

A Review on Yield Prediction of Various Techniques and Features

Neha Sharma

Dept. of Information Technology
UIT, BU
Bhopal, M.P., India

Abstract- Agriculture is one of the major revenue producing sectors of India and a source of survival. Various seasonal, economic and biological factors influence the crop production but unpredictable changes in these factors lead to a great loss to farmers. This paper presents a survey on the various models used for crop yield prediction. So this paper focuses on this problem of increasing the size of the data. Here various approaches adopted by researchers are detailed with their field of accuracy for prediction. Some of the issues related to the papers are also discussed. Techniques of knowledge extraction and storage were discussed in this work. Here details of features required to analyze the crop yield are present with their calculation and requirements.

Index Terms- Crop yield prediction, Data mining, Machine Learning, Vegetation Index.

I. INTRODUCTION

Agriculture is the backbone of Indian Economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depends on weather conditions. Rainfall conditions also influence the rice cultivation. In this context, the farmers necessarily require a timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production in their crops. Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about to expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. The volume of data is enormous in Indian agriculture. The data when become information is highly useful for many purposes.

Data Mining is widely applied to agricultural problems. Data Mining is used to analyze large data sets and establish useful classifications and patterns in the data sets. The overall goal of the Data Mining process is to extract the information from a data set and transform it into understandable structure for further use. The intention of this paper is to give details about different data mining techniques in perspective of agriculture domain so researchers can get details about appropriate data mining technique in context to their work area.

Data mining tasks can be classified into two categories: Descriptive data mining and Predictive data mining. Descriptive data mining tasks characterize the general properties of the data in the database while predictive data mining is used to predict explicit values based on patterns determined from known results. Prediction involves using

some features or fields in the database to predict unknown or future values of other variables of interest. As far as data mining technique is concerned in the most of cases predictive data mining approach is used. Predictive data mining technique is used to predict future crop, weather forecasting, pesticides and fertilizers to be used, revenue to be generated and so on.

Whole paper is organized into few sections where second section gives explanation of various prediction techniques used by researcher for yield prediction. While third section gives summary of the work done by different authors in this field of crop yield prediction, here comparison table of researcher approach was shown with their limitations. Fourth section explained various features on which production of any crop is dependent such as NDVI, VCI, etc. Finally problem still present in the work which needs to be covered is summarized with respective solutions.

II. TECHNIQUES OF YIELD PREDICTION

In the existing system the yield prediction is done with the single algorithm based classification in which the single algorithm is used with the single dataset by that we can get only the single output with the single dataset. By these process we cannot get the 100% result hence in proposed system we are using the hybrid model for the classification and by the hybrid model we are increasing the accuracy level of the result and 100% guaranteed the accurate output. And the second problem in the existing system is that it acquires a more time for processing because these process is going through the single

algorithm based classification i.e the single input and the single output. Data Mining Methods

1. Association Rule Mining - Association rule mining technique is one of the most efficient techniques of data mining to search unseen or desired pattern among the vast amount of data. In this method, the focus is on finding relationships between the different items in a transactional database. Association rules are used to find out elements that co-occur repeatedly within a dataset consisting of many independent selections of elements (such as purchasing transactions), and to discover rules. An application of the association rules mining is the market basket analysis, customer segmentation, store layout, catalog design, and telecommunication alarm prediction. An application of the association rules mining is the market basket analysis, customer segmentation, store layout, catalog design, and telecommunication alarm prediction. The different association rule mining algorithm are Apriori Algorithm(AA), Partition, Dynamic Hashing and Pruning(DHP), Dynamic Itemset Counting(DIC), FP Growth(FPG), SEAR, Spear, Eclat&Declat, MaxEclat.

2. Classification- Classification is the data mining techniques used to predict group membership for data instances. It has a training set containing data that have been previously categorize. Some classification algorithm requires training data. Classification is supervised learning technique use to assign predefined tag to instances on the basis of feature. The different classification techniques for discovering knowledge are Rule Based Classifiers, Bayesian Networks (BN), Decision Tree (DT), Nearest Neighbours(NN), Artificial Neural Network(ANN), Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, Genetic Algorithms.

3. Clustering- The process of grouping a set of physical or abstract object into classes of similar objects is called clustering. Clustering is unsupervised technique used to group similar instances on the basis of feature. It has no labels required. Clustering does not require training data. Each cluster that is form can be viewed as a class of object, form which rule can be derived The different clustering methods are Hierarchical Methods(HM), Partitioning Methods (PM), Density-based Methods(DBM), Model-based Cluster Methods(MBCM), Grid-based Methods and Soft-computing Methods [fuzzy, neural network based], Squared Error-Based Clustering (Vector Quantization), network data and Clustering graph.

III. RELATED WORK

In [1] Pritam Bose et. al. illustrates this concept with the introduction of the first SNN computational model for crop yield estimation from normalized difference vegetation index image time series. It presents the development and testing of a methodological framework

which utilizes the spatial accumulation of time series of Moderate Resolution Imaging Spectra radio meter 250-m resolution data and historical crop yield data to train an SNN to make timely prediction of crop yield. The research work also includes an analysis on the optimum number of features needed to optimize the results from our experimental data set. The proposed approach was applied to estimate the winter wheat yield in Shandong province, one of the main winter-wheat-growing regions of China. Our method was able to predict the yield around six weeks before harvest with a very high accuracy.

In [2] Michele Meroni et. al. The present study complements this quality assessment by analyzing the effect of using PV NDVI instead of VGT NDVI for operational crop monitoring and yield forecasting activities. The study makes use of paired observations from the two instruments collected during the overlap period when both satellite systems were active. This overlap period, ranging from end of October 2013 to end of May 2014, covers the time period used by North African countries to monitor crop development (January to May) and to provide timely yield forecast of the major grain crops (barley, soft wheat, and durum wheat).

In [3] N. Gandhi et al (2016), presented the overview on utilization of machine learning system for Indian rice editing ranges. Machine learning systems can be used to enhance forecast of harvest yield under various climatic situations. This paper examines at the exploratory outcomes acquired by applying SMO classifier utilizing the WEKA apparatus on the dataset of 27 areas of Maharashtra state, India. Those dataset acknowledged to the rice trim yield forecast might have been sourced from openly available Indian organization records. The parameters recognized for those review were precipitation, base temperature, Normal temperature, most extraordinary temperature Furthermore reference trim evaporation, territory, arrangement Furthermore yield for those Kharif season (June to November) for the years 1998 to 2002.

In [4] M.C.S.Geetha (2015), examined about the piece of information mining in context of cultivating field and furthermore presents around a few information mining systems and their related work by a few creators in setting to horticulture area. It additionally examines on various information mining applications in tackling the distinctive horticultural issues. Horticulture is the most noteworthy application region especially in the creating nations like India. Utilization of data innovation organization over agribusiness might change those circumstance from claiming choice making Furthermore farmers could yield clinched alongside An exceptional approach. Information mining assumes an significant part to choice making looking into a few issues identified with farming field.

This paper gives a overview of different information mining systems utilized within agribusiness. Research paper expects in discovering suitability information models that attain An secondary precision What's more a secondary sweeping statement for admiration to four parameters in particular rainfall, year, handling also region from claiming sowing. To this purpose, distinctive sorts from claiming information mining systems were assessed around separate information sets.

In [5] D. Ramesh and B. Vardhan (2015), exhibited An short Investigation about crop yield prediction utilizing Density based clustering technique and Multiple Linear Regression (MLR) for the selected region. A recent development in Information Technology for horticulture field has turned into an intriguing exploration region to anticipate the harvest yield. The issue from claiming yield forecast may be a significant issue that remains will be tackled In view of accessible information. Diverse information mining strategies are utilized Also assessed in agribusiness assessing what's to come year's harvest creation. At first the factual model Multiple Linear Regression strategy is connected on existing information. The impacts so gotten were checked Also investigated utilizing the information mining framework to be particular Density-based grouping strategy. In this system the aftereffects of two strategies were looked at as expressed by particular locale

In [10] S. Dahikar and S. Rode (2014), proposed simulated neural system approach for horticultural product yield expectation. By considering different circumstances of climatologically marvels influencing neighborhood climate conditions in different parts of the world. These climate conditions directly affect edit yield. Different examines have been done investigating the associations between extensive scale climatologically marvels and product yield. shown to be intense devices for displaying and expectation, to expand their adequacy.

Edit forecast procedure is utilized to foresee the appropriate harvest by detecting different parameter of soil and furthermore parameter identified with climate parameters. For that reason we are utilized Artificial Neural Network (ANN). Authors inferred that ANN is valuable apparatus for product expectation. In this paper incorporates the parameter of their provincial soil parameter. At that point it is examine by utilizing encourage forward back engendering ANN. Break down in tangle lab ANN way to deal with make it more effective.

Table 1 Comparison of Different Techniques outcomes and Limitation.

Author	Technique	Outcome	Limitations
Pritam Bose et. al. 2016	Spiking Neural Networks for Crop Yield	Accuracy of 95.64%, with an average error of prediction	Whole prediction done by using NDVI
Hemageet haa, 2016	Naïve Bayes, Appriori algorithm are used	Focuses mainly on various soil parameters like pH,	Only 77% of accuracy is achieved.
Bendre, 2015	Map Reduce and Linear Regression algorithm	The effective model to improve the accuracy	The forecasting is done based on only a
Fathima, 2014	k means and Appriori algorithm are used.	Crop type and Irrigation parameters are considered	Focus on the policies that government
Raorane, 2012	Artificial Neural Network (ANN), Decision	The techniques used for crop production is	Respective methodologies does not give efficient

IV. MODELS OF CROP YIELD

1. Attribute selection- The dependable attributes can be difficult to find. Several methods of predicting and modeling crop yields have been used in the past with varying success. Farmer has to face the different problems due to various factors which affect the planning made by him in advance. These factors do not have the fixed type of impact, it varies time to time, year to year depends on the situation, climatic nature, increase in costs of various constraints under uncertain environment, ambiguity and vagueness. Fuzzy logic [4] modeling provides the formulation of mathematical modeling to find the interface results in uncertain situations. Statistical models often don't take into account characteristics of the plants, the weather, or the soil attributes limiting their usefulness. Some models are based on information from just a single year or location. When a model is developed using single location or year data, it will have limited practical applications, therefore variability from multiple environments must be included.

2. Crop yield prediction- The crop yield prediction comprises of mostly all essential parameters that are needed for the better yield of crop. This enhances the classification results of the crop yield. All the essential parameters are considered as inputs to the model-ANFIS. In general, one of the difficulties faced in the prediction process is that most of the essential parameters that are

necessary to consider for the accurate prediction are not consider. It reduces the efficiency of the predicted results which in turn leads to lack of proper forecasting of the crop yield. It is also more complex to predict the optimized number of input parameters that are to be considered in the prediction process.

3. ANFIS models- ANFIS model is one of the efficient ways which is used for prediction, by imposing most of the essential parameters as inputs, it improves the accuracy of prediction results which has the property of learning by artificial neural network [5]. ANFIS presents some linearity with respect to some of its parameters, hence it increases the overhead of computation process without increasing the efficiency. The ANFIS failed to optimize the fuzzy rules in ANFIS which degrades the performance of prediction. Inputs that are to be consider and selected depends on the heuristics. It also decreases the performance by degrading the efficiency of the prediction process.

4. Fuzzy inference system- The implementation of Fuzzy Inference System (FIS) [6] as a technique for predicting the onset of rainy season based on the Southern Oscillation Index (SOI) data. Fuzzy set memberships and the set of rules are designed by investigating the two sets of data (via visualization and clustering). The prediction system is verified by using the actual data from the district. This explains an implementation of a Fuzzy Inference System for the prediction of rainy season onset. A fuzzy system was developed based on the knowledge gained from data exploration in a certain district to cover a larger range of area. Also, the computation of the output from the input data is based on a logical knowledge backed by observation data. Accordingly, it is expected to have a better accuracy result than the previous works.

5. Neuro fuzzy technique- The model explains the multi objective linear programming problem [7] by optimization technique and ANFIS model on neuron fuzzy technique for prediction of crop yield. In this the author explains, Fuzzy logic is utilized for the purpose of effective feature extraction and classified the crop yield data with the help of extracted features using ANFIS model. The ANFIS classified the data based on the fuzzy sets created in every input variable. It analysis various optimization methods and it is proved that OR optimization solution gave better results than the other optimization solution in fixed situation. The author explains deals with MOLPP by optimization technique and ANFIS model on Neuro-Fuzzy technique. Fuzzy Logic is used for effective feature extraction and Adaptive Neuro-Fuzzy Inference System (ANFIS) is considered for the classifier model.

V.CONCLUSION

Agriculture is the most important application area particularly in the developing countries like India. Use of information technology in agriculture can change the scenario of decision making and farmers can yield in better way. For decision making on several issues related to agriculture field; data mining plays a vital role. This paper displays the survey of Data Mining Techniques for crop yield prediction and concentrates the advantage of utilizing it. The paper gives an overview of available literature works of a few algorithm utilized by various specialists to use different information mining systems, for crop yield Prediction. The work that has been finished by different analysts in this field has been looked in a forbidden frame. This paper integrates the work of various authors in one place so it is useful for researchers to get information of current scenario of data mining techniques and applications in context to agriculture field.

REFERENCES

1. Pritam Bose, Nikola K. Kasabov, Fellow, IEEE, Lorenzo Bruzzone, Fellow, IEEE, and Reggio N. Hartono. "Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series". IEEE Transactions On Geoscience And Remote Sensing 2016.
2. Michele Meroni, Dominique Fusbender, Riad Balaghi, Mustapha Dali, Myriam Haffani, Ismael Haythem, Josh Hooker, Mouanis Lahlou, Raul Lopez-Lozano, Hamid Mahyou, Moncef Ben Moussa, Nabil Sghaier, Talhaoui Wafa, and Olivier Leo. "Evaluating NDVI Data Continuity Between SPOT-VEGETATION and PROBA-V Missions for Operational Yield Forecasting in North African Countries". IEEE Transactions On Geoscience And Remote Sensing, Vol. 54, No. 2, February 2016 795.
3. N.Gandhi, L.J. Armstrong, O. Petkar and A. Tripathy, "Rice Crop Yield Prediction in India using Support Vector Machines", IEEE The 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Thailand, 2016
4. M.C.S.Geeta, "A survey on data mining techniques in Agriculture", International Journal of Innovative Research in Computer and Communication Engineering, vol. 3, No. 2, pp. 887-892, 2015
5. D. Ramesh and B. Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of Research in Engineering and Technology, vol. 4, no.1, pp. 47-473, 2015
6. R. Medar, V. Rajpurohit, "A survey on data mining techniques for crop yield prediction", International Journal of Advance Research in Computer Science and Management Studies, vol. 2, no. 9, pp. 59-64, 2014.
7. R.Kalpana, N.Shanti and S.Arumugam, "A survey on data mining techniques in Agriculture", International

- Journal of advances in Computer Science and Technology, vol. 3, No. 8, 426-431, 2014
8. G.N.Fatima, R.Geeta, "Agriculture crop pattern using data mining techniques", International Journal of Advanced Research in in Computer Science and Software Engineering, vol. 4, No. 5, 781-786, 2014
 9. Mr. Abhishek B. Mankar, Mr. Mayur S. Burange, "Data Mining - An Evolutionary View of Agriculture", International Journal of Application or Innovation in Engineering & Management (IJAIEEM), vol. 3, No. 3, pp. 102-105, 2014
 10. S. Dahikar and S. Rode, "Agricultural crop yield prediction using artificial neural network approach", International Journal of Innovative Research in Electrical, Electronic Instrumentation and Control Engineering, vol. 2, no. 1, pp. 683-686, 2014
 11. A.A.Raorane and R.V.Kulkarni, "Review-Role of data mining in Agriculture", International Journal of Computer Science and Information Technology, vol. 4, No. 2, pp. 270-272, 2013
 12. Yetheraj .N.J, "Applying data mining techniques in the field of agriculture and allied sciences", International Journal of Business Intelligent, vol. 1, No. 2, pp. 72-76, 2012.
 13. Raorane, A.A., Kulkarni R.V., "Data Mining: An effective tool for yield estimation in the agricultural sector", International Journal of Emerging Trends & Technology in Computer Science(IJETTCS), Vol. 1, issue 2, pp.75-79, 2012.
 14. E. Manjula, S. Djodiltachoumy "A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017 ISSN: 2349-6363 298