iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VI **Month of publication:** June 2019

DOI: http://doi.org/10.22214/ijraset.2019.6200

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

Data Mining Techniques for Prediction of Wheat Crop Production

Akanksha Gahoi[1], Akhilesh Tiwari[2]
[1, 2]Department of CSE & IT, Madhav Institute of Technology & Science, Gwalior (M.P), India

Abstract: ***"Agriculture is the art and science of growing plants and other crops and for food, other human needs, or economic gain". India is an agricultural country.***

Two- third of the population is dependent on agriculture directly or indirectly. In this research work, the main focus is on analysis of prediction of wheat crop production of Madhya Pradesh state on the basis of area harvested, rainfall and soil type by using data mining techniques and also build an efficient model for prediction of wheat crop production.

Keywords: Data Mining, Artificial Neural Network (ANN), Generalized Linear Model (GLM), Random Forest, Recursive Partitioning and regression tree (RPART), Crop Prediction

I. INTRODUCTION

India is an agricultural country. Agriculture provides the highest contribution to national income. There are many serious problems like- Erosion, Diseases, Pests, Weeds, Drought, Rainfall that people face trying to grow food today.

Data Mining is simply an approach for extracting intention based knowledge from the raw data set. Data mining can also be used to explore huge data, the most frequent set of patterns in a dataset.

The main aim behind the actual Data mining procedure is to mine the information data from a large collection of data and change it into an explainable framework for additional use.

Data Mining techniques are mostly used in the scenario where prediction based problems need to be handled. Traditional approaches of data mining mostly suitable for basic clustering, forecasting and classification problems. Data mining can be used for Prediction, Identification, Classification, and Optimization.

Applying the Data Mining classification and prediction techniques on climate-related data and crop-related data. Various predictions can be made on the basis of predicted result which can help in increasing crop production rate. Data mining plays a crucial role in decision making on several issues related to the agriculture field.

Applicability of data mining techniques in agriculture, crop planning, and crop management can change the way of farming farmers can yield in a much efficient way.

II. LITERATURE REVIEW.

A. In 2011 D. Rajesh described "Application of Spatial Data Mining for Agriculture" In this paper author have proposed a method to extract a pattern from spatial database using k-means clustering algorithm [6].

B. In 2015 A.T.M Shakil Ahamed et. al tested few data mining techniques for prediction of the annual yield of major crops in Bangladesh. In this paper, the clustering technique is used to predict the results. There are parameters such as temperature, humidity, minimum temperature, maximum temperature, average sunshine, Soil PH and salinity are used to predict the Annual crop production. K- means clustering is used by the author for recommending plant crops in the districts of Bangladesh[4].

C. In 2018 Belabed Image et. al., proposed an approach for extracting information by using data mining approaches in three domains, bioinformatics, medicine, and agriculture industry. Initially, the variables are clustered to increase the functionality, and the association rules are used between the target variables and the previously identified set of variables [5].

III. RESEARCH METHODOLOGY

A. Dataset

Data used in this experimental is verified from- data.gov.in and indiawateerportal.org. There were wheat crop data of 5 years of 50 districts of Madhya Pradesh. The attributes are in this data set are- district, year, area (in hectare), production (in metric ton), rainfall (in mm) and soil type.

Parameters included are:

1) District
2) Year
3) Area of Harvest
4) Production of Wheat
5) Rainfall
6) Type of Soil

Data Set may contain other information according to crop behavior.

	A	B	C	D	E	F
1	District	Year	Area	Productio	Rainfall	soiltype
2	Allirajpur	2007	14048	25310	922.2	Deep Medium Black
3	Anuppur	2007	11964	8230	948.2	Deep Medium Black
4	Ashoknag	2007	104899	119871	810	Deep Medium Black
5	Balaghat	2007	14672	13166	1257.1	Deep Medium Black
6	Barwani	2007	34398	51908	733.6	Deep Medium Black
7	Betul	2007	99067	153407	1301	Shallow & Medium Black
8	Bhind	2007	73949	102508	478.6	Alluvial Soil
9	Bhopal	2007	68225	120087	952.8	Deep Medium Black
10	Bhurhanp	2007	10380	19911	1060	Deep Medium Black
11	Chhatarpu	2007	57707	60129	575.2	Mixed Red & Black
12	Chhindwa	2007	110732	223408	927.5	Shallow & Medium Black
13	Damoh	2007	63063	99767	888.9	Deep Medium Black
14	Datia	2007	88482	151820	568.4	Deep Medium Black
15	Dewas	2007	118746	215031	1013.3	Deep Medium Black
16	Dhar	2007	216267	500035	1161	Deep Medium Black
17	Dindori	2007	29692	14958	975.2	Mixed Red & Black
18	Guna	2007	79286	113969	765.8	Alluvial Soil
19	Gwalior	2007	67870	111365	540.5	Alluvial Soil
20	Harda	2007	125786	412619	1140.6	Deep Medium Black
21	Hoshanga	2007	217612	698551	1092.8	Deep Medium Black
22	Indore	2007	126748	244343	960.5	Deep Medium Black
23	Jabalpur	2007	85511	126551	1033.4	Deep Medium Black

Dataset

Figure. 2 (Wheat Crop Data Set)

IV. PROPOSED METHODOLOGY

For testing of this model raw agricultural data is taken. Other agricultural data can also be used as an input in this model. The data should be authentic and verified by any governmental organization or an agricultural research institute. The authentic data generates more accurate results and those results can lead towards the efficient generation of rules for optimized farming.

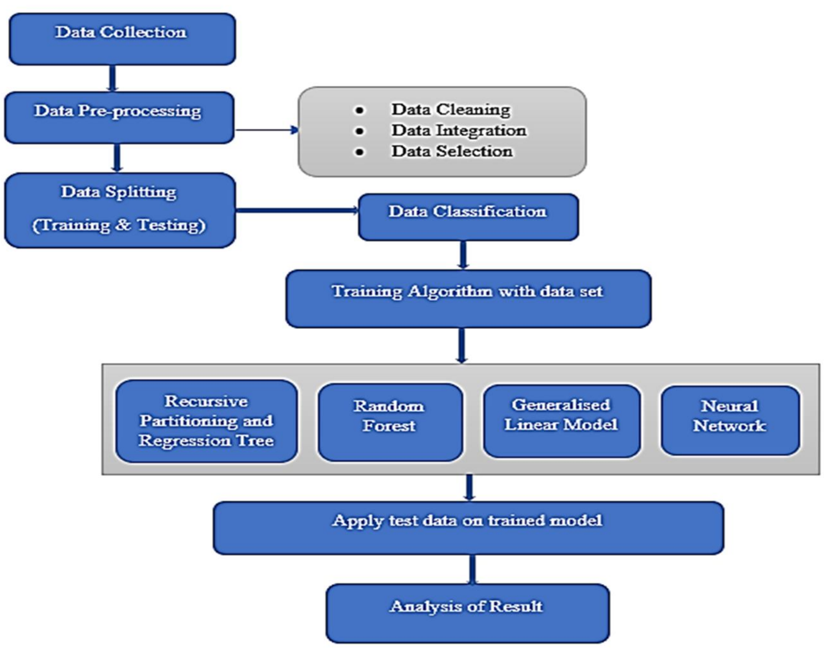


Figure. 1 (Work Flow of Proposed methodology)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

1) *Data Collection:* Raw Data generated every year contains lots of information about the crop. Certain features associated with a crop makes the data quite large.
2) *Data Preprocessing:* This step makes the data ready to be processed. Processes include handling of noisy values, removal of redundant values. Selection of proper attributes, etc.
3) *Data Splitting:* For training of the model certain amount of data is required.similarly for testing and validation the remaining amount of data is provided. Data Splitting includes the ratio in which training data and testing data is separated.
4) *Classification:* Classification is the process of classifying data into class labels. These labels are generated on the basis of parameters.
5) *Prediction:* Predictive approaches are applied to determine the valid functions on the basis of their continuity.

The classification and predictive approaches tested for wheat crop data are as follows:

a. *RPART:* RPART is a *"Recursive Partitioning and Regression Tree"*. The RPART approach classifies the data on regression models. It includes 2 stage procedure, the output can be represented as a binary tree.
b. *Random Forest:* Belongs to the family of supervised learning approaches, suitable for the classification and regression problems as well. Basic working ideas behind this approach are multiple collections of tree-structured classifiers.
c. *Generalized Linear Model:* Generalised linear model, applicable to the data contains numerical as well as the continuous target variable. The approach computes the response of the explanatory variable by modeling a linear function along with the error term association.
d. *Artificial Neural Network:* An artificial neural network often just called a neural network. ANN is an efficient and most used approach in the field of the computing system. The idea behind the ANN is based on Biological Neurons in the human brain.

V. EXPERIMENT

From fig. 3-6 shows the comparison between actual production and predicted production by using RPART, Random Forest, Generalized Linear Model, and Neural Network.

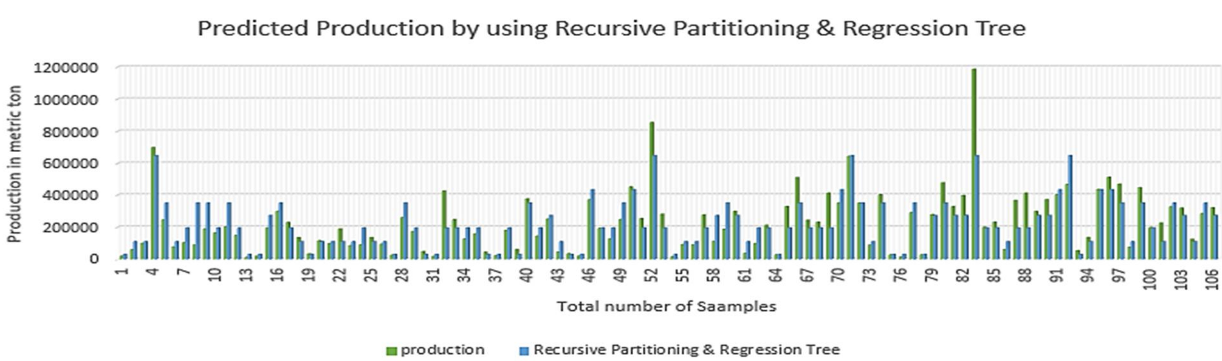


Figure. 3 Actual Production Vs Predicted Production by RPART

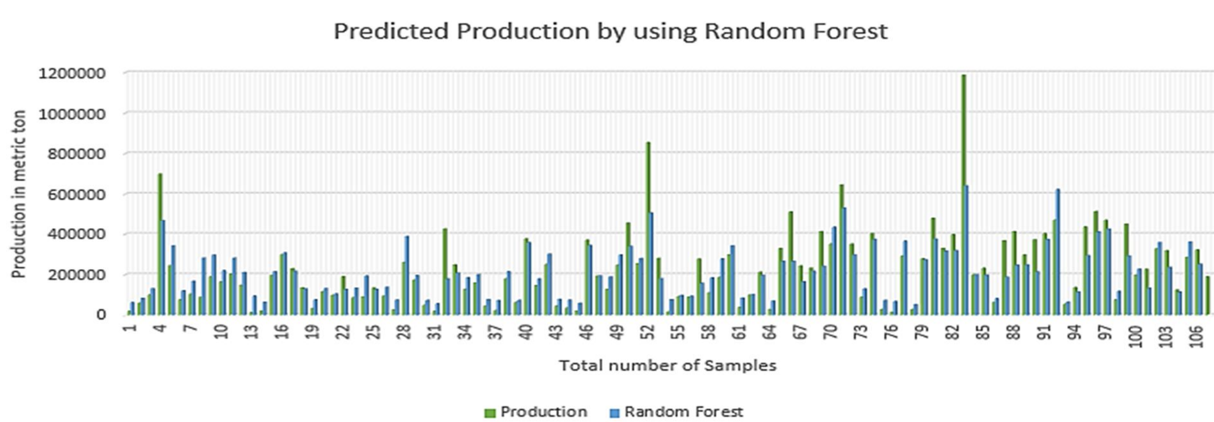


Figure. 4 Actual Production Vs Predicted Production by Random forest

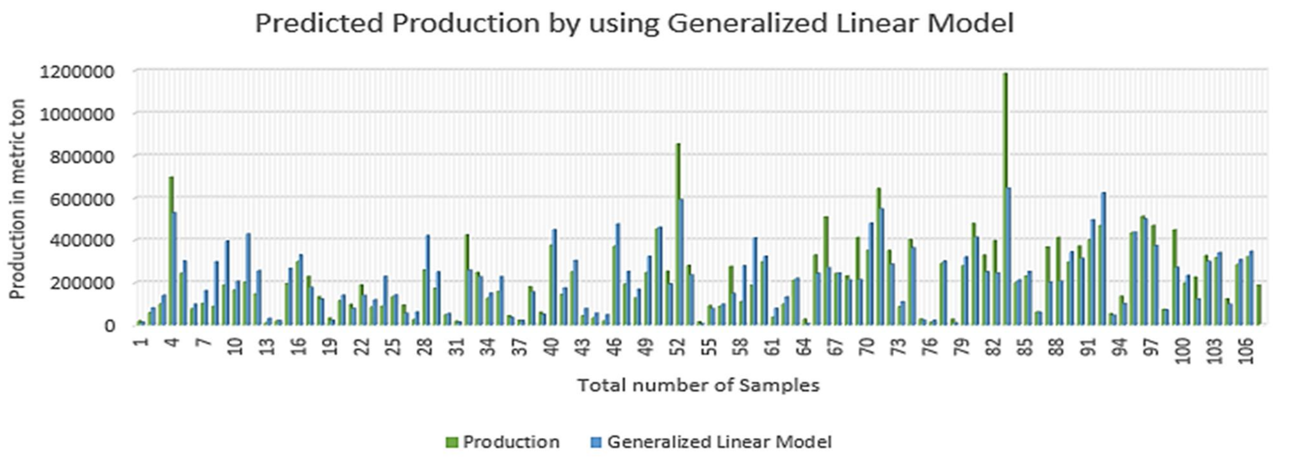


Figure. 5 Actual Production Vs Predicted Production by GLM

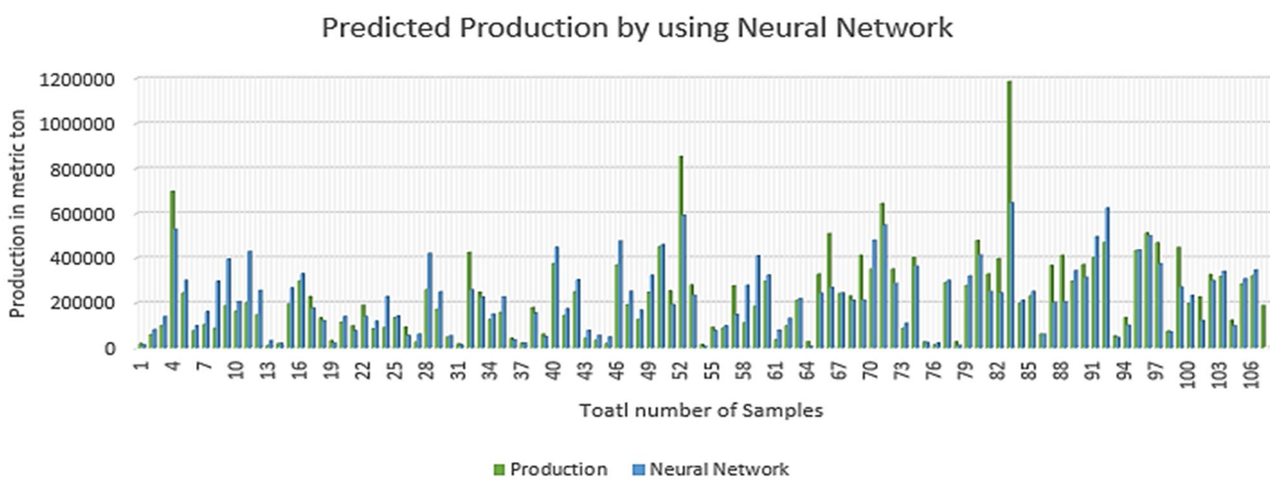


Figure. 6 Actual Production Vs Predicted Production by Artifical Neural Network

VI. RESULTS

Figure. 7 show the comparison between the actual product and predicted the production of the wheat crop on the basis of area, rainfall, and soil type by using four classifiers- RPART, Random Forest, Generalised Linear Model, and Neural Network. And fig. 8 show the accuracy and error comparison between these classifiers.

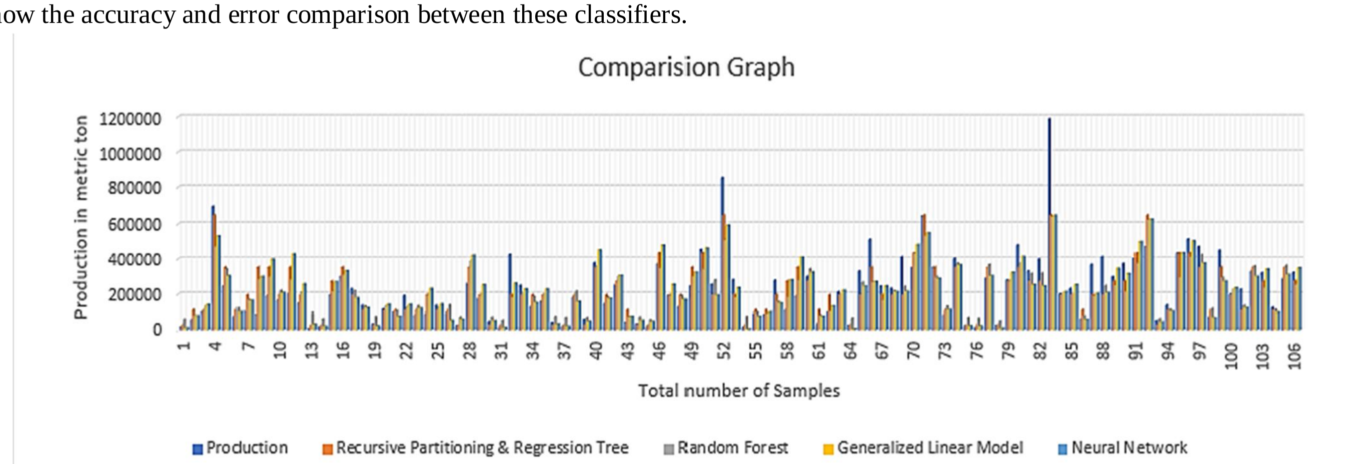


Figure. 7 Comparison Graph between Actual Production and Predicted Production

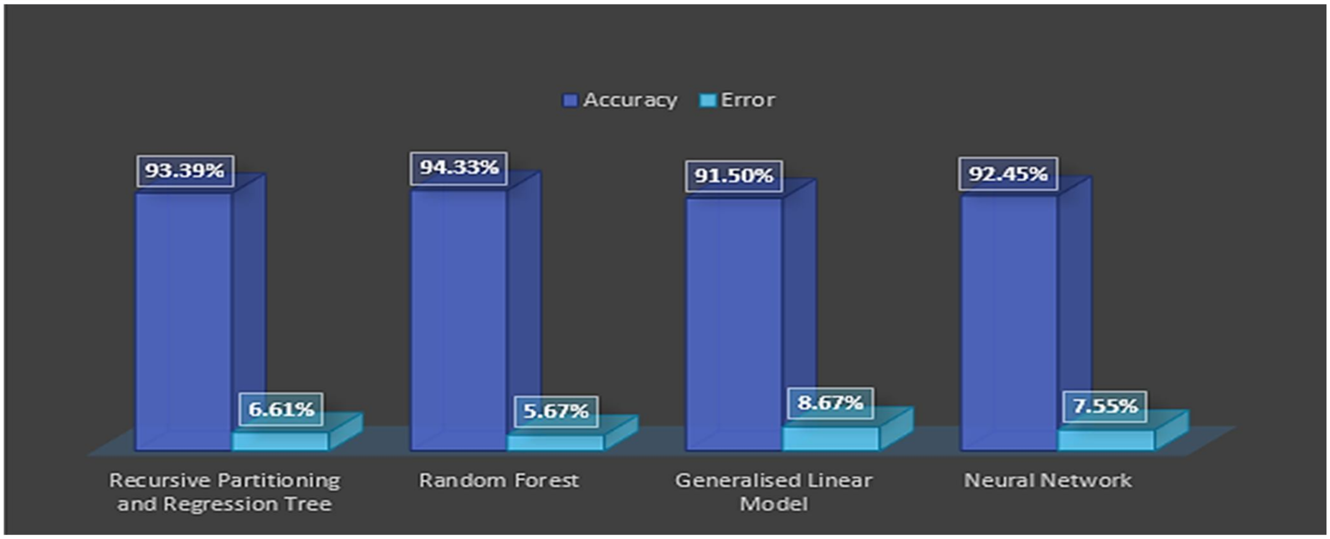


Figure. 8 Accuracy Comparison

VII. CONCLUSION

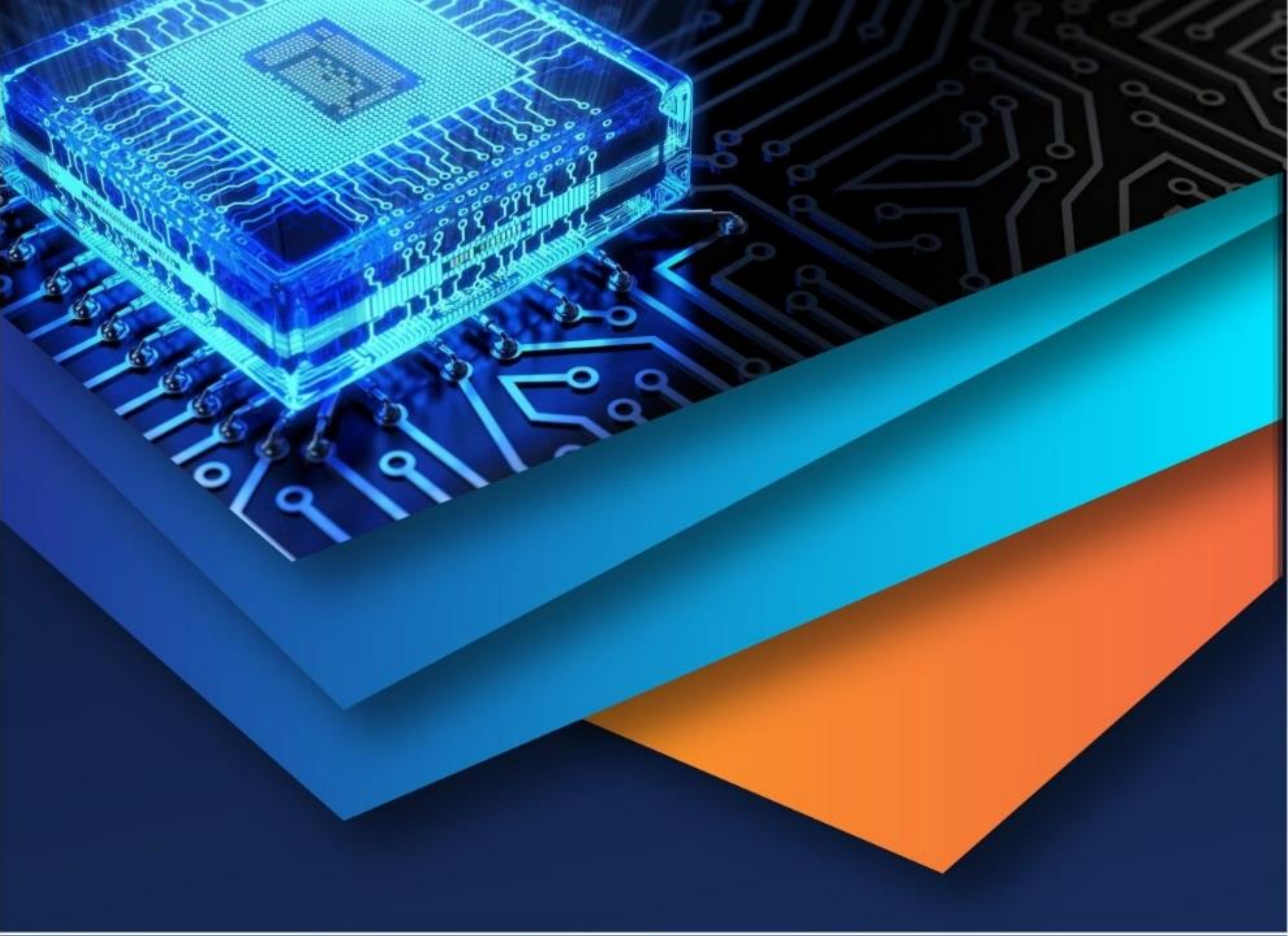
In this research, four classifiers are used namely are Recursive Partitioning & Regression tree, Random forest, Generalized linear model and Neural Network. By providing mentioned parameters, the above study shows that Random Forest is an efficient classifier for prediction of wheat crop production. The accuracy of this model is 94.33% and the error is 5.67%.

VIII. FUTURE WORK

The efficiency of this model can be increased by integrating Machine learning algorithms in the future to get more accurate results. These classifiers can be applied for prediction of other crops produced in India and other countries as well. For improving efficiency, existing optimization techniques can also be applied after slight parameter tuning.

REFERENCES

[1] Shruti Mishra, Priyanka Paygude, Snehal Chaudhary and Sonali Idate, "Use of Data Mining in Crop Yield Prediction", Second International Conference on Inventive System and Control (ICISC 2018).
[2] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh", IEEE 2018.
[3] Belabed Imane, Talibi Alaoui Mohammed, Talibi Alaoui Youssef, "Data mining approach based on clustering and association rules applicable to different fields", IEEE 2018.
[4] D.Rajesh, "Application of Spatial Data Mining for Agriculture", International Journal of Computer Applications (0975 – 8887) Volume 15– No.2, February 2011.
[5] Vrushali Y Kulkarni, Pradeep K Sinha, "Effective Learning and Classification using Random Forest Algorithm", IJEIT, Volume 3, Issue 11, may 2014.
[6] Miss.Snehal S.Dahikar, Dr.Sandeep V.Rode, "Agricultural Crop Yeild Prediction Using Artificial Neural Network Approach", IJIREEICE, Vol. 2, Issue 1, January 2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY