

## Crop Yield Prediction using Random Forest

R. Vidhya, Assistant Professor, Dept. Computer Science Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai

Pragya Mathur, Shivani Sai Valluri

Dept. Computer Science Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai

### Abstract

*Agriculture is the most vital sector of Indian Economy. This industry is one of the biggest sources of income for the people. Ours is a country which suffers from natural calamities which leads to the crop getting destroyed. We are prone to getting affected by floods, droughts and other such calamities and this in turn leads to great financial losses for the farmers. Predicting the crop yield well in advance can assist the farmers in making suitable plans for things like storing, selling etc. Recent development in the domain of data mining has enabled us to reach a point where we can research about prediction of crop yield. Prediction of crop yield deals with large set of data which makes this prediction system a suitable one for application of data mining. This paper offers a quick evaluation of crop yield prediction using Random Forest technique for a specific region.*

**Keywords** – Agriculture, Data Mining, Random Forest, Prediction.

### I. INTRODUCTION

Data Mining is nothing but a method that offers with analysis, extraction and prediction of some useful significant statistics from massive sets of information. It is a procedure which essentially gives meaning to any uncooked facts. The method of Data Mining includes various steps. The first step desires to be accumulating data that desires to be processed. Once we've got all of the facts with us the following step is pre-processing of that records that's the step in which we can take away anomalies. Pre-processing facts approach locating any lacking values, any null values or locating rows with zeros.

The very last step is reworking the facts to get styles that may then be used to are expecting useful insights basically giving which means to the uncooked information. Data mining techniques can be applied at the climate and crop production related records amassed over years and many predictions can be made on the idea of this. These predictions will help us in increasing the production of crop in a extra green way. In a rustic like ours in which agriculture is a prime source of residing, introducing era in agriculture. It will help in benefiting a large sector of our society.

We feel that the scope of this project is not just restricted to crop yield prediction but we can extend the technology being used here in things like Weather forecasting, water management etc which will help the farmers of our country. At this point we will be concentrating on just one area or state but with time we would want to cover all the states where agriculture plays an important role and include as many crops as possible too.

The Government has started a scheme where they are collecting soil samples and testing it for farmers, it's called soil health cards. Researchers can make use of the data gathered from this and give it more meaning. Weather forecasting is something that is already being done and we can extend the purpose of our project to have a greater meaning by extending the technology to things like smart irrigation systems. Crop yield prediction is just one part of a way bigger picture and with the help of researchers we can help the economy of our country by helping in the field of agriculture.

### II. LITERATURE SURVEY

Agriculture is a major source of income of several Indians. There are many reasons to make crop cultivation strong. There are many ways to make use of the data mining techniques in the vast field of agriculture. Researchers in [5] divided land into clusters. In this they have assumed that the lands which contains similar values of the significant attributes should ideally be part of the same cluster. This crop yield prediction model [6] makes use of Multiple Linear Regression where Production is what we are

predicting and to do so we are making use of a number of factors like Rainfall, Year, Area of Sowing, Yield and few Fertilizers.

Various techniques like Multiple Linear Regression, Random Forest and other algorithms on the dataset of different crops like rice, maize etc were done [2] and comparison was done between actual production and predicted production. For analysis WEKA tool was used [4]. WEKA is an open source software which is used in solving problems related to data mining and machine learning and for visualisation. It performs various tasks some of which are pre-processing, classification, regression, clustering and visualization. The data set is given as an input into the software and then we select the task we want to implement. It provides a variety of classifiers that can be used for building models and solving analytical problems. It has an interactive interface which contains all the alternatives that are available for data analysis. Feature extraction or Attribute selection [1] aims at recognizing and using the most significant attribute that we can get from the dataset. Through this process we primarily get rid of any irrelevant and redundant information that is there.

### III. PROPOSED WORK

We need to look for a robust and different classifier which will help us in improving the performance of the prediction system. We can make use of sensors to degree the moisture and the vitamins inside the soil, this information also can be used to assist out the farmers which is the number one aim behind this prediction device. The proposed set of rules for crop yield prediction facilitates in creating a decision assist gadget for green farming. Therefore, this paper indicates an concept to predict the yield for particular states the use of some of various factors which effect the manufacturing of crop. Farmers can employ this facts concerning the predicted yield of the crop for that particular kingdom and based totally on which could determine to domesticate it.

Decision tree classifiers appoint grasping technique. This method that when we make a selection of an characteristic within the very first step it may not be used anywhere anymore. This will restrict us in masses of strategies due to the fact can be the use of that characteristic in a later step also can provide us better effects. Also, over-fitting may additionally stand up which can supply bad outcomes. As we will see there are certain obstacles worried and using an ensemble model is a better approach in order to conquer those barriers.

### IV. IMPLEMENTATION

An ensemble version sincerely a mixture of numerous other models. The result we get by way of the use of an ensemble version is advanced compared to that obtained from any other model.

Random Forest comes under the category of ensemble classifiers. It uses a number of decision trees to predict the result. There are several subsets of training data and each time we select a different subset to train each tree. A forest as the name suggests is a collection of trees and these trees are then trained with the help of the different subsets of training data. The subsets are selected at random and that's why it's far known as random forest. We can use this for both classification in addition to for regression issues. The output is mode of the instructions of the individual bushes. The number of bushes in the wooded area is what defines the robustness of the prediction. So, extra the trees, greater robust the prediction. Random Forest can perform each category and regression.

- One major advantage of this technique is that it is parallelizable. This means that we the process can be split and can be run of a number of machines simultaneously. This results in faster computation time. It generally operates by constructing several decision trees at training.
- Over fitting is less of an issue in Random Forest. Hence, pruning need not be done.
- The variety of trees is what makes RF more effective than individual decision trees
- It manages large datasets with high accuracy because we are working with subsets of data.
- 

### V. RESULT

The use of a data set with more consequences in a better fee of accuracy. Random Forest indicates to be a advanced prediction algorithm when as compared to other methods like decision trees, a couple

of linear regression and so forth. By using a couple of capabilities to predict the production fee for a selected country in a specific 12 months for a certain crop.

## VI. CONCLUSION

The use of a dataset with more features results in a higher rate of accuracy. Random forest shows to be a superior prediction algorithm when compared to other methods like decision trees, multiple linear regression etc. Our dataset includes a lot more factors which has led to better prediction

## VII. REFERENCES

1. Mrs.K.R. Sri Preethaa M. E, S. Nishanthini, D. Santhiya, K. Vani Shree,” CROP YIELD PREDICTION”, Arasur, Coimbatore, International Journal On Engineering Technology and Sciences – IJETS, March- 2016
2. E. Manjula, S. Djodiltachoumy,” A Model for Prediction of Crop Yield”, International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017
3. Suvidha Jambekar, Shikha Neman and Zia Saquib, “Prediction of Crop Production in India Using Data Mining Techniques”, Pune, India, IEEE, 16-18 Aug. 2018.
4. R. Sujatha and P. Isakki, “A study on crop yield forecasting using classification techniques”, Kovilpatti, India, IEEE, 7-9 Jan. 2016.
5. Shruti Mishra, Priyanka Paygude, Snehal Chaudhary and Sonali Idate, “Use of data mining in crop yield prediction“, Coimbatore, India, IEEE, 19-20 Jan. 2018.
7. A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, and Rashedur M Rahman, “Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh” , Takamatsu, Japan, IEEE, 1-3 June 2015.
8. D Ramesh, B Vishnu Vardhan, “Analysis of Crop Yield Prediction using Data Mining Techniques”, International Journal of Research in Engineering and Technology (IJRET), Vol.4, 2015.
9. M. Gunasundari Ananthara, Dr. T. Arunkumar and Ms. R. Hemavathy, “CRY — An improved crop yield prediction model using bee hive clustering approach for agricultural data sets”, Salem, India, IEEE, 21-22 Feb. 2013
10. Yvette Everingham, Justin Sexton, Danielle Skocaj and Geoff Inman-Bambe , “Accurate prediction of sugarcane yield using a random forest algorithm”, Accepted:22 March 2016/Published online: 19 April 2016, INRA and Springer-Verlag France 2016.
11. E. Manjula, S. Djodiltachoumy,” A Model for Prediction of Crop Yield”, International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017
12. 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. 5964, 2014.
13. M.C.S.Geetha, “A Survey on Data Mining Techniques in Agriculture”, International Journal of Innovative Research in Computer and communication Engineering, Vol. 3, No. 2, pp. 887892, February 2015.
14. Yethiraj N G , ” Applying data mining techniques in the field of Agriculture and allied sciences”, Vol 01, Issue 02, December 2012
15. Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, “Analysis of agriculture data using datamining techniques: application of big data”Majumdar et al. J Big Data (2017) 4:20 DOI 10.1186/s40537-017-0077-4
16. V. Anandhi and R.M. Chezian, “Support vector Regression in forecasting”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, no. 10, October 2013.