Comparative study on Enhanced Random Forest Algorithm by Crop Prediction

[*]Priyankar Ravindra Tiwari, [#]Dr. Anand Khandare

[*]M.E. Scholar, [#]Assistant Professor, Thakur College of Engineering and Technology, Mumbai, India, [*]priyankartiwarimy@gmail.com, [#]anand.khandare1983@gmail.com

Abstract: Earlier yield prediction was performed by considering the farmer's experience on a particular field, weather and crop. Since farmers don't have knowledge about the presence of the nutrients and they don't have the idea about the crop to plough and pesticides to be used due to which performance of agriculture is degrading in economy and farmers getting into loss which they have to pay from their own pocket, which is also a root cause of farmers suicide. This makes the problem of predicting the yielding of crops an interesting challenge.

This work presents a system, which uses machine learning techniques called Random Forest with enhanced performance in order to predict the category of the analysed soil datasets. The category, thus predicted indicates the yielding of crops. The problem of predicting the crop yield is formalized as a classification and regression rule, where Enhanced Random Forest is divided into classification and regression, used for categorization of soil and predicting rainfall and both the result will provide the predicted crop to plough.

Keywords — Random Forest, classification, regression, enhanced Random Forest, decision tree, ID3, crop prediction

I. Introduction

Agriculture is the backbone of the Indian. The agriculture data increases day by day. Since an outsized population lives in rural areas and is directly or indirectly captivated with agriculture for a living. Outlay from farming forms the main source for the farming community. The essential requirements for harvesting are water resources and ability to buy seeds, fertilizers, pesticides, labour etc. Most farmers raise the required capital by compromising on other essential expenditures, and when it is still insufficient, they resort to credit from sources like banks and private commercial institutions. In such a situation, the repayment is dependent on the success of the harvest. If the harvest fails even once because of many factors, like atmospheric condition pattern; soil type; improper, excessive, and ill-timed application of each fertilizers and pesticides; debased seeds and pesticides etc. Most power of soil in nature comes from soil survey efforts. Soil survey, or soil mapping, is the process of determining available nutrients in soil or other holding of the soil cover over a landscape, and mapping them for others to understand and use. Primary data for the soil survey is acquired by area sampling and supported by remote sensing. As the volume of data increase, it requires involuntary way for these data to be extracted when needed. Machine Learning can be used for pretend the next trends of agricultural processes. Every soil is a mixture of these component: Nitrogen, Phosphorus, Potassium, pH Value and Electrical Conductivity. Based on these factors we predict the soil fertility level and crop for a particular soil sample.

In this context, the goal of this paper is to provide a comprehensive, comparative and self-contained analysis of a class of algorithms known as ID3 decision trees and random forests and enhanced random forest. These methods have proven to be a robust, accurate and successful tool for solving countless of machine learning tasks, including classification, regression, density estimation, manifold learning or semi-supervised learning.

1.1 Problem Definition:

A brief study of problems related to maximization of the productivity and prediction of crop yield has been done by going through the related literature review, and with the brief discussions with soil analysts and farmers and broader view of research problem has been gained. Yield prediction is incredibly well-liked among farmers currently, that notably contributes to the right choice of crops for sowing. This makes the problem of predicting the yielding of crops an interesting challenge. Earlier yield prediction was performed by considering the farmer's expertise on a selected field and crop. This work presents a system, which uses Machine Learning techniques in order to predict the category of the analysed soil datasets. The category, thus predicted indicates the yielding of crops.

DOI : 10.35291/2454-9150.2019.0291

1.2 Motivation:

Farmers in India, specially Vidarbha region in Maharashtra state faces drought due to which their crop and yielding is getting degraded. They don't have any idea about availability of nutrient in their field. They use their own experience to plough the crop which have very less success ratio. Due to less success ratio they are unable to pay their loan amount sanctioned for their crop. In unsuccessful for their repayment of the loan amount they attempt to suicide which is a main reason for highly rising ratio in farmers suicide.

To help the farmers to decide the crop to be plough for their benefits I am motivated to build this system. This system collects the data from the soil testing laboratory supported by Department of Agriculture, Government of India. This dataset consists of the available nutrient for farmers' soil and rainfall for particular region.

II. LITERATURE REVIEW

Random Forest (RF) is an ensemble classifier proposed by Breiman (2001) which consists of many sub-models. The predictions and other quantities of interest are obtained by combining the outputs of all the sub-models. The sub-models for Random Forest are classification and regression trees (CART) which is the key for understanding the Random Forest.

In the past decade, various methods have been proposed to grow a random forest (Breiman, 2001; Dietterich, 2000; Ho, 1998). Among these methods, Breiman's method (Breiman, 2001) has gained increasing popularity because it has higher performance against other methods (Banfield et al., 2007).

Let D be a training dataset in an M-dimensional space X, and let Y be the class feature with total number of c distinct classes. The method for building a random forest (Breiman, 2001) follows the process including three steps (Baoxun Xu et al., 2012):

Step 1: Training data sampling: use the bagging method to generate K subsets of training data {D1, D2, ..., DK} by randomly sampling D with replacement;

Step 2: Feature subspace sampling and tree classifier building: for each training dataset Di (1≤ i ≤ K), use a decision tree algorithm to grow a tree. At each node, randomly sample a subspace Xi of F features (F << M), compute all splits in subspace Xi, and select the best split as the splitting feature to generate a child node. Repeat this process until the stopping criteria is met, and a tree hi(Di, Xi) built by training data Di under subspace Xi is thus obtained;

Step 3: Decision aggregation: ensemble the K trees {h1(D1, X1), h2(D2, X2),...hK(DK, XK)} to form a random forest and use the majority vote of these trees to make an ensemble classification decision. (i.e., majority votes for classification, average for regression).

The algorithm has two key parameters, i.e., the number of K trees to form a random forest and the number of F randomly sampled features for building a decision tree. According to Breiman (2001), parameter K is set to 100 and parameter F is computed by F=[log2M + 1]. For large and high dimensional data, a large K and F should be used.

The estimation of the error rate can be obtained based on the training data as follows:

1. At each bootstrap iteration, predict the data not in the bootstrap sample (what Breiman calls "out-of-bag", or OOB data) using the tree grown with the bootstrap sample.

2. Aggregate the OOB predictions. (On the average, each data point would be out-of-bag around 36% of the times, so aggregate these predictions.) Calculate the error rate, and call it the OOB estimate of error rate.

2.1 Advantages of Random Forest:

1. Accuracy is as good as Adaboost and sometimes better.
2. It is faster than bagging or boosting.
3. It gives useful internal estimates of error, strength, correlation and variable importance.
4. It is simple and easily parallelized

2.2 Disadvantage of Random Forest:

1. Models in Random Forest which has been overfit will have poor predictive performance as it doesn't generalize well. Generalization means how well model makes prediction for the cases that are not in training set.
2. In Random Forest Algorithm we need to choose number of trees.
3. Large number of attributes for prediction and large number of trees makes algorithm slower.
4. For data including categorical variables with different number of levels, random forests are biased in favour of those attributes with more levels. Therefore, the variable importance scores from random forest are not reliable for this type of data.

2.3 Related Work:

Over the past decade, some research was invested in boosting the performance of RF. One of the earliest to be reported is by Latinne *et al.* (2001). A method based on the McNemar non-parametric test of significance was proposed. The method a priori determines the minimum number of trees in the RF to use in order to obtain prediction accuracy comparable to the one obtained with larger ensembles. In addition to maintaining accuracy with fewer trees, the method significantly improves classification speed and reduces memory costs.

Robnik-Šikonja (2004) investigated new ways to improve the performance of RF. By using several attribute evaluation measures instead of just one, the correlation

DOI : 10.35291/2454-9150.2019.0291

between trees is decreased without any loss in their strength. Another way to improve the performance of RF is to change the voting method. Instead of using majority voting, weighted voting is used. With this voting technique, internal estimates are used to identify instances most similar to the instance being labeled. The votes of the corresponding trees are then weighted with the strength they demonstrate on these near instances. Improvements were demonstrated on several classification data sets.

Tsymbal *et al.*, (2006) found a way to improve the performance of RF on some data sets by replacing majority voting with more sophisticated dynamic integration techniques. Three techniques were used: Dynamic Selection (DS), Dynamic Voting (DV), and Dynamic Voting with Selection (DVS). Using DV and DVS integration strategies, experimental studies showed that dynamic integration was able to improve the accuracy of RFs on 12 out of 27 data sets.

III. METHODOLOGY

Improving accuracy in classification and prediction has been grasping a lot of attention from many researchers all over the world. Random Forest is a new approach to data exploration, data analysis, and predictive modelling. This research work focuses on improving the performance of random forest in three aspects.

3.1 Enhanced Random Forest Algorithm:

The standard algorithm has two key parameters, *i.e.*, the number of *n* trees to form a random forest and the number of *F* randomly sampled features for building a decision tree. According to Breiman (2001), parameter *K* is set to 100 and parameter *F* is computed by F=[log2M + 1].

To enhance the algorithm, the samples feature should not be limited. For this, in enhanced algorithm number of F max_feature is randomly is entered by algorithm and the best one is selected for the system. Also the standard algorithm need to be entered the number of "*n*" Trees to a random forest and the larger the number of the trees slower the speed of algorithm. Thus, to resolve this enhanced algorithm is implemented with random function which randomly choose the number of trees and checks the result for each and select the number of tree which has the best result and uses that for all the further steps.

1. Randomly select "k" features from total "m" features.
   Where k << m
2. Among the "k" features, check for every feature and select "f" best feature.
3. For the feature "f" calculate the node "d"
4. Split the node "d" into daughter node "l" using best split.
5. Repeat 1 to 3 steps until "l" number of nodes has been reached.
6. Randomly put the value of number of trees and select "n" the number giving best result
7. Build forest by creating "n" number of trees.

The above algorithm generates forest by following above algorithm which is enhanced to improve performance. After building forest, classification and regression is applied as below:

1. Draw n_{tree} bootstrap samples from the original data.
2. For each of the bootstrap samples, grow an unpruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample m_{try} of the predictors and choose the best split from among those variables. (Bagging can be thought of as the special case of random forests obtained when, m_{try} = p, the number of predictors.)
3. Predict new data by aggregating the predictions of the n_{tree} trees (i.e., majority votes for classification, average for regression).

3.2 Flowchart of Enhanced Random Forest Algorithm:

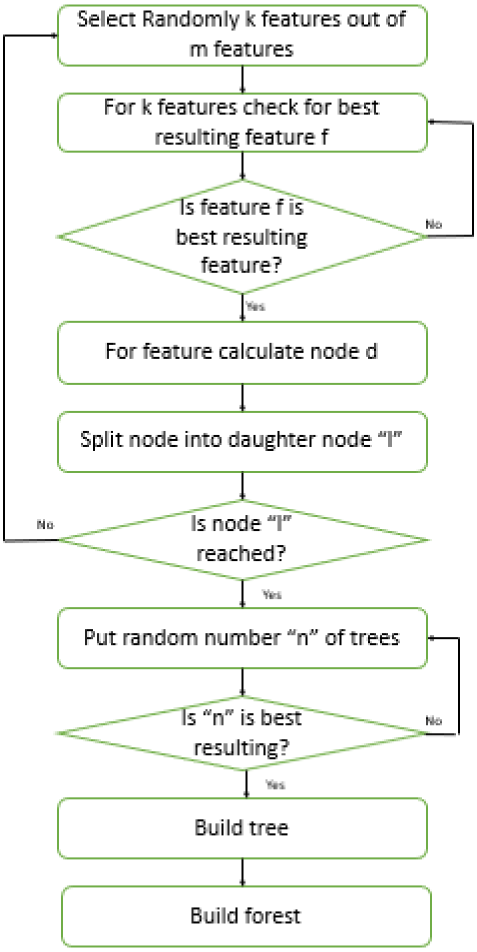

Select Randomly k features out of m features
For k features check for best resulting feature f
Is feature f is best resulting feature?
No
Yes
For feature calculate node d
Split node into daughter node "l"
Is node "l" reached?
No
Yes
Put random number "n" of trees
Is "n" is best resulting?
No
Yes
Build tree
Build forest

Fig 1: Flowchart of Enhanced Random Forest Algorithm

IV. RESULT

The algorithm measure performance on different parameters. Few parameters like accuracy, OOB Error Rate, Confusion Matrix, Error Rate, Mean Squared Error, R2 Score.

DOI : 10.35291/2454-9150.2019.0291

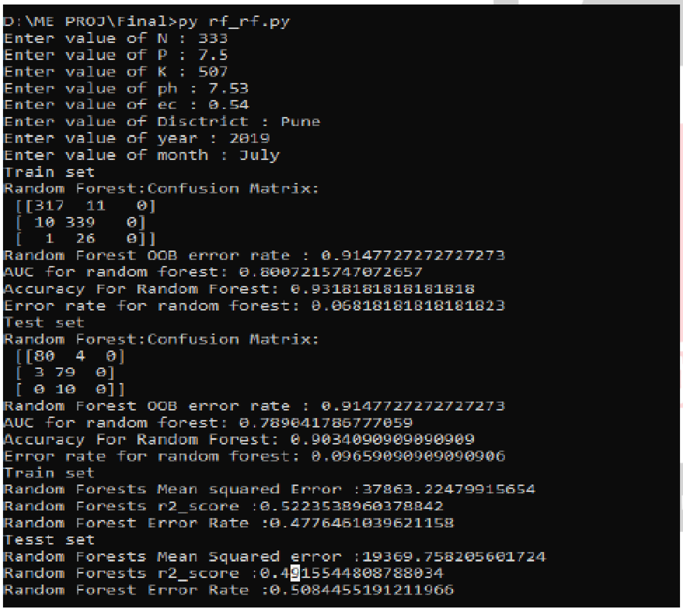
4.1 *Dataset:*

Datasets used in this system numeric value which consist of the nutrient value in different unit and rainfall in different area is recorded in mm (Millimeter). The soil dataset consist of the nutrient value for the classification of fertility level. There are total 880 soil sample have been gathered and based on that these data sets have been prepared. The dataset have been devided into 80:20 for training and testing respectively.

The rainfall dataset consist of rainfall record of maharashtra region in different state. There are 3168 record have been used for the prediction and it also have been devided into 80:20 for training and testing. The rainfall data contains rainfall record month wise from year 2010 to 2017. It is used to predict the rainfall for the required year and month.

Crop dataset which records the list of crop, required minimum and maximum rainfall and fertility level of the soil in which it can be grown.

4.2 *Result for Standard Random Forest:*



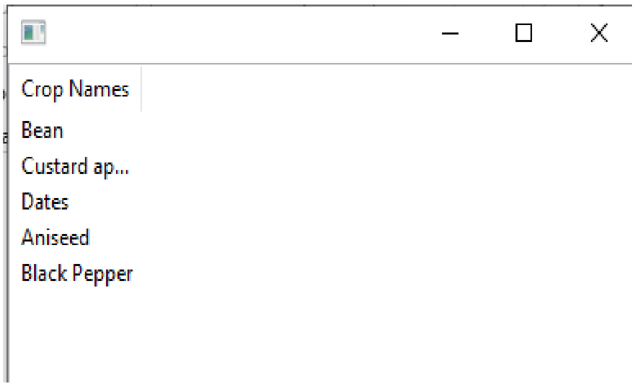
```
D:\ME PROJ\Final>py rf_rf.py
Enter value of N : 333
Enter value of P : 7.5
Enter value of K : 507
Enter value of ph : 7.53
Enter value of ec : 0.54
Enter value of Disctrict : Pune
Enter value of year : 2019
Enter value of month : July
Train set
Random Forest:Confusion Matrix:
 [[317  11   0]
 [ 10 339   0]
 [  1  26   0]]
Random Forest OOB error rate : 0.9147727272727273
AUC for random forest: 0.8007215747072657
Accuracy For Random Forest: 0.9318181818181818
Error rate for random forest: 0.06818181818181823
Test set
Random Forest:Confusion Matrix:
 [[80  4  0]
 [ 3 79  0]
 [ 0 10  0]]
Random Forest OOB error rate : 0.9147727272727273
AUC for random forest: 0.789041786777059
Accuracy For Random Forest: 0.9034090909090909
Error rate for random forest: 0.09659090909090906
Train set
Random Forests Mean squared Error :37863.22479915654
Random Forests r2_score :0.5223538960378842
Random Forest Error Rate :0.4776461039621158
Tesst set
Random Forests Mean Squared error :19369.758205601724
Random Forests r2_score :0.4915544808788034
Random Forest Error Rate :0.5084455191211966
```

Fig 2: Standard Random Forest Result

Table 1: Standard Random Forest Performance

<table>
<tr><th colspan="7">Random Forest Classification</th></tr>
<tr><th colspan="3">ROC</th><th>OOB Error Rate</th><th>AUC</th><th>Accuraacy</th><th>Error Rate</th></tr>
<tr><td>Class 0</td><td>Class 1</td><td>Class 2</td><td></td><td></td><td></td><td></td></tr>
<tr><td>0.99</td><td>0.99</td><td>0.89</td><td>0.91</td><td>80.07</td><td>93.18</td><td>0.07</td></tr>
<tr><th colspan="7">Random Forest Regression</th></tr>
<tr><th colspan="3">Mean Squared Error</th><th colspan="3">R^2 Score</th><th>Error Rate</th></tr>
<tr><td colspan="3"><b>19369.75</b></td><td colspan="3"><b>0.49</b></td><td><b>0.51</b></td></tr>
</table>

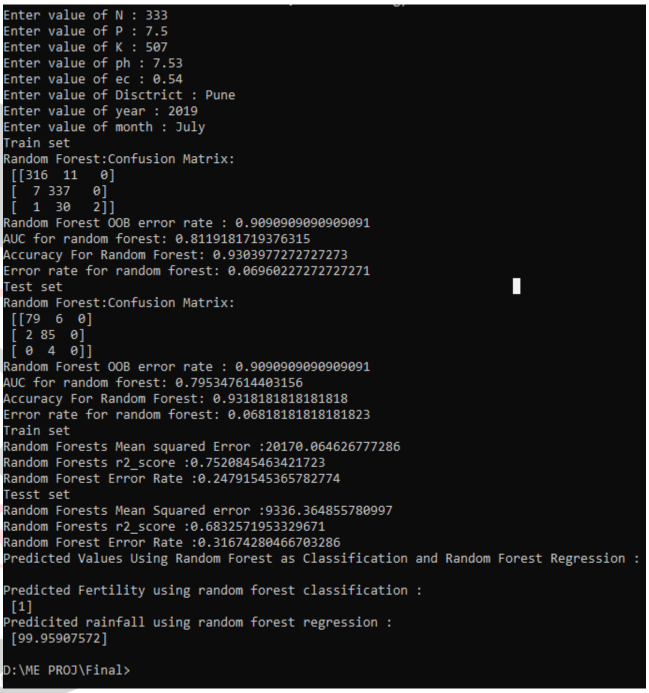
The predicted values also shown in above snap. Predicted crop will be displayed in another window as shown below.



Crop Names
Bean
Custard ap...
Dates
Aniseed
Black Pepper

Figure 3: Predicted Crop by Standard RF

4.3 *Result for Enhanced Random Forest:*



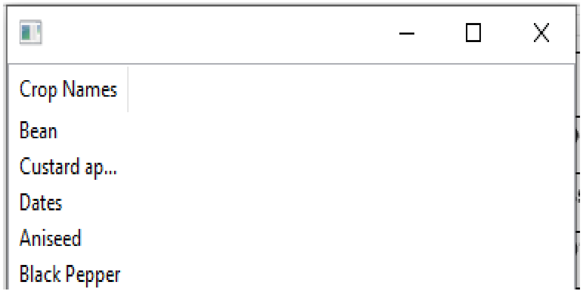
```
Enter value of N : 333
Enter value of P : 7.5
Enter value of K : 507
Enter value of ph : 7.53
Enter value of ec : 0.54
Enter value of Disctrict : Pune
Enter value of year : 2019
Enter value of month : July
Train set
Random Forest:Confusion Matrix:
 [[316  11   0]
 [  7 337   0]
 [  1  30   2]]
Random Forest OOB error rate : 0.9090909090909091
AUC for random forest: 0.8119181719376315
Accuracy For Random Forest: 0.9303977272727273
Error rate for random forest: 0.06960227272727271
Test set
Random Forest:Confusion Matrix:
 [[79  6  0]
 [ 2 85  0]
 [ 0  4  0]]
Random Forest OOB error rate : 0.9090909090909091
AUC for random forest: 0.795347614403156
Accuracy For Random Forest: 0.9318181818181818
Error rate for random forest: 0.06818181818181823
Train set
Random Forests Mean squared Error :20170.064626777286
Random Forests r2_score :0.7520845463421723
Random Forest Error Rate :0.24791545365782774
Tesst set
Random Forests Mean Squared error :9336.364855780997
Random Forests r2_score :0.6832571953329671
Random Forest Error Rate :0.31674280466703286
Predicted Values Using Random Forest as Classification and Random Forest Regression :

Predicted Fertility using random forest classification :
 [1]
Predicited rainfall using random forest regression :
 [99.95907572]

D:\ME PROJ\Final>
```

Figure 4: Performance Measurement for Enhanced RF

Table 2: Performance table for Enhanced Random Forest

<table>
<tr><th colspan="7">Random Forest Classification</th></tr>
<tr><th colspan="3">ROC</th><th>OOB Error Rate</th><th>AUC</th><th>Accuraacy</th><th>Error Rate</th></tr>
<tr><td>Class 0</td><td>Class 1</td><td>Class 2</td><td></td><td></td><td></td><td></td></tr>
<tr><td>1.00</td><td>0.99</td><td>0.98</td><td>0.90</td><td>79.54</td><td>93.18</td><td>0.07</td></tr>
<tr><th colspan="7">Random Forest Regression</th></tr>
<tr><th colspan="3">Mean Squared Error</th><th colspan="3">R^2 Score</th><th>Error Rate</th></tr>
<tr><td colspan="3"><b>9336.36</b></td><td colspan="3"><b>0.68</b></td><td><b>0.31</b></td></tr>
</table>

DOI : 10.35291/2454-9150.2019.0291



Crop Names
Bean
Custard ap...
Dates
Aniseed
Black Pepper

Figure 5: Predicted crop using Enhanced RF

4.4 Comparison:

Table 3: Comparison of RF and ID3 against Enhanced RF

	OOB Error rate			AUC			Accuracy			Error Rate			Mean Squared Error			R2 Score			Error Rate		
	Standard Random Forest	ID3	Enhanced Random Forest	Standard Random Forest	ID3	Enhanced Random Forest	Standard Random Forest	ID3	Enhanced Random Forest	Standard Random Forest	ID3	Enhanced Random Forest	Standard Random Forest	ID3	Enhanced Random Forest	Standard Random Forest	ID3	Enhanced Random Forest	Standard Random Forest	ID3	Enhanced Random Forest
Sample 1	0.91	0.91	0.9	0.79	0.72	0.81	90.3	89.3	93.2	0.09	0.08	0.07	19369.76	19475.35	9336.36	49.15	50.26	68.32	0.51	0.52	0.25
Sample 2	0.89	0.9	0.9	0.78	0.76	0.8	91.2	90.2	93.4	0.09	0.09	0.08	15647.23	20365.95	7645.56	50.12	48.46	70.42	0.48	0.65	0.21
Sample 3	0.92	0.9	0.89	0.9	0.8	1	91.7	91.5	92.3	0.08	0.09	0.08	17534.35	18356.24	7546.3	50.1	49.64	71.64	0.5	0.49	0.23
Sample 4	0.88	0.91	0.9	0.99	0.93	0.9	91.9	91.2	92.8	0.08	0.08	0.07	16547.24	17356.32	8343.16	48.16	48.65	72.48	0.52	0.49	0.23
Sample 5	0.92	0.92	0.88	0.95	0.94	0.97	90.7	89.5	92.2	0.09	0.09	0.08	13247.23	19648.65	9435.23	50.16	47.35	65.56	0.57	0.54	0.21
Sample 6	0.9	0.89	0.87	0.79	0.72	0.98	90.6	89.6	93.4	0.09	0.1	0.08	16326.23	20315.94	7653.64	50.75	49.85	78.56	0.47	0.51	0.22
Sample 7	0.87	0.9	0.88	1	0.94	0.9	91.2	88.7	92.1	0.1	0.09	0.08	17453.12	17546.65	8323.8	49.26	47.67	78.48	0.49	0.57	0.2
Sample 8	0.92	0.88	0.85	0.89	0.9	0.91	90.5	90.1	92.2	0.09	0.09	0.08	17354.32	18356.26	8542.35	46.82	51.65	68.45	0.54	0.51	0.24
Sample 9	0.9	0.91	0.87	0.88	0.81	0.97	89.3	90.2	92.4	0.08	0.1	0.07	18975.64	17365.32	9356.64	50.1	52.26	67.58	0.51	0.49	0.21
Sample 10	0.89	0.9	0.88	0.86	0.85	0.94	91.8	89.2	92.8	0.08	0.09	0.08	19346.64	18653.23	7564.65	51.46	48.65	72.65	0.48	0.52	0.23

V. Discussion

As compared in comparison table, Enhanced Random forest classification algorithm gives higher ROC Value, Less OOB Error Rate, higher AUC, Accuracy and less Error Rate against Random Forest Regression in comparison to ID3 algorithm as regression. Even ID3 didn't perform well neither against Random Forest nor Enhanced Random Forest algorithm. Also, in Comparison table, Enhanced Random Forest perform very well. It returns adequate lower Mean Squared Error compared to ID3 regression against Random Forest algorithm. Also, it returns with the higher R^2 Score and lower Error rate in compared to any other classification algorithm.

VI. Future Scope

The future of the Random Forest as classification and Regression involves predicting the pesticides and fertilisers to be used to improve fertility level of soil based on current micro and macro nutrient available in soil. Random Forest as classification and Regression is also helpful in predicting the rainfall for coming years based on previous rainfall trend.

VII. Conclusion

Earlier yield production was decided based on farmers experience where technology involvement was not there which gives accurate answer to decide the crop to plough. Therefore, in order to help farmers to decide the crop to plough for their financial as well as social benefits crop prediction system make use of Random Forest as classification as well as regression. Classification algorithm classifies the soil sample based on the available nutrient in soil into different class of soil where as regression predicts the expected rainfall for the entered year and month in which farmer want to plough.

Enhanced Random Forest classification and regression which performed in comparison. The classification comparison is based on the parameter ROC Curve, AUC, OOB Error Rate, Accuracy and Error Rate, where as Regression comparison is based on parameter Mean Squared error, R^2 Score and Error rate.

The planned model work presents comparison of Random forest Classification combined with Random Forest Regression and ID3 Regression and Enhanced Random Forest. Different soil sample have been used to compare the algorithm and it is concluded that Enhanced Random Forest as classification and Regression performed better in term of RUC Curve, AUC, Accuracy, Error Rate and OOB Error Rate. Accuracy is most important parameter that demonstrate the performance of any algorithm. It is observed that accuracy of Enhanced Random Forest as Classification and Regression combined is better in classifying the dataset and predicting the result.

DOI : 10.35291/2454-9150.2019.0291

REFERENCES

[1] Supriya D M "Analysis Of Soil Behavior And Prediction Of Crop Yield Using Data Mining Approach" In International Journal Of Innovative Research In Computer And Communication Engineering, Vol. 5, Issue 5, May 2017.

[2] Vaneesbeer Singh, Abid Sarwar "Analysis Of Soil And Prediction Of Crop Yield (Rice) Using Machine Learning Approach" In International Journal Of Advanced Research In Computer Science, Volume 8, No. 5, May – June 2017.

[3] Profile Of Maharashtra And Selected Districts, Shodhganga.Inflibnet.Ac.In/Bitstream/10603/121515/13/13_Chapter4.Pdf.

[4] Andrew.W "Moore Professor School Of Computer Science Carnegie Mellon University", Naïve Bayes Classifiers, Www.Cs.Cmu.Edu/~Awm Awm@Cs.Cmu.Edu.

[5] Andrew.W "Moore Professor School Of Computer Science Carnegie Mellon University", Naïve Bayes Classifiers, Www.Cs.Cmu.Edu/~Awm Awm@Cs.Cmu.Edu.

[6] B. Bhattacharya, D.P. Solomatine "Machine Learning In Soil Classification" Elsevier 2006 Special Issue, Neural Networks 19 (2006) 186–195.

[7] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio "Generative Adversarial Nets"

[8] Kriegel, Hans-Peter; Schubert, Erich; Zimek, Arthur (2016). "The (Black) Art Of Runtime Evaluation: Are We Comparing Algorithms Or Implementations?". Knowledge And Information Systems. 52: 341–378. Doi:10.1007/S10115-016-1004-2. Issn 0219-1377

[9] Mackay, David (2003). "Chapter 20. An Example Inference Task: Clustering" (Pdf). Information Theory, Inference And Learning Algorithms. Cambridge University Press. Pp. 284&Ndash, 292. Isbn 0-521-64298-1. Mr 2012999.

[10] Coates, Adam; Ng, Andrew Y. (2012). "Learning Feature Representations With K-Means" (Pdf). In G. Montavon, G. B. Orr, K.-R. Müller. Neural Networks: Tricks Of The Trade. Springer.

[11] Csurka, Gabriella; Dance, Christopher C.; Fan, Lixin; Willamowski, Jutta; Bray, Cédric (2004). Visual Categorization With Bags Of Keypoints (Pdf). Eccv Workshop On Statistical Learning In Computer Vision.

[12] Coates, Adam; Lee, Honglak; Ng, Andrew Y. (2011). An Analysis Of Single-Layer Networks In Unsupervised Feature Learning (Pdf). International Conference On Artificial Intelligence And Statistics (Aistats). Archived From The Original (Pdf) On 2013-05-10.

[13] Schwenker, Friedhelm; Kestler, Hans A.; Palm, Günther (2001). "Three Learning Phases For Radial-Basis-Function Networks". Neural Networks. 14 (4–5): 439–
458. Citeseerx 10.1.1.109.312. Doi:10.1016/S0893-6080(01)00027-2

[14] Geetha Mcs. Implementation Of Association Rule Mining For Different Soil Types In Agriculture. International Journal Of Advanced Research In Computer And Communication Engineering. 2015 Apr; 4(4):520–2.

[15] Knowledge Discovery And Data Mining To Identify Agricultural Patterns, Kulwant Kaur, Maninderpal Singh, Ijesrt [1337-1345], March, 2014

[16] G.Kesavaraj, Dr.S.Sukumaran "A Study On Classification Techniques In Data Mining" Ieee – 31661.

DOI : 10.35291/2454-9150.2019.0291