A Study on Various Data Mining Techniques for Crop Yield Prediction

Yogesh Gandge
Department of Master of Computer Applications
B.K.I.T. Bhalki,
Bidar, India
yogeshvin22@gmail.com

Sandhya
Department of Studies in Computer Applications
Visvesvaraya Technological University,
Post graduate Centre, Mysuru, India.
sanjoshi17@yahoo.com

***Abstract*— India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. It is also one of the country which suffer from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide. Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc. Predicting a crop well in advance requires a systematic study of huge data coming from various variables like soil quality ,pH ,EC,N,P,K etc. As Prediction of crop deals with large set of database thus making this prediction system a perfect candidate for application of data mining. Through data mining we extract the knowledge from the huge size of data. This paper presents the study about the various data mining techniques used for predicting the crop yield. The success of any crop yield prediction system heavily relies on how accurately the features have been extracted and how appropriately classifiers have been employed. This paper summarizes the results obtained by various algorithms which are being used by various authors for crop yield prediction, with their accuracy and recommendation.**

***Keywords*— crop yield prediction, data mining, data mining algorithm, models, accuracy and Recommendation.**

I. INTRODUCTION

India is an agriculture based country where most of the people derive their living from this sector. Agriculture is having a great impact on the country's economy. In the last decade India has seen serious natural calamities like drought or flood. Due to such disasters there is a huge loss to crop production and ultimately to the farmers. Due to such financial loss many farmers are committing suicide. If natural calamities are not present then there may be sudden pest attack destroying the crop. In any case farmer and the crop are always at the edge of risk. Government policies are there but that is not sufficient. Figure 1. shows the major crop producing states of India.

Prediction of crop yield in advance can help the farmers and the Government bodies to plan for storage, selling, fixing minimum support price, importing /exporting etc.

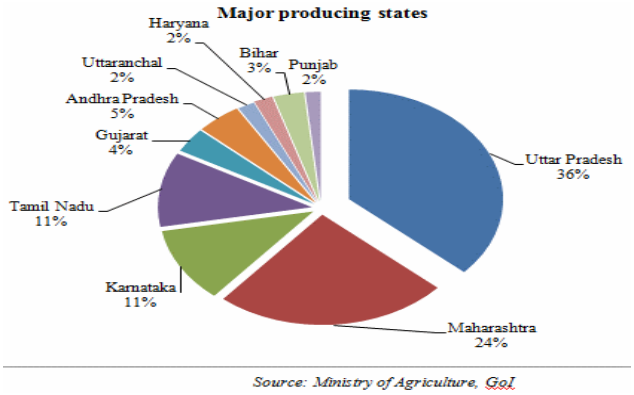

Figure 1: Major crop producing states

Till now only the past experience of the farmer was used which consist of randomly counting the number of seed buds that a plant is having and within each seed bud the number of seed it holds. Then based on experience the farmer used to predict the crop yield. Figure 2 shows an example of the (split pigeon peas) toor crop with seed buds.

Figure 2 Toor crop seed buds

Information technology can be used to avert the risk associated with the agriculture and it can also be used to predict the crop yield more accurately prior to harvest .Yield prediction needs different kinds of data gathered from different sources like meteorological data, agri- meteorological, soil (pH,N,P,K) data, remotely sensed data,agricultural statistics etc[1].To handle such a huge data the best option we have is Data Mining. Data mining

978-1-5386-2361-9/17/$31.00 ©2017 IEEE

is a method by which one can extract the knowledge from the huge bulk of data.

Data mining techniques are mainly divided in two groups:

- Classification
- Clustering.

Classification techniques are intended to categorize unknown samples using information provided by a set of classified samples. This set is usually referred to as the training set it is used to train the classification technique to perform its classification.

In case a training set is not there, then there is no knowledge about the data to categorize. In such cases clustering technique can be used to split a set of unknown samples into clusters[2].

Data mining procedure is separated into seven methods [13]:

- Data cleaning
- Data integration
- Data selection
- Data transformation
- Data mining
- Pattern estimation
- Knowledge display

II. METHODOLOGY

Through crop yield prediction system better planning and decisions can be chalked out for enhancing the yield

A. Input

Most of the research papers that were studied have considered some climatic parameters like temperature, humidity, rainfall. Some agronomical parameters like soil, nutrient contents like N, P, K, and pesticides etc.The values of these variables have been taken as input.

B. Preprocessing (Noise Removal)

For the successful application of data mining a huge set of dataset is required. The data which is acquired from various resources are sometime in raw form. It may contain some incomplete, redundant, inconsistent data. Therefore in this step such redundant data should be filtered. Data should be normalized.

C. Feature Extraction (Attribute Selection)

This step aims at identifying and using most relevant attribute from the dataset. Through this process irrelevant and redundant information is removed for the application of classifiers

D. Output

The output is the crop yield prediction per acre with some recommendation. A broad outline of the crop prediction approach is shown in figure 3.

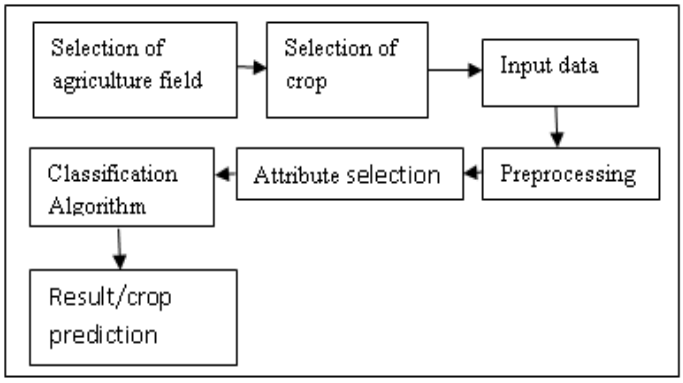

Figure 3 crop prediction system

A Brief overview of the Crop Prediction System:

1. Selection of agriculture field: Consider any agriculture field for the crop prediction system.
2. Selection of crop: consider any crop of choice which will be sown in that field.
3. Input data: Data may include information regarding soil (Nitrogen (N),Phosphorus(P),Potassium(K) content), Micronutrients present in soil, Moisture in soil etc which is collected over some period of time.
4. Preprocessing: Data which is collected should be preprocessed redundant data, inconsistent should be taken care.
5. Attribute Selection: Important Features have to be extracted.
6. Classification Algorithm: An appropriate and efficient algorithm should be employed.
7. Result: prediction or recommendation can be provided to the farmers based on the results obtained.

TABLE I. Different approaches used for crop yield prediction

Algorithm /models used	Crop type	Accuracy/ Recommendation
Multiple Linear Regression [3]	Rice yield	90%-95%
Decision tree analysis and ID3[4]	Soybean	The rules formed from the decision tree are helpful in predicting the conditions responsible for the high or low soybean crop productivity under given climatic

		parameters
Support Vector Regression model [5]	For any crop	The results show that support vector regression can serve as a better reference model for yield prediction. It is computationally less demanding.
Three models used APAR,SEBAL, Carnegie Institution Stanford model[6]	Wheat,rice ,sugarcane, cotton	Successful for wheat ,rice ,sugarcane but not successful for cotton. The proposed technology can significantly contribute to quantitatively describing yield variations across the Indus Basin
Neural Networks [7]	Corn yield	95%
C4.5 algorithm and decision tree [8]	Soyabean, paddy, maize	For soyabean=87% For paddy=85% For maize=76%
Harmonic Analysis of NDVI Time-Series algorithm [9]	sugarcane	86.5%
Gaussian Processes [10]	Wheat yield	Wheat yield is expected to increase with an increase in temperature but there can be an increasing under estimation error in predicting the wheat yields
Relational cluster Bee Hive algorithm [11]	Any crop	This crop yield prediction model (CRY) performs 12% better than cluster & Regression Tree algorithm
K-Means algorithm for clustering And for classification Linear Regression,k-NN, ANN model [12]	Wheat , potato	90% - 95%
J48, LADTree [14]	Rice	100%
K-Nearest Neighbor (KNN) and Naive Bayes (NB) [15]	Any crop	classification of soil into low, medium and high categories are done

III. FUTURE WORK

It can be observed from table 1, there are still many challenges in this research area. To meet such challenges a more careful study has to be carried out. We should explore for a robust and novel classifier to improve the performance of the prediction system. We should be able to suggest based on the nutrient contents of the soil which crop is most suitable for a farmers land. In addition we should be suggesting water tolerant seed variety for sowing so that in case of flood and drought the crop could withstand the damage. Sensors can be used to measure the moisture and the nutrients in the soil, this information can also be used to guide the farmers

IV. CONCLUSION

According to our study, there is still scope for the improvement in result. During the study which we have carried out it is observed that the algorithm which is used by most of the authors does not uses a unified approach where in all the factors affecting the crop yield can be utilized simultaneously for predicting the crop yield. There is still further scope of improvement as the dataset which is considered is small in some cases. Therefore the result can be improved by using a large dataset.

REFERENCES

[1] Maria Rossana C.de Leon,Eugene Rex L.Jalao, "A prediction model framework for crop yield prediction," Asia Pacific Industrial Engineering and Management System.

[2] Ramesh A. Medar,Vijay S. Rajpurohit, "A survey on Data Mining Techniques for crop yield prediction", International Journal of Advance Research in Computer Science and Management Studies.Volume 2 , Issue 9, Sept 2014.

[3] D Ramesh, B Vishnu Vardhan, "Region specific crop yield Analysis :A data Mining approach ",UACEE International Journal of Advancees in computer Science and its Applications-IJCSIA volume 3:issue 2.

[4] S.Veenadhari, Dr Bharat Mishra, Dr C D Singh, "Soyabean productivity modeling using Decision Tree Algorithms," International Journal of Computer Applications Vol 27-No7 Aug 2011.

[5] George RuB, "Data Mining of Agricultural Yield Data:A comparison of Regression models.

[6] Wim G.M. Bastiaanssen, Samia Ali, "A new crop yield forecasting model based on satellite measurements applied across the Indus Basin,Pakistan," Agriculture, Ecosystems and Environment 94 (2003) 321–340.

[7] Sudhanshu Sekhar Panda, Daniel P. Ames , and Suranjan Panigrahi, "Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques" , Remote Sensing 2010, 2, 673-696; doi:10.3390/rs2030673

[8] S.Veenadhari,Dr. Bharat Misra ,Dr. CD Singh," Machine Learning Approach for forecasting crop yield based on climatic parameters", 978-1-4799-2352-6/14/$31.00 ©2014 IEEE

[9] Jeferson Lobato Fernandes; Jansle Vieira Rocha," Sugarcane yield estimates using time series analysis of spot vegetation images", Rubens Augusto Camargo Lamparelli, Sci. Agric. (Piracicaba, Braz.), v.68, n.2, p.139-146, March/April 2011

[10] Yunous Vagh, Jitian Xiao," Mining temperature profile data for shire-level crop yield prediction", 978-1-4673-1487-9/12/$31.00 ©2012 IEEE, Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012

[11] M. Gunasundari, Dr. T Arunkumar, Ms. R. Hemavathy," CRY – An improved Crop Yield Prediction model using Bee Hive Clustering

Approach for Agricultural data sets", Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22,978-1-4673-5845-3/13/$31.00©2013 IEEE

[12] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman," Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh", 978-1-4799-8676-7/15/$31.00 copyright 2015 IEEE SNPD 2015, June 1-3 2015, Takamatsu, Japan

[13] R. Sujatha, Dr. P.Isakki,A Study on Crop Yield Forecasting Using Classification Techniques, 978-1-4673- 8437-7/16/$31.00 ©2016 IEEE.

[14] Niketa Gandhi, Leisa Armstrong," Applying Data Mining Techniques to predict yield of Rice in Humid Subtropical Climatic Zone of India". 978-9-3805-44212/16/$31.00_c 2016 IEEE.

[15] Monali Paul, Santosh K. Vishwakarma, Ashok Verma," Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" ,2015 International Conference on Computational Intelligence and Communication Networks, 978-1-5090-0076-0/15 $31.00 © 2015 IEEE, DOI 10.1109/CICN.2015.156