of best science practices are thought to be the important parts for low harvest yields in Asian country. On the off likelihood that harvest generation is to increment each yield creation and therefore the administration of trimming region should move forward. the employment of innovation and totally different computer code engineering procedures to anticipate crop productivity in numerous atmospheric condition will facilitate so muchmers' different partners in crucial basic leadership as far as scientific discipline and yield call. Varieties in atmosphere, soil, water, trouble attack and completelydifferentparts boost the compl--exness of the difficulty.

Data mining (DM) is outlined as a method of distinctive antecedently unknown inferences from the large volume of obtainable knowledge. It detects application in market research, fraud detection, customer retention, etc. DM software analyses relationships and patterns in hold on dealings knowledge supported open finished user queries.

On the premise of the character of information being deep-mined there square measure 2 classes of functions concerned in data processing particularly, Descriptive operate that deals with general properties of knowledge and Prediction operate that identifies the trends supported offered data. As way as agriculture worries prophetic varieties that embody classification,

association, cluster and regression square measure used as shown in fig1.

The main techniques for DM include Association rules mining, Classification, Clustering and Regression. The different DM techniques used for solving different agricultural problem has been discussed [1].

A. Association rules mining (ARM)

ARM method is a standout amongst the most effective strategies of DM to look concealed or desiredpattern among the immense measure of information. In this strategy, the attention is on discovering connections between the diverse things in a value-based database. Association rules are utilized to discover components that co-happen over and again inside a dataset comprising of numerous free choices of components, (for example, obtaining exchanges), and to find rules.The distinctive affiliation lead mining calculation are Apriori Algorithm(AA), Partition, Dynamic Hashing and Pruning(DHP), Dynamic Itemset Counting(DIC), FP Growth(FPG), SEAR, Spear, Eclat and Declat, MaxEclat [2]

B. Classification

There are two types of data analysis namely classification and prediction that can be utilized to extricate models portraying important data classes or to anticipate future information patterns. It is a procedure in which a model figures out how to anticipate a class mark from an arrangement of preparing information which would then be able to be utilized to foresee discrete class names on new specimens. To expand the prescient exactness acquired by the classification demonstrate while characterizing cases in the test set inconspicuous amid preparing is one of the significant objectives of characterization calculation. DM grouping calculations can take after three diverse learning approaches: semi-regulated learning, administered learning and unsupervised learning. The different classifition systems for finding information are Rule Based Classifiers (RBC), Bayesian Networks(BN), Decision Tree (DT), Nearest Neighbour(NN), Artificial Neural Network(ANN), Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, Genetic Algorithms [3]

C. Clustering

In clustering, the attention is on finding a partition of information records into groups to such an extent that the focuses inside each group are near each other. It clusters the information occasions into subsets in such a way, to the point that comparable examples areassembled together, while different occurrences have a place with assorted gatherings. Since the point of clustering is to discover another arrangement of classifications, the most recent gatherings are of enthusiasm for themselves, and their appraisal is inborn. [4] There is no earlier learning about information. The distinctive grouping techniques are Hierarchical Methods(HM), Partitioning Methods (PM), Density-based Methods(DBM), Model-based Clustering Methods(MBCM), Grid-based Methods and Soft-computingMethods [fuzzy, neural system based], Squared Error—Based Clustering (Vector Quantization), organize information and Clustering chart [5]

D. Regression

Regression is taking in a capacity that maps an information thing to a genuine esteemed expectation variable. The diverse applications of regression are foreseeing the measure of biomass display in a backwoods, assessing the likelihood of patient will survive or not on the arrangement of his symptomatic tests, anticipating customer interest for another product.[6] Here the model is prepared to foresee a persistent target. Regression undertakings are regularly regarded as arrangement assignments with quantitative class tag. The techniques for forecast are Nonlinear Regression(NLR) and Linear Regression (LR).

The contents of the paper are organized as follows: Section 2 discuss the existing techniques used for crop yield prediction, section 3 describe the measurement parameters, and finally, we conclude our work in section 4.
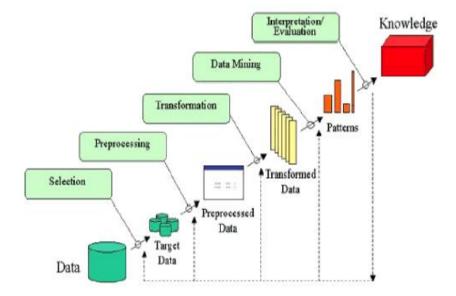

Fig. 1: Data MiningProcess [22]

Table 1: Comparison among the existing techniques

Author& Year	Data mining methodologies	Applications
2001,K. Verheyen [7]	Fuzzy set	Yield Prediction in agriculture
2006,Veenadhari, S. [8]	K-nearest Neighbor	Scale back procedure burdenof k-nearest neighbor algorithm
2012,Sanjay D. Sawaitul [9]	Neural Networks	Focuses on weather forecasts
2013, V.R Thakare [10]	fuzzy system	Soil Classification
2015,Aditya Shastry [11]	Fuzzy Logic, Adaptive Neuro Fuzzy Inference System (ANFIS)	crop yield in agriculture
2016,Hemageethaa [12]	Naïve Bayes, Apriori algorithm	Market Basket Analysis
2016,Niketa Gandhi [13]	BayesNet	rice crop yield for Maharashtra
2017,Umid Kumar Dey [14]	Nonlinear Regression, SVM	forecast the yield of rice with astute analysis

2. Existing Techniques

Ashwani Kumar et al. [15] proposed a new "Agro algorithm" to process huge volume of agricultural data using Big data Hadoop platform. With this method predict the crop yield for agriculture and advise the farmer with best one crop by which quality and profitability also improved. In this work also analysis the type of soil that best suits which particular crop is suitable for a soil. In farming, Climatic condition is much important, so predict the type of crop for soil and improve the crop quality by using weather and disease related data sets.

Ankalaki et al. [16] presents a comparative study on DBSCAN and AGNES algorithm for clustering. Crop yield is forecasted using MLR (Multiple Linear Regression) and a formula is derived for each crops.

Gayatri et al. [17] use IOT and web services to handle large amount of data. Sensors are used to collect the data and pass the data to data centre. Agriculture field images are captured and GPS is used to accurately feed the data into repositories along with their location. Far and near nodes are communicated through cloud .

Sellam et al [18] clarify the different ecological parameters like the Area under Cultivation (AUC), Annual Rainfall (AR) and Food Price Index (FPI) that impacts the yield of product and the relationship among these parameters is set up. Utilizing Regression Analysis (RA), Linear Regression (LR) the different natural variables and their curse on crop yield is examined .

Naushina et al. [19] proposed to implement K-NN technique to create clusters and making predictions on the basis of large

amount of data. Also use of Geospatial analysis for crop yield prediction, GSA is used to identify a pattern that helps farmers with information about fertilizers.

S. Poongodi et al. [20] proposed to help the farmer for choosing the appropriate crop for their cultivation based on this data. In feature selection, genetic algorithm is implemented for selecting the optimal features. Then improved C4.5 with ANFIS classifier is used for classifying the data based on region wise. The experiment results of proposed work shows 92.50 % better accuracy than existing classifier.

D. Ramesh et al. [21] used Data Mining techniques: classification and clustering, and Neural Networks to picture the applications of Data Mining in Agricultural Yield. While, selecting Rainfall as the dependent attribute and Year, Area of sowing and Production as the independent attributes, in proposed use of K-Means Algorithm and Multiple Linear Regression to make their prediction. With an accuracy of 98%, they concluded that MLR technique was more precise than K-Means Algorithm, which had a success rate of 96%.

3. Performance measure

Root Mean Square Error (RMSE) of prediction: It defined the error measure between observed and predicted values.

$$RMSE = \sqrt{m^{-1} \sum_{j=1}^{m} (x_j - y_j)^2}$$

where m is number of years j, y_j and x_j are observed and predicted values.

Peak Signal Noise Ratio (PSNR) of prediction: PSNR defined the improved prediction value and is the most wanted feature. The higher the value, the processing efficiency is better of the approach. It can be calculated in decibels (dB) and it is given by

$$PSNR(x) = \frac{10 \, X \log\left((255.^2)\right) / \log 10}{RMSE(x)}$$

4. Conclusion

This paper overviews different techniques specialized accomplishments in the field of harvest yield forecast. Plates philosophy, far reaching overview of different proposed strategies to anticipate crop yield and applications. It likewise talks about different information digging procedures utilized for forecast of harvest yield. Developing better systems to predict trim profitability in different climatic conditions can help agriculturist and diverse accomplices in fundamental essential administration to the extent agronomy and item choice.Further this analysis results can be directly available to farmers through web which can help to increase production and can get better price.

References

[1] Mucherino, A., Papajorgji, P., & Pardalos, P. (2009), "Data mining in agriculture" (Vol. 34), Springer.

[2] Zaki, M J (1999), "Parallel and distributed association mining: A survey". IEEE concurrency, 7(4), 14-25.

[3] Beniwal, S. & Arora, J. (2012), "Classification and feature selection techniques in data mining", International Journal of Engineering Research& Technology (IJERT), 1(6), 1-6.

[4] Lior Rokach, Oded Maimon, "Clustering Methods", Chap-15

[5] Xu, R & Wunsch, D (2005), "Survey of clustering algorithms", Neural Networks, IEEE Transactions on, 16(3), 645-678.

[6] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996), "From data mining to knowledge discovery in databases". AI magazine, 17(3), 37.

[7] K. Verheyen, D. Adriaens, M. Hermy, and S. Deckers, "High resolution continuous soil classification using morphological soil profiledescriptions", Geoderma, vol. 101, pp. 31-48, 2001.

[8] Veenadhari, S. 2007, "Crop productivity mapping based on decision tree and Bayesian classification". Unpublished M.Tech Thesis submittedto Makhanlal Chaturvedi National University of Journalism and Communication, Bhopal.

[9] Sanjay D. Sawaitul, Prof. K.P. Wagh, Dr. P.N. Chatur, "Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, Issue 1, January 2012, pp. 110-113.

[10] V.R.Thakare and H.M. Baradkar. "Fuzzy System forMaximum Yield from Crops", International Journal ofApplied Information Systems, pp. 4-9 , March 2013.

[11] Aditya Shastry , Sanjay H A and Madhura Hegde," A Parameter based ANFIS Model for crop yield prediction", 2015 IEEE,pp 253-257.

[12] Hemageetha, N., "A survey on application of data mining techniques to analyze the soil for agricultural purpose", 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp.3112-3117, 2016.

[13] Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong," PredictingRice Crop Yield Using Bayesian Networks", ICACCI,IEEE 2016,pp 795-799.

[14] Umid Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin, "Rice Yield Prediction Model Using Data Mining", ECCE, IEEE 2017, pp 321-326.

[15] Ashwani Kumar Kushwaha, SwetaBhattachrya " Crop yield prediction using Agro Algorithm in Hadoop" IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 5, No2, April 2015.

[16] Ankalaki, S., Chandra, N., Majumdar, J., "Applying Data Mining Approach and Regression Model to Forecast Annual Yield of Major Crops in Different District of Karnataka", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Special Issue 2, pp.25-29, 2016.

[17] Gayatri, M.K., Jayasakthi, J., Anandha Mala, G.S., "Providing Smart Agricultural Solutions to Farmers for better yielding using IoT", IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR), pp.40-43, 2015.

[18] Sellam,V., Poovammal, E., "Prediction of Crop Yield using Regression Analysis", Indian Journal of Science and Technology, Vol. 9(38), pp.1-5, 2016.

[19] Naushina Farheen Imam Shaikh, Prof. R. V. Argiddi "Annual Crop Yield Prediction and Recommend Planting of Different Crops by Using Data Mining Technique" International Journal of Innovative Research in Computer and Communication Engineering Vol. 4, Issue 10, October 2016

[20] S. Poongodi and M. Rajesh Babu " Prediction of Crop Production using Improved C4.5 with ANFIS Classifier" International Journal of Control Theory and Applications, Vol.10(21), pp-121-132,2017.

[21] D. Ramesh and B. V. Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.

[22] Monali Paul, Santosh K. Vishwakarma, Ashok Verma "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" International Conference on Computational Intelligence and Communication Networks , 2015.

[23] Harikishore Kakarla, Madhavi Latha M and Habibulla Khan, "Transition Optimization in Fault Free Memory Application Using Bus-Align Mode", European Journal of Scientific Research, Vol.112, No.2, pp.237-245, ISSN: 1450-216x135/1450-202x, October 2013.

[24] T. Padmapriya and V. Saminadan, "Improving Throughput for Downlink Multi user MIMO-LTE Advanced Networks using SINR approximation and Hierarchical CSI feedback", International Journal of Mobile Design Network and Innovation-Inderscience Publisher, ISSN : 1744-2850 vol. 6, no.1, pp. 14-23, May 2015.

[25] S.V.Manikanthan and K.srividhya "An Android based secure access control using ARM and cloud computing", Published in: Electronics and Communication Systems (ICECS), 2015 2nd International Conference on 26-27Feb. 2015, Publisher: IEEE, DOI:10.1109/ECS.2015.7124833.