Advanced Science Letters
E-ISSN:1936-7317

Crop Yield Prediction by Using Machine Learning Techniques

Dr.M.Purushotham[1] ,K.Hima Varshini [2], R.Dhatri [2], Afeera Samreen[2] , M.Pallavi Reddy[2]
[1]Professor, Department of CSE,[2]Final year B.Tech. Students, Department of CSE,
Sreyas Institute of Engineering and Technology, Hyderabad, Telangana, India

Abstract

The nature and magnitude of these impacts depends both on the evolution of the climate system, Accurate yield estimation is essential in agriculture. This Project develops machine learning based techniques for accurate crop yield prediction. The paper concludes that the rapid advances in sensing technologies and ML techniques will provide cost-effective and comprehensive solutions for better crop and environment state estimation and decision making. We are using two different algorithms, Artificial Neural Network and Random Forest to find out the best applicable one. Also providing an option to predict the future production based on user input.

Keywords: Crop Yield Prediction, K-NN, Random Forest.

1. Introduction

This paper focuses on the latter—yield prediction from weather. Accurate models mapping weather to crop yields are important not only for projecting impacts to agriculture, but also for projecting the impact of climate change on linked economic and environmental outcomes, and in turn for mitigation and adaptation policy. These approaches have relied on classical econometric methods. Recent work has sought to fuse crop models with statistical models, variously by including crop model output within statistical models (Roberts et al 2017), and by using insights from crop models in the parameterization of statistical models (Roberts et al 2012, Urban et al 2015).

In parallel, machine learning (ML) techniques have advanced considerably over the past several decades. ML is philosophically distinct from much of classical statistics, largely because its goals are different—it is largely focused on prediction of outcomes, as opposed to inference into the nature of the mechanistic processes generating those outcomes. (We focus on supervised ML—used for prediction—rather than unsupervised ML, which is used to discover structure in unlabelled data.)

2. Problem Statement

Crop yield prediction is an important agricultural problem. Farmers need information regarding crop yield before sowing seeds in their fields. The agricultural yield primarily depends on

weather conditions, pests and planning of harvest operation. Accurate information about history of crop yield is an important thing for making decisions related to agricultural risk management. So we have used random forest and k-nearest neighbour algorithms on agriculture data in order to suggest top 3 best crops along with the prediction(in hectares)of given crop.

3.Objective

The main objective is to compare the output of KNN and Random forest to verify whether the results in crop prediction are accurate. This paper uses crop yield prediction techniques to forecast the appropriate crop by identifying land parameters in acres, crop, state, district, year and seasonal parameters.

4.Existing System

Remote sensing (RS) systems are being more widely used in building decision support tools for contemporary farming systems to improve yield production and nitrogen management while reducing operating costs and environmental impact.RS based approaches require processing of enormous amounts of remotelysensed data from different platforms and, therefore greater attention is currently being devoted to machine learning methods. This is due to the capability of machine learning based systems to process a large number of inputs and handle non-linear tasks.

5.Proposed System

5.1.Easily Identifies Trends and Patterns

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviours and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

5.2.No Human Intervention Needed (Automation)

With ML, you don’t need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on

their own. A common example of this is anti-virus softwares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

5.3.Continuous Improvement

As **ML algorithms** gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

5.4.Handling multi-dimensional and multi-variety data

Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

5.5.Wide Applications

You could be an e-trailer or a healthcare provider and make ML work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

6. **Architecture**

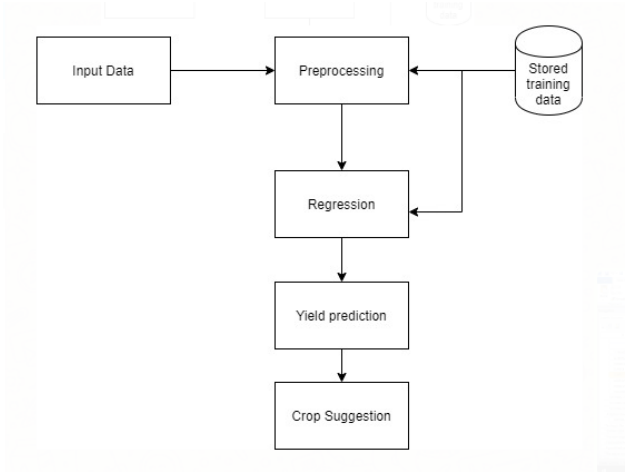


Figure 1:Architecture of Crop Yield Prediction

7.Algorithms

7.1.K-Nearest Neighbour

K-NN is a non-parametric and lazy learning algorithm. Non-parametric means there is no assumption for underlying data distribution i.e. the model structure determined from the dataset. For implementing any algorithm, we need dataset. we must load the training as well as test data then choose the value of K i.e. the nearest data points. Then for each point calculate the distance between test data and each row. The most commonly used method to calculate distance is Euclidean. Now based on the distance value, sort them in ascending order. Next, it will choose the top K rows from the sorted array.

Now, it will assign a class to the test point based on most frequent class of these rows.

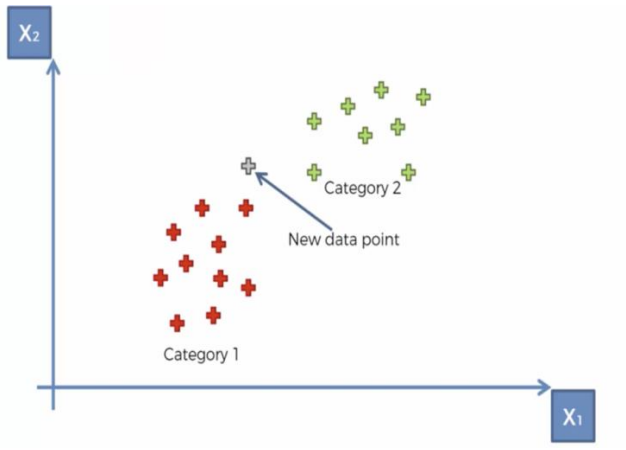


Figure 2: K-Nearest Neighbour

7.2.Random Forest

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.Random.forest is an unsupervised learning algorithm which uses both classification and regression.

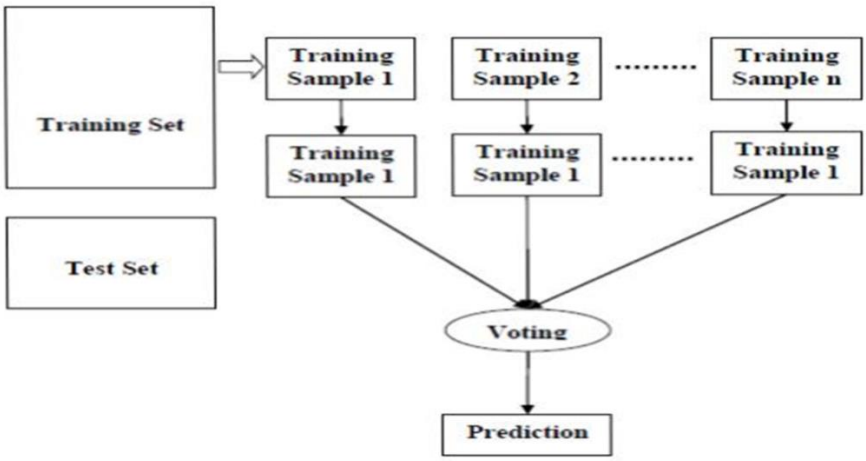


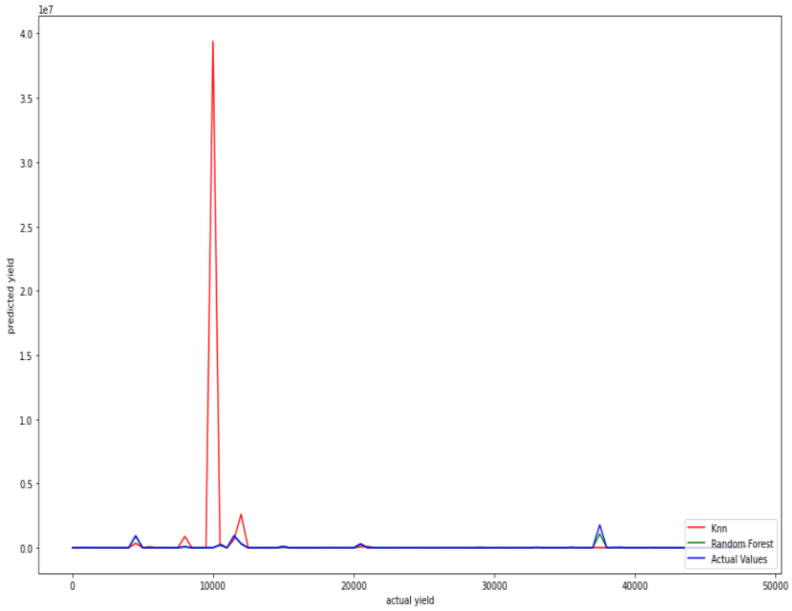
Figure 3: Random Forest

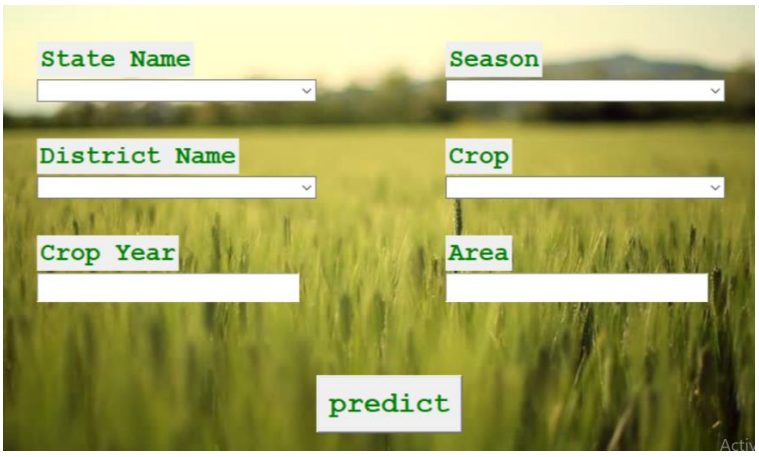
8.Test Cases And Scenario

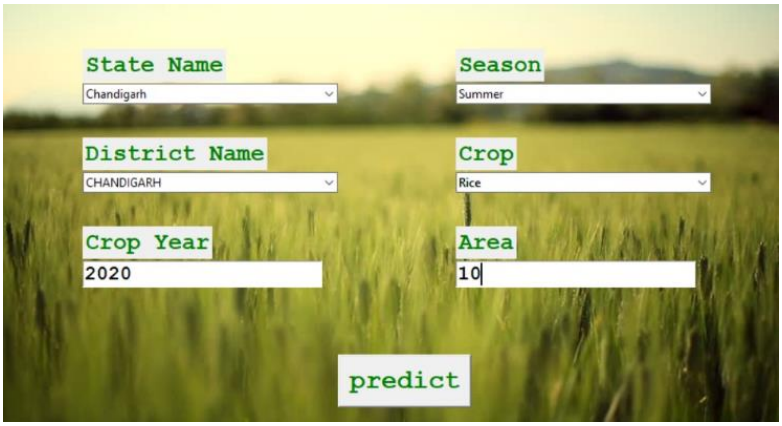
TestCase Id	Test Case Description	Expected Output	Actual Output	Result
TestCase 1	Outputis perfectly displayed if all the fields are selected	production is displayed	Production is displayed	Pass
TestCase 2	It displays the following if any one of the field is left empty	"please enter the values and do not leave anything empty"	"please enter the values and do not leave anything empty"	Fail
TestCase Id	**Test Case Description**	**Expected Output**	**Actual Output**	**Result**
TestCase 3	Until and unless state field is selected we cannot choose district	It asks to Choose state name first	It asks to Choose state name first	Fail

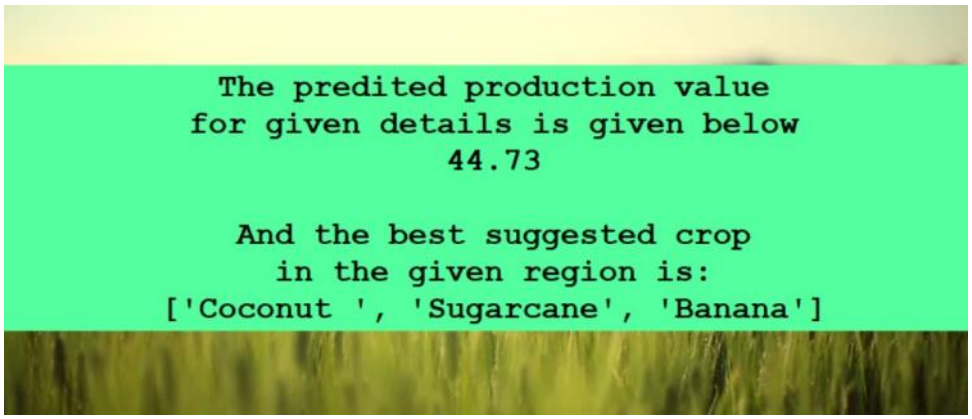
TestCase Id	Test Case Description	Expected Output	Actual Output	Result
TestCase 4	User can directly use keyboard in to provide input rather than dropbox	production is displayed	production is displayed	Pass

9.Results









10.Conclusion

Accurate yield estimation is essential in agriculture. This Project develops machine learning based techniques for accurate crop yield prediction. We are using two different algorithms here they are K-Nearest Neighbour and Random Forest to find out the best applicable one and also suggest the best crop that is suitable. We are doing this project to improve the crop production of farmers with the help of the data available from the past experiences of farming particular crops during particular seasons.

11.References

1. Aditya Shastry, H.A Sanjayand E.Bhanushree, “Prediction of crop yield using Regression Technique”, International Journal of computing12 (2):96-102 2017,ISSN:1816-9503

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. Mrs.K.R.Sri Preethi, S.Nishanthini, D.SanthiyaK.Vani Shree, "CropYield Prediction", InternationalJournal on Engineering Technology and Sciences – IJETS™ISSN (P): 2349-3968, ISSN (O):2349-3976 Volume III, Issue III, March- 2016

4. Askar Choudhury, James Jones, "CROP YIELD PREDICTION USING TIME SERIESMODELS"

5. 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

6. 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.

7. Yethiraj N G," Applying data mining techniques in the field of Agriculture and allied sciences", Vol 01, Issue 02, December 2012.

8. Zelu Zia (2009). An Expert System Based on Spatial Data Mining used Decision Tree for Agriculture Land Grading. Second International Conference on Intelligent Computation Technology and Automation.Oct10-11, China.

9. Dhanya, C.T. and D. Nagesh Kumar, 2009. Data mining for evolution of association rules for droughts and floods in India using climate inputs. J. of Geo. Phy.Res.114:1-15.

10. Fernandes F.L., Jansle V.R., Rubens Augusto and Camargo L. (2011). Sugarcane yield estimates using time series analysis of spot vegetation images. Sci. Agric. (Piracicaba, Braz.)vol.68no.2.