

Review on Use of “BAGGING” Technique in Agriculture Crop Yield Prediction

Dipika H. Zala¹ M.B. Chaudhari²

¹PG Student ²Professor

^{1,2}Department of Computer Engineering

^{1,2}Government Engineering College Gandhinagar, Gandhinagar, Gujarat, India

Abstract— In the countries like India, Agriculture is a most important contributor for economy. In now days, there are irregular environmental parameter affects badly to the yield of crop in agriculture due to harmful pollution. So, it cannot be predicted easily that which crop should be taken on which environmental parameter. Problem can be resolved by taking past year data analysis and making decision according to that. Research done here is for yield prediction of groundnut, is by studying past data set and analyzing that data. And so that DATA MINING concept comes to resolve this problem. After studying various technique and reading previous research on agriculture yield prediction using different data mining technique, our research uses data mining BAGGING TECHNIQUE for yield prediction of groundnut. Dataset used in research area has parameters like Minimum Temperature, Average Temperature, Maximum Temperature, Area, production, evapotranspiration. The main goal of this research is for getting better efficiency of crop prediction using “BAGGING” technique.

Key words: Data Mining, Crop Yield Prediction, Bootstrap Aggregating Technique

I. INTRODUCTION

Crop yield refers to the production of crop in Kilogram per hectare. It also referred as crop productivity. Crop Yield prediction is the forecasting of crop yield before harvesting of crop by analyzing past yield, weather data.

Crop yield depends on many different factors. Now a day's crop yield prediction becomes important as environment changed drastically and there is need to crop yield predict accurately. Agriculture is one of the major revenues 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. These risks can be quantified when appropriate mathematical or statistical methodologies are applied on data related to soil, weather and past yield. With the advent of data mining, crop yield can be predicted by deriving useful insights from these agricultural data that aids farmers to decide on the crop they would like to plant for the forthcoming year leading to maximum profit. Reliable forecasts of crop production before the harvest constitute a problem of topical interest. Such forecasts are needed by the Government, policy makers, agro-based industries, traders and agriculturists alike.

The government needs these for use as a basis for its policy decisions in regard of procurement, distribution, buffer-stocking, import export, price fixation and marketing of agricultural commodities while agro-based industries, traders and the agriculturists need forecasts for planning their operations properly. In view of this, there is a need for an objective methodology for pre harvest crop yield prediction.

This involves building up suitable forecast model(s) which has certain merits over the traditional forecasting method. These merits include the objectivity of the forecast and its ability to provide a measure of reliability which a traditional forecast method cannot provide. This, as such, calls for the necessity of objective methods for pre-harvest forecast of crop yields.

Different ways used for crop yield prediction: Various organizations in India and abroad are engaged in developing methodology for pre-harvest forecast of crop yields using various approaches. The main factors affecting crop yield are inputs and weather. Use of these factors forms one approach for forecasting crop production. The other approach uses plant vigor measured through plant characters. It can be assured that plant characters are integrated effects of all the factors affecting production. Yet another approach is measurement of crop vigor through remotely sensed data. Since area under the crop is available in advance of harvest through timely reporting scheme (TRS) and remotely sensed data, the problem mainly reduces to forecast of yield rate. Statistical forecast models applied in agricultural systems for forecasting crop yield are as follows

- 1) Regression Models
 - Multiple linear regression (MLR) models using plant characters
 - Weather indices based MLR models
 - Logistic regression models
- 2) Time Series Models
 - Exponential smoothing models
 - Auto-regressive integrated moving average (ARIMA) models
- 3) Probabilistic Models
 - Markov chain models
 - Naive Bayes Probabilistic model
 - Bootstrap Aggregating (Bagging)

Agriculture plays vital role in the Indian economy. Over 70 per cent of the rural households depend on agriculture. Agriculture contributes about 17% to the total GDP and provides employment to over 60% of the population. Agriculture yield depends on different factors of environment, farming strategy and planning. Agricultural crop yield can be predicted using historical crop production and climate data. Crop yield prediction is important to increase crop yield, farmer can decide farming strategy based on predicted yield. India is the second largest country in production of groundnut.

II. LITERATURE SURVEY

A. Agriculture Crop Yield Prediction

Agriculture is backbone of counties like India. Flow and methodology for agriculture yield prediction is given as:[8]

- Selection of Crop Region
- Apply Crop Yield Model
- Input Datasets
- Apply Model
- Preprocessing
- Evaluate Results

Crop yield can be predicted using historical data of crop production and weather data. A crop production dataset may contain parameters like year, area, production, and yield. A weather dataset may contain parameters like minimum temperature, maximum temperature, average temperature, precipitation, evapotranspiration, and reference crop evapotranspiration. These are the basic parameters of weather dataset for crop yield prediction; weather dataset may also contain more parameters.

B. K-Means & Clustering

The K-Means algorithm can partition the samples in to various clusters [9] The k-means algorithm is used for partitioning, in which each cluster's center is represented by the mean value of the objects in the cluster

- Input: k: no. of clusters,
- D: a dataset containing n objects.
- Output: A set of k clusters.

C. Bagging Technique Used On Crop Yield Prediction

Bootstrap aggregating proposed by Leo Breiman in 1994 for improve the classification by combining classifications of randomly generated training sets. Bootstrap aggregating (bagging) is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid over fitting. Although it is usually applied to decision tree methods, it can be used with any type of method. Bagging is a special case of the model averaging approach. Given a standard training set D of size n, bagging generates m new training sets D_i , each of size $n' < n$, by sampling from D uniformly and with replacement. By sampling with replacement, some observations may be repeated in each D_i . If $n'=n$, then for large n the set D_i is expected to have the fraction $(1 - 1/e)$ of the unique examples of D, the rest being duplicates. This kind of sample is known as a bootstrap sample. The m models are fitted using the above m bootstrap samples and combined by averaging the output voting. The Bagging Algorithm Training phase [5]

- 1) Initialize the parameter $D=\emptyset$, the ensemble. Lk the number of classifiers to train.
- 2) For $k=1 \dots L$, take a bootstrap sample S_k from Z. Build a classifier D_k using S_k as the training set. Add the classifier to the current ensemble, $D=D \cup D_k$.
- 3) Return D. Classification phase
- 4) Run D_1, \dots, D_k on the input x.
- 5) The class with the maximum number of votes is chosen as the label for x.

Snap of clustering:

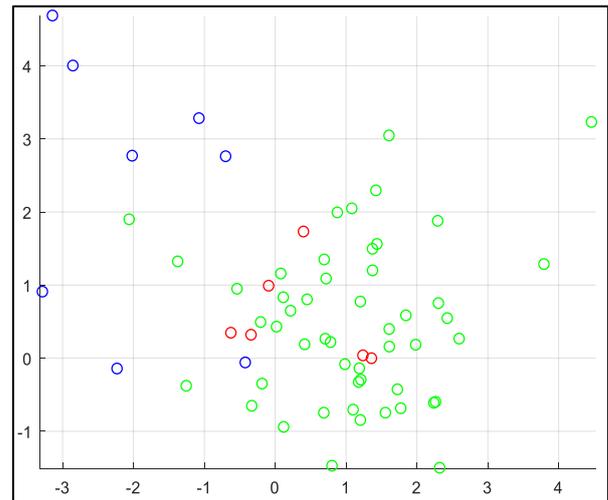


Fig. 1: Cluster 1

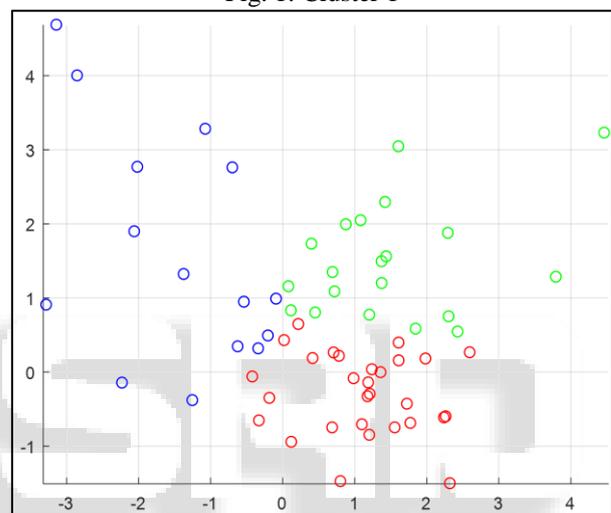


Fig. 2: Cluster 2

Bootstrap aggregating is a technique used for enhancing results of prediction algorithms. This method was formulated by Leo Breiman and its name was deduced from the phrase "bootstrap aggregating" [1]. More detail of Bootstrap aggregating can be understood by [2], [3] and [4].

Classification algorithms that are used for yield prediction in past research work are Support Vector Machine, Random Forest, Neural Network, REP Tree, Bagging, and Bayes. Crop yield prediction can be done using different classification algorithm. Best result of all the available algorithm is using Bagging technique is explained by detailed analysis and plotted graph in past research work [6]

By using parameters for crop yield prediction like soil type, soil pH, pest details, weather, water level, and seed type gives the measurement of crop yield which is used for a food grains, legume and is usually calculated in metric tons per hectare. Crop harvest is able to refer the real seed invention from the plant. [7]

Crop yield prediction is done by many data mining techniques, SVM method, naïve Bayes method, REP tree, neural network, BAGGING technique. In prediction of crop yield, BAGGING technique gives better accuracy then others [6].

D. Bagging Technique Result

1) Future work

In future work, the prediction of Groundnut crop yield in the relationship of area and production can be done by composing another technique to bagging technique and enhance the result.

III. CONCLUSION

The application of crop yield prediction is an important concern in data mining. It can be easily decidable that which climate or weather is suitable for groundnut crop. The accurate prediction of crop yield can help governments and authorities to have strategic decision in policy making and also helpful for farmers.

In this paper, we propose an efficient Bootstrap aggregating technique for the prediction of crop yield. Voting cast of bagging technique with different parameters will improve the accuracy in the prediction of groundnut crop yield.

REFERENCES

- [1] Breiman, L.: Bagging predictors. Technical Report 421, Department of Statistics, University of California at Berkeley, 1994.
- [2] Friedman, J., Springer, L.: <http://www-stat.stanford.edu/people/faculty/friedman.html>
- [3] Hastie, T.: <http://www-stat.stanford.edu/%7Ehastie/> Robert Tibshirani, R.: <http://www-stat.stanford.edu/%7Etibshirani/> Jero Abhijit A. Sawant¹, P. M. Chawan² "Study of Data Mining Techniques used for Financial Data Analysis", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, May 2013
- [4] Anshal Savla¹, Parul Dhawan², Himtanaya
- [5] Bhadada³, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture", International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS), 2015.
- [6] R.Sujatha¹, Dr.P.Isakki², "A Study on Crop Yield Forecasting Using Classification Techniques", IEEE International conference on IEEE, 2016.
- [7] Aakunuri Manjula, "XCYPF: A Flexible and Extensible Framework for agricultural Crop Yield Prediction", IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO), 2015
- [8] D Ramesh¹, B Vishnu Vardhan², "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering Vol.2, September 2013, pp.3477-3480
- [9] Niketa Gandhi, "Applying Data Mining Techniques to Predict Yield of Rice in Humid subtropical Climate Zone of India", IEEE International conference on computing for sustainable global development, 2016.
- [10] A.T.M Shakil Ahamed¹, Navid Tanzeem Mahmood², Nazmul Hossain³, Mohammad Tanzir Kabir⁴, Kallal Das⁵, Faridur Rahman⁶, Rashedur M Rahman⁷, "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh", IEEE, 2015
- [11] Indian Brand Equity Foundation (2015) Indian Agriculture Industry: An Overview. www.ibef.org/industry/agriculture-india.aspx
- [12] Importance of Agriculture In The Indian Economy www.publishyourarticles.net/knowledge-hub/essay/essay-on-the-importance-of-agriculture-in-the-indian-economy/3300/
- [13] Madhusudan L, "Agriculture role on Indian economy", Business and Economics Journal, pp. 20th Aug 2015
- [14] "GROUNDNUT CROP AND ITS IMPORTANCE AMONG THE OIL SEEDS" http://shodhganga.inflibnet.ac.in/bitstream/10603/33332/10/10_chapter%204.pdf
- [15] Jiawei Han¹, Micheline Kamber², "Data Mining: Concepts and Techniques", 3rd Edn., Elsevier and Morgan Kauffman Publishers, 2012.
- [16] Microsoft, "Data Mining Concepts", <https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/data-mining-concepts>