YIELD PREDICTION USING MACHINE LEARNING: ANN AND RF COMPARISON

Annie Christila S
Asst.Professor, St. Francis de Sales College, Bangalore
anniechristilas@sfscollege.in

Lakshmi CB
Asst.Professor, St. Francis de Sales College, Bangalore
lakshmicb@sfscollege.in

Abstract

Crop yield estimation is an active field of study and experimentation in agriculture to promote an efficient decision support tool including what crops to grow during appropriate climatic conditions. This paper discusses about the comparison of artificial neural network and Random Forest algorithm to predict the Crop yield. The Study states that evaluating the performance of actual target values to the predicted values by the machine learning models infer that both ANN & RF have better performance in crop yield estimation considering error rate ANN shows better prediction results.However, accurate yield estimation for agricultural planning is a critical problem. This paper has identified ANN algorithm has more accuracy that tends to boost crop yield estimation and to investigate in the agricultural environment that machine learning algorithms need to focus on and resolve. From different outputs, it shows that Random Forest is an efficient learning algorithm to analyse crop at current climatic condition and has a huge exactness in data investigation. The chosen variety crops are cereals depending on factors such as soil moisture, Moisture rainfall, Average Humidity, Temperature, Soil texture like alkaline, chalky, and sandy. Study also focusses on the important parameters required for the study thereby soil moisture content is taken as the key parameter.

Keywords: *Crop yield, Machine learning, Artificial Neural Network, Random Forest, Predictive Analytics.*

INTRODUCTION

Agriculture determines the survival and development of the Earth's inhabitants. Crop yield dependent on various agronomic factors such as soil moisture, Moisture rainfall, Average Humidity, Temperature, Soil texture like alkaline, chalky, and sandy. Crop yield forecasting is critical artificial neural networks and random forests are being used. Most attempts in crop yield forecast have been heavily based on the empirical relationships of environmental parameters such as rainfall, temperatures, environmental indexes etc. with respect to the crop yield (Isard et al., 1995;Jain and Ranjana, 2000;Kandiannan et al., 2002). In an machine learning a model can be either descriptive or predictive, descriptive models help us in obtaining knowledge from the data collected in order to make predictions for future.

ML models vary based on the research question and research perspective. To choose right ML algorithm is always an crucial part of research as handling huge data is difficult. Machine learning techniques results in cost effective solutions in agricultural sector. Random Forest and ANN are used to build a model to predict the crop yield of millet. ANN uses various learning algorithms to model input and output relationships.

Non-linear function to an individual neuron have little intrinsic convergence when neurons work together. In given scenario various Machine learning algorithms are being used to develop statistical model thus ANN inspired or partially modelled by biological neural networks. The study is carried out on datasets of 800 samples of cereals and different types of millets. Crop production varies accordingly by various climatic change affects the crop yield like dried period, increasing in temperatures are a challenge to agriculture workers, governments and traders to strengthen exactness & analysing crop yield in different climatic conditions.

In this system, a machine-learning method, ANN & Random Forest algorithm has an ability to analyse crop growth to derive confusion matrix to establish an error rate. The Datasets are obtained from varied sources which are involved for both training & testing process.

REVIEW OF LITERATURE

Soil moisture is one of the main factors in agricultural production and hydrological cycles, and its precise prediction is important for the rational use and management of water resources.[1] Technique inferred to predict soil moisture using machine learning equipment known as KD2 Pro, Using soil thermal properties portable equipment. [2].Soil moisture is a crucial parameter for crops growth and crop yield prediction, soil moisture is predicted using various mathematical models like PTFs (Pedotransfer Functions) [3].Yield prediction is an challenging area in terms of precision agriculture, carious models are proposed and validated till day, the key challenge is acquiring various datasets and dependency factors like climate, weather, soil ,use of fertilizer and seed variety [4]. In order to estimate crop yield many prediction models are proposed but a better performance model is still desirable in terms of actual yields [5]. Crop yield prediction learning model is designed based on various climatic factors this paper also advises to further expand over various parameters to get better accuracy[6]. ANN predicting system is used to predict non-linear interaction based on a given inputs. ANN depicts a neural pathway of brain, in order to predict the output a neural network requires training the data. Input data may be complex, multivariate and non-linear neural networks produces an optimal results. [8, 9,10]. Soil moisture index is an important parameter to be evaluated and validated for the crop estimation [11].

SYSTEM MODEL

Machine learning approaches are used to make crop yield prediction using parameters like Moisture rainfall, Average Humidity, Temperature, soil texture like alkaline, chalky, and sandy crop yield prediction for millet using machine learning method. The primary goal is to collect data and prepare for future use. Predictive analytics is used to predict future outcomes.

Hidden Layers ANN

b represents the bias associated with a node

h1 represents hidden layer node 1

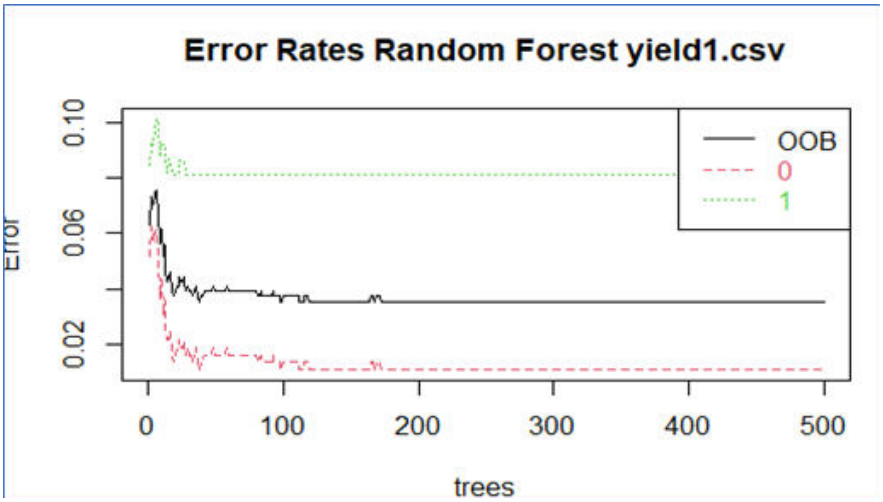
i1 represents input node 1 (i.e., input variable 1)

o represents the output node

Table 1. Input for Crop Simulation Model

Factor	Description
Soil moisture	Moisture content of the soil subject to hydrological changes
Weather	Daily global soil radiation, daily maximum and minimum temperature, daily rainfall Average Humidity
Rainfall	Soil type, soil depth (divided by n layers), soil texture, soil organic carbon, bulk density, soil nitrogen, pH
Soil type	Sandy Chalky Alkaline
Outcomes	Yield prediction high or low

Fig 3 Proposed Block Diagram For the Study



FINDINGS:

The following table gives the Confusion matrix of random forest and Neural Networks and its error rate.

Table 2. Error Rate

Error matrix for the Neural Net model (test) (counts):	Error matrix for the Neural Net model [test] (proportions):	Overall error
Predicted Actual 0 1 Error 0 160 0 0 1 8 72 10	Predicted Actual 0 1 Error 0 66.7 0 0 1 3.3 30 10	3.3%, Averaged class error: 5%
Random Forest	Confusion matrix: 0 1 class. Error 0 359 4 0.01101928 1 16 181 0.08121827	OOB estimate of error rate: 3.57%

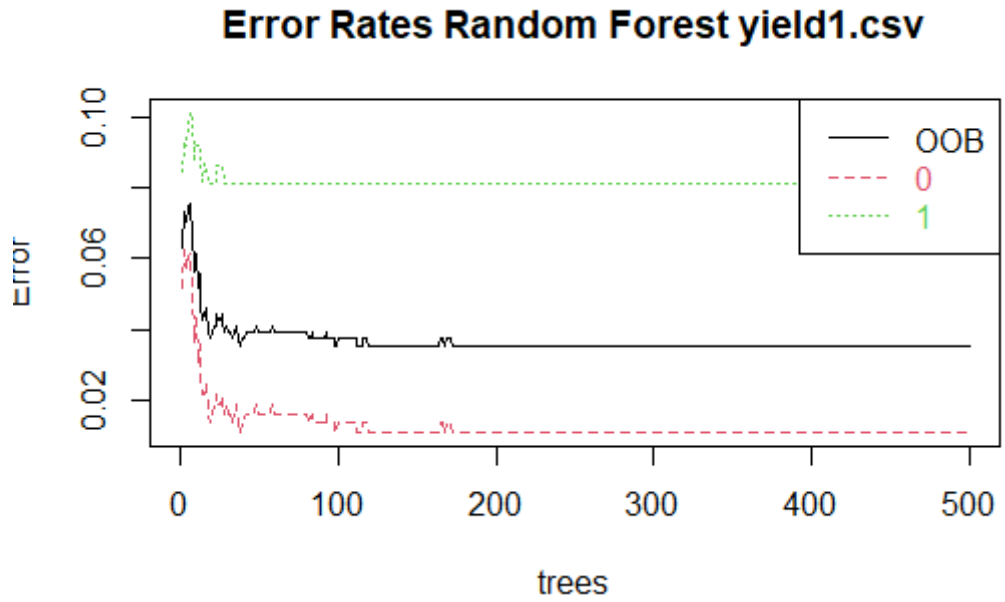


Fig 3.1 The Error rate of RF Confusion Matrix

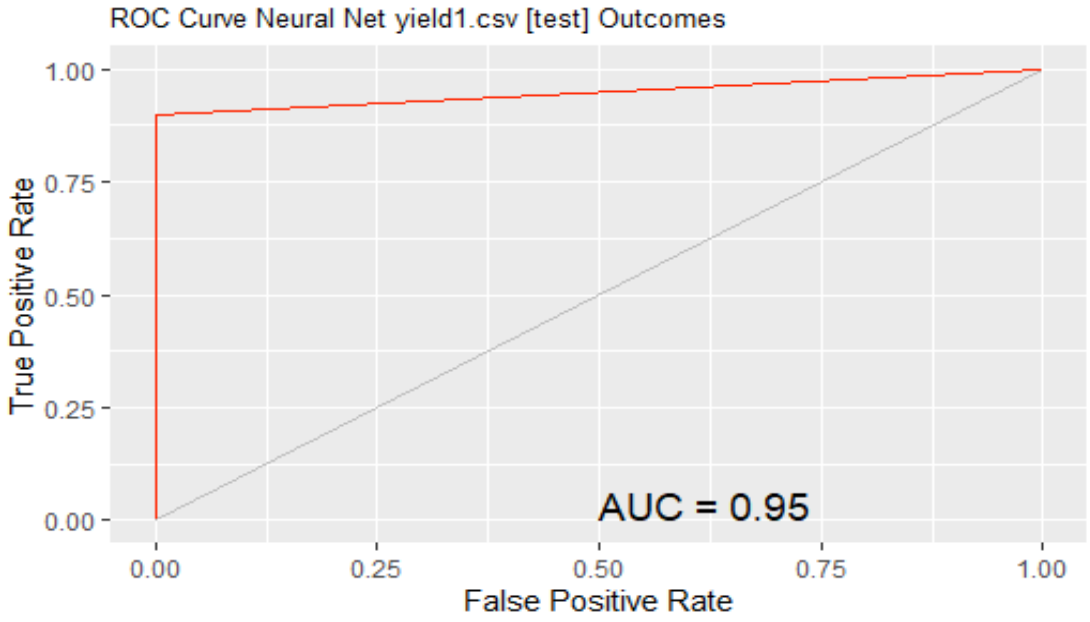


Fig 3.2 AUC CURVE of NN(Artificial Neural Networks)

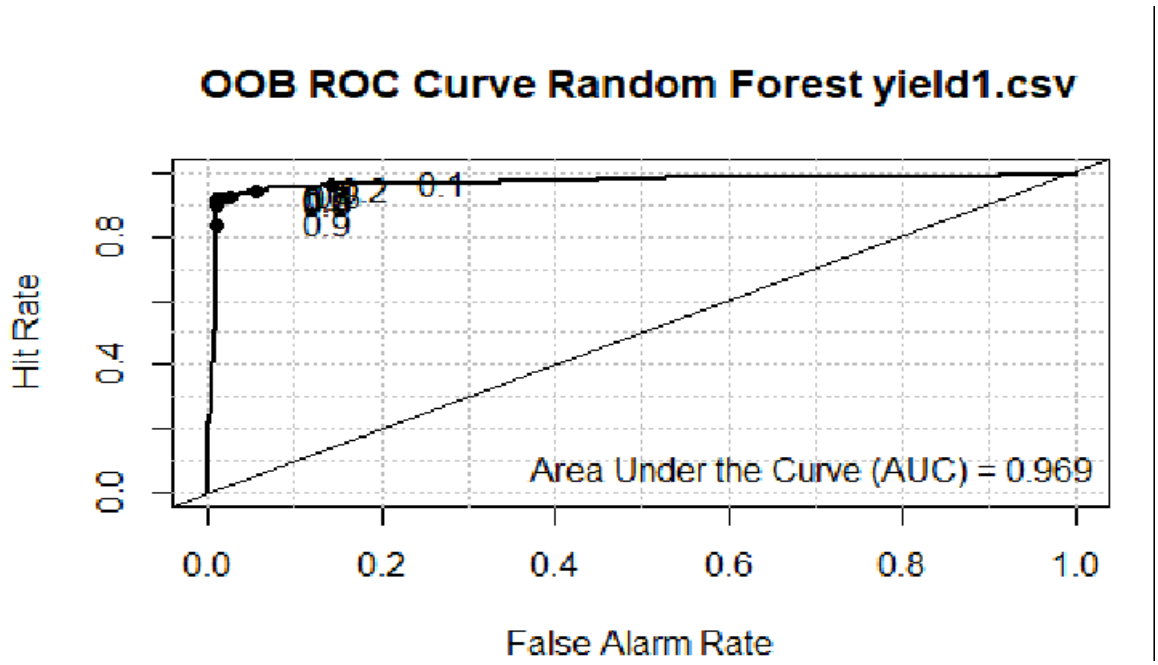


Fig 3.2 AUC CURVE of RF(Random Forest)

VARIABLE IMPORTANCE OF RANDOM FOREST

	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
Sandy	33.51	51.97	45.93	85.67
Alkaline	12.49	20.34	18.08	21.95
Clay	11.37	17.97	15.95	14.90
Mean Temp	11.48	8.56	14.61	3.57

Min Temp	12.38	6.87	14.28	3.56
Millet Yield	12.31	6.15	12.55	8.20
Avg. Humidity	10.67	5.54	11.60	5.58
Chalky	9.94	10.81	11.50	3.41
Max.Temp	8.72	4.96	10.56	4.41
Rainfall	7.52	4.04	8.33	2.78
Moisture	1.29	2.73	2.73	5.05

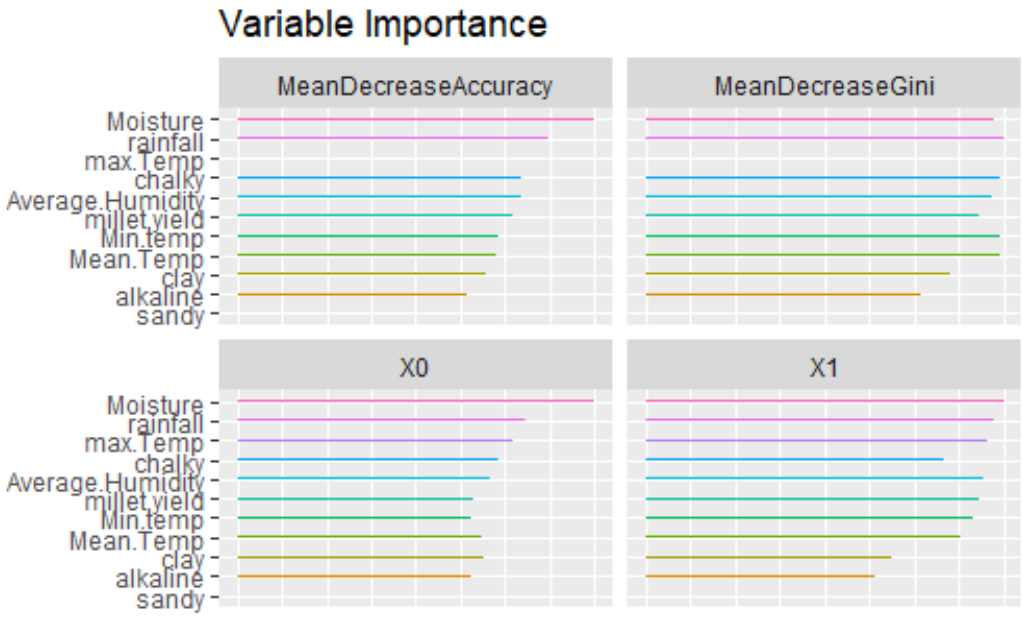


Fig 3.4 Variable of Importance NN

Random Forest classifier has huge ability to predict crop yield. From different outputs, it shows that Random Forest is an efficient learning algorithm to analyse crop at current climatic condition and has a huge exactness in data investigation. The Study states that evaluating the performance of actual target values to the predicted values by the machine learning models infer that both ANN & RF have better performance in crop yield estimation considering error rate ANN shows better prediction results.

CONCLUSION

In future research can be elaborated by applying many more machine learning models like logistic regression, decision trees or CHAID. Machine learning approaches are used to estimate crop yield in agriculture. The random forest method and artificial neural networks are used for the study and identified ANN algorithm has more accuracy in crop yield estimation. Though Random forest and neural networks are different domains they belong to same class of Machine Learning algorithms.

REFERENCES

[1] Cai Y, Zheng W, Zhang X, Zhangzhong L, Xue X (2019) Research on soil moisture prediction model based on deep learning. PLoS ONE 14(4): e0214508. https://doi.org/10.1371/journal. pone.0214508

[2] Sanuade, O. A., Hassan, A. M., Akanji, A. O., Olaojo, A. A., Oladunjoye, M. A., & Abdulraheem, A. (2020). New empirical equation to estimate the soil moisture content based on thermal properties using machine learning techniques. Arabian Journal of Geosciences, 13(10), 377. https://doi.org/10.1007/s12517-020-05375-x

[3] Kaingo, J., Tumbo, S. D., Kihupi, N. I., & Mbilinyi, B. P. (2018). Prediction of Soil Moisture-Holding Capacity with Support Vector Machines in Dry Subhumid Tropics. Applied and Environmental Soil Science, 2018. https://doi.org/10.1155/2018/9263296

[4] X. Xu, P. Gao, X. Zhu, W. Guo, J.Ding, C. Li, X. Wu

[5] Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China Ecol. Ind., 101 (2019), pp. 943-953

[6] P. Filippi, E.J. Jones, N.S.Wimalathunge, P.D.S.N. Somarathna, L.E. Pozza, S.U. Ugbaje, T.F.A. BishopAn approach to forecast grain crop yield using multi-layered, multi-farm datasets and machine learning. Precis. Agric., 1–15 (2019)

[7] Alpaydin, E., 2010. Introduction to Machine Learning, 2nd ed.

[8] D. Elavarasan, D.R. Vincent, V.Sharma, A.Y. Zomaya, K. Srinivasan Forecasting yield by integrating agrarian factors and machine learning models: a survey

[9] Kim, N and Lee, Y.W 2016, "Machine learning approaches to corn yield estimation using satellite images and climate data: A case of Iowa State", J. Korean Soc. Surv. Geod. Photogramm. Cartogr, 34, 383–390.

[10] K. Kuwata and R. Shibasaki 2015, "Estimating crop yields with deep learning and remotely sensed data," 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, pp. 858-861. doi: 10.1109/IGARSS.2015.7325900.

[11] V. Geetha, A. Punitha, M. Abarna, M. Akshaya, S. Illakiya and A. P. Janani, "An Effective Crop Prediction Using Random Forest Algorithm," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-5, doi:1109/ICSCAN49426.2020.9262311.

[12] Lakshmi CB, Dr.Malarvizhi. S, "SOIL MOISTURE CONTENT PREDICTION USING MACHINE LEARNING: A REVIEW", IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.6, Issue 2, Page No pp.117-125, June 2019, http://www.ijrar.org/IJRAR19K4693.pdf