International Journal of Pure and Applied Mathematics
Volume 118 No. 20 2018, 255-262
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

WEATHER PREDICTION ANALYSIS USING RANDOM FOREST ALGORITHM

[1]S. Karthick, [2]D. Malathi, [3]C.Arun
[1,3]Department of Software Engineering, SRM University, Kattankulathur,
Chennai - 603203, Tamil Nadu, India
[2]Department of Computer Science and Engineering, SRM University,
Kattankulathur, Chennai - 603203, Tamil Nadu, India
[1]karthik.sa@srmuniv.ac.in, [2]malathi.d@ktr.srmuniv.ac.in
[3]arun.c@ktr.srmuniv.ac.in

Abstract: One of the greatest challenge that meteorological department faces are to predict weather accurately. These predictions are important because they influence daily life and also affect the economy of a state or even a nation. Weather predictions are also necessary since they form the first level of preparation against the natural disasters which may make difference between life and death. They also help to reduce the loss of resources and minimizing the mitigation steps that are expected to be taken after a natural disaster occurs. This research work focuses on analyzing algorithms that are suitable for weather prediction and highlights the performance analysis of C4.5 with Random Forest algorithms. After a comparison between the data mining algorithms and corresponding ensemble technique used to boost the performance, a classifier is obtained that will be further used to predict the weather.

Keywords: Data mining, Decision Tree, Ensemble Technique, Pre-processing, Weather prediction.

1. Introduction

Weather prediction has been a difficult task in the meteorological department since years due to varied reason such as the drastically unpredictable behavior of climate. Even with the latest advancement in the technology, it's been difficult, even if the prediction can be made the accuracy in prediction of weather is a questionable factor. Even in current date, this domain remains as a research topic in which scientists and mathematicians are working to produce a model or algorithm that will accurately predict the weather. There have been immense improvements in the sensors that are responsible for recording the data from the environment and cancel the noise present in them; along with this new models have been proposed which include different attributes related for making the accurate prediction.

Currently one of the most widely used techniques for weather prediction is mining information represent the weather. Data mining offers a way of analyzing the data statistically and extract or derive such rules that can be used for predictions. Data mining is exploration technique has been widely used in major fields where the prediction of future is highly desired such as prediction of prices of stocks over a period of time in future. Scientists have now realized that data mining can be used as a tool for weather prediction as well.
The basic entity requires to mine is the meteorological data gathered over a definite period of time. Data might be a piece of information which provides details of the climatic condition such moisture, humidity, etc. The data collection has been happened in various places using the varied types of devices and gets stored in some form which could be rendered to extract the prediction results. Weather prediction is usually done using the data gathered by remote sensing satellites. Various weather parameters like temperature, rainfall, and cloud conditions are projected using image taken by meteorological satellites to access future trends

The Data mining offers a variety of technique to predict the results such Classification, Clustering, Regression, etc. Data Classification is the process of unifying the data into different categories or groups based on the relations among them The term data mining refers to the techniques that are used to extract the required information from the given set of data that might be useful for statistical purpose or making predictions by learning patterns in data and correlation between different parameters. The accuracy of the prediction widely depends on knowledge of prevailing weather condition over a wide area.

2. Related Work

Weather is one of the most influential factors that decide the life activity of the members of the ecosystem, which influence the economic factor of country and people of the region. To eliminate the disaster due to weather, an activity has to be activated to predict the weather uncertainty behavior in future. Usually two main techniques are used for weather

forecasting, one involves usage of large amount of data collected over a period of time and analyze the data to gain knowledge about future weather and the other involves construction of equations that will help predict weather by identifying different parameters and substituting the values to obtain the desired resulti.e by means of constructing the regression techniques. Decades of research work has been done in the field of meteorology. Recently researchers have started highlighting the effectiveness of data mining algorithms in predicting the weather. One of the latest research works includes a paper [1] by MsAshwiniMandale, MrsJadhawar B.A. this paper makes a mention of Artificial Neural Networks and Decision Tree algorithms and their performance in prediction of weather. ANN finds a relationship between the weather attributes and builds a complex model, whereas C5 decision tree learns the trend of data and accordingly builds a classifier tree that can be used for prediction. Another well-known data mining technique, the CART was used by Elia G. Petre in her paper [2]. A decision tree was produced as an output and its performance was calculated using evaluation metrics which included parameters like precision, accuracy, FP rate, TP rate, F-measure, and ROC Area.

Since numerous data mining algorithms are available for use, it is necessary to find the appropriate technique that will be suitable for the domain it is being applied to. In certain cases, regression technique proves to be more effective whereas in other cases, rule-based technique and decision tree algorithms give an accurate result with a low computational cost. In [3], Divya Chauhan and Jawahar Thakur have reviewed various data mining techniques and gave a performance comparison between algorithms like C4.5, CART, k-means clustering, ANN, and MLR when they were used for weather prediction. They made a conclusion that k-means and decision tree algorithms perform better than other algorithms in case of weather prediction. In [5], [6], [7], and [12] in-depth performance comparison has been between C4.5 and Random Forest algorithm, includes discussion over the suitability of algorithm when applied to the different dataset.

3. Approach

The methodology used in this paper consists of certain steps that are usually used in data mining applications the steps are as follows: [8] [9]

Data Collection & Retrieval- Data used for the research work was obtained from a meteorological tower of SRM University Chennai, India. The format of data was in CSV format and included parameters like humidity, temperature, cloud cover, wind speed, etc. [4]

Data Transformation- The CSV file was first converted to the .arff format to feed it into the Data **Mining tool –** WEKA. The conversion to .arff format was implemented through code written in Java. Two separate files were maintained as weather.arff and predict weather.arff in which weather.arff consists of the actual data collected over a period of 2 years and the predicted weather .arff file contained sample data used for prediction. [4]

Data Pre-processing- The weather.arff file was used as a source file and then Resample technique was applied to the data present in it. Resample technique involves choosing instances from the dataset on a random basis with or without replacement.

Feature Extraction- Among all the parameters considered which consisted of max temperature, "min" temperature, mean temperature, max humidity, min humidity, mean humidity, wind speed, cloud cover, and rainfall. These parameters were used for further processing in the application as they were mutually exclusive and no redundancy was present between them. [1]

Data mining- This stage consisted of analyzing the given dataset with different algorithms like Random Forest and C4.5 (J48) algorithm and then choosing the better one for further predictions. Then the dataset was split into the training set for making the machine learn and the testing dataset along with cross-validation. Then the patterns were recorded to make further predictions. Additionally few ensemble algorithms like boosting and bagging were applied to improve the results. [4]

Table 1. Attributes of Weather.arff [1] [4] [9]

Min Temperature	Numeric
Mean Temperature	Numeric
Max Temperature	Numeric
Min Humidity	Numeric
Mean Humidity	Numeric
Max Humidity	Numeric
Wind Speed	Numeric
Cloud Cover	Numeric
Rainfall	Boolean

4. Comparison between Random Forest and C4.5 Decision Tree Algorithm

4.1 Random Forest

Random Forest is believed to be one of the best ensemble classifiers for high-dimensional data. Random forests are a mixture of tree predictors such that each tree depends on the values of a random vector sampled autonomously and with the same distribution for all trees in the forest. The generalization error for forests converges to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the

individual trees in the forest and the association between them. A different subset of the training data is selected, with replacement, to train each tree. Remaining training data are used to estimate error and variable importance. Class assignment is made by the number of votes from all of the trees and for regression, the average of the results is used. [12] [13]

Algorithm:

Each tree is constructed using the following algorithm:
1. Let the number of training cases be N, and the number of variables in the classifier is M.
2. We are told the number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M.
3. Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.
4. For each node of the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set.
5. Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier).
For prediction, a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is iterated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction. [12] [13]

4.2 C4.5 Decision Tree

Unlike Random Forest, the C4.5 is a classification algorithm used to generate a decision tree for the given dataset. It is based on the information entropy concept. Construction of the decision tree is done by selecting the best possible attribute that will be able to split set of samples in most effective manner. The attribute having the highest entropy difference or normalized information gain is selected as the splitting criteria for that particular node. Similar fashion is followed and nodes are added to the decision tree. Each penultimate node carries the last attribute or multiple attributes for making the final decision of the problem. [6] [7]

Algorithm:

INPUT
D // Training data
OUTPUT
T // Decision tree
DTBUILD (*D)
T = Null;
T = Create root node and label with splitting attribute;
T = Add arc to root node for each split predicate and label;
For each arc do
D = Database created by applying splitting predicate to D;
If stopping point reached for this path, then
T'= Create leaf node and label with appropriate class;
Else
T' = DTBUILD (D);
T = Add T' to arc; [5] [6] [7]

Correctly Classified Instances	516	82.428%
Incorrectly Classified Instances	110	17.572%

Kappa statistic	0.6478
Mean absolute error	0.2022
Root mean squared error	0.3929
Relative absolute error	40.5318%
Root relative squared error	78.6639%
Total Number of Instances	626

Detailed Accuracy By Class:

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.832	0.185	0.832	0.832	0.832	0.863	yes
	0.815	0.168	0.815	0.815	0.815	0.863	no
Weighted Avg.	0.824	0.177	0.824	0.824	0.824	0.863	

Confusion Matrix

a	b	Classified as
273	55	a=yes
55	243	b=no

Figure 1. Evaluation of training set for C4.5 (J48) algorithm. [2] [10] [14]

Correctly Classified Instances	545	87.061%
Incorrectly Classified Instances	81	12.939%

Kappa statistic	0.7405
Mean absolute error	0.1858
Root mean squared error	0.3131
Relative absolute error	37.2431%
Root relative squared error	62.6905%

Total Number of Instances				626			
Detailed Accuracy By Class:							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.881	0.141	0.873	0.881	0.877	0.931	yes
	0.859	0.119	0.868	0.859	0.863	0.929	no
Weighted Avg.	0.871	0.13	0.871	0.871	0.871	0.93	
Confusion Matrix							
a	b	Classified as					
289	39	a=yes					
42	256	b=no					

Figure 2. Evaluation of training set for Random Forest algorithm. [2] [10] [14]

After the performance comparison, the Random Forest algorithm proved to be better than J48 algorithm providing better results. Both algorithms were further used studying the legacy data concerning weather. The resample filter was used for data pre-processing, furthermore for the data selection step, from a total of 9 parameters. This resulted in consideration of linear parameters like max temperature, min temperature, mean temperature, max humidity, min humidity, mean humidity, wind speed, cloud cover, and rainfall.[1] [2]
The next step in the implementation was to apply the decision tree algorithm to the dataset. The filtered data was given as an input to the algorithm and a decision tree was expected as an output to the J48 algorithm, while Random Forest by definition generated a forest including several smaller decision tree units. Each decision tree will contain a collection of nodes and each node consists of attributes or set of attributes as criteria to split the node for further classification. Since the tool does not give the facility to visualize the forest generated by the algorithm, the resultant consisted of the Random forest of 10 trees, each constructed while considering 4 random features with an out of the bag error of 0.1597. [1] [2] [11]

Table 2. Performance Comparison [1] [2] [3] [4]

Parameters	**Random Forest**	**C4.5 Decision Tree**
Correctly Classified Instances	545	516
Incorrectly Classified Instances	81	110
Kappa Statistic	0.7405	0.6478
Mean Absolute Error	0.1858	0.2022
Root Mean Squared Error	0.3131	0.3923
Relative Absolute Error	37.2431%	40.5318%
Root Relative Squared Error	62.6905%	78.6639%
F-Measure	0.871	0.824
Precision	0.871	0.824
Time is taken to Build Model	0.11	0.11

5. Resultant C4.5 Decision Tree

```
MinTemp<= 24
|  CloudCover<= 4
|  |  WindSpeed<= 17
|  |  |  WindSpeed<= 11: yes (10.0)
|  |  |  WindSpeed> 11
|  |  |  |  MinTemp<= 20
|  |  |  |  |  MeanTemp<= 24
|  |  |  |  |  |  MeanHumidity<= 66
|  |  |  |  |  |  |  MeanTemp<= 23
|  |  |  |  |  |  |  |  MinTemp<= 14: yes (2.0)
|  |  |  |  |  |  |  |  MinTemp> 14: no (2.0)
|  |  |  |  |  |  |  MeanTemp> 23: yes (25.32)
|  |  |  |  |  |  MeanHumidity> 66
|  |  |  |  |  |  |  MinTemp<= 19: no (15.0/1.0)
|  |  |  |  |  |  |  MinTemp> 19
|  |  |  |  |  |  |  |  MeanHumidity<= 82: yes (11.0/3.0)
|  |  |  |  |  |  |  |  MeanHumidity> 82: no (4.0)
|  |  |  |  |  MeanTemp> 24: no (13.0)
|  |  |  |  MinTemp> 20
|  |  |  |  |  MinTemp<= 23: yes (36.77/3.0)
|  |  |  |  |  MinTemp> 23: no (2.0)
|  |  WindSpeed> 17
|  |  |  CloudCover<= 2
|  |  |  |  MaxHumidity<= 72: yes (8.85)
|  |  |  |  MaxHumidity> 72
|  |  |  |  |  MaxTemp<= 35
|  |  |  |  |  |  CloudCover<= 1
|  |  |  |  |  |  |  WindSpeed<= 19: no (14.0)
|  |  |  |  |  |  |  WindSpeed> 19
|  |  |  |  |  |  |  |  MinHumidity<= 31
|  |  |  |  |  |  |  |  |  MinHumidity<= 18: no
(2.18/0.18)
|  |  |  |  |  |  |  |  |  MinHumidity> 18
|  |  |  |  |  |  |  |  |  |  WindSpeed<= 24: yes (4.0)
|  |  |  |  |  |  |  |  |  |  WindSpeed> 24: no (2.0)
|  |  |  |  |  |  |  |  MinHumidity> 31: no (10.0)
|  |  |  |  |  |  CloudCover> 1
|  |  |  |  |  |  |  MeanHumidity<= 57
|  |  |  |  |  |  |  |  WindSpeed<= 23: no (19.38/0.38)
|  |  |  |  |  |  |  |  WindSpeed> 23: yes (2.0)
```

```
| | | | | | | MeanHumidity> 57
| | | | | | | | MeanHumidity<= 62: yes (13.0/1.0)
| | | | | | | | MeanHumidity> 62
| | | | | | | | | WindSpeed<= 20: yes (4.0)
| | | | | | | | | WindSpeed> 20: no (30.0/5.0)
| | | | | MaxTemp> 35: yes (5.98/0.85)
| | | CloudCover> 2
| | | | MaxHumidity<= 94
| | | | | MeanTemp<= 23
| | | | | | MaxTemp<= 26: no (3.0)
| | | | | | MaxTemp> 26: yes (13.96)
| | | | | MeanTemp> 23
| | | | | | MaxTemp<= 28
| | | | | | | MeanHumidity<= 72: yes (2.0)
| | | | | | | MeanHumidity> 72: no (15.0)
| | | | | | MaxTemp> 28
| | | | | | | WindSpeed<= 40
| | | | | | | | MinTemp<= 21
| | | | | | | | | WindSpeed<= 27
| | | | | | | | | | WindSpeed<= 20: yes (2.0)
| | | | | | | | | | WindSpeed> 20: no
(14.43/2.45)
| | | | | | | | | WindSpeed> 27: yes (7.0)
| | | | | | | | MinTemp> 21
| | | | | | | | | WindSpeed<= 34
| | | | | | | | | | MinTemp<= 22: no (11.49/1.0)
| | | | | | | | | | MinTemp> 22
| | | | | | | | | | | WindSpeed<= 29
| | | | | | | | | | | | WindSpeed<= 26
| | | | | | | | | | | | | WindSpeed<= 24
| | | | | | | | | | | | | | MeanHumidity<= 61:
no (6.0)
| | | | | | | | | | | | | | MeanHumidity> 61
| | | | | | | | | | | | | | | MaxTemp<= 31: no
(2.0)
| | | | | | | | | | | | | | | MaxTemp> 31: yes
(2.0)
| | | | | | | | | | | | | | WindSpeed> 24: yes
(4.0)
| | | | | | | | | | | | | WindSpeed> 26: no (4.0)
| | | | | | | | | | | | WindSpeed> 29: yes (2.0)
| | | | | | | | | | | WindSpeed> 34: no (9.0)
| | | | | | | | | WindSpeed> 40
| | | | | | | | | MeanHumidity<= 69: no (2.0)
| | | | | | | | | MeanHumidity> 69: yes (6.0)
| | | | | MaxHumidity> 94
| | | | | | WindSpeed<= 26
| | | | | | | MaxTemp<= 34
| | | | | | | | CloudCover<= 3
| | | | | | | | | MinTemp<= 20
| | | | | | | | | | MinHumidity<= 34
| | | | | | | | | | | WindSpeed<= 20: yes (10.0)
| | | | | | | | | | | WindSpeed> 20: no (4.0/1.0)
| | | | | | | | | | MinHumidity> 34: no (13.0)
| | | | | | | | | MinTemp> 20: yes (9.0/1.0)
| | | | | | | | CloudCover> 3
| | | | | | | | | WindSpeed<= 24: no (15.0/1.0)
| | | | | | | | | WindSpeed> 24: yes (5.0/1.0)
| | | | | | | MaxTemp> 34: yes (3.0)
| | | | | | WindSpeed> 26
| | | | | | | MeanHumidity<= 78: yes (27.0)
| | | | | | | MeanHumidity> 78
| | | | | | | | WindSpeed<= 33: yes (4.0)
| | | | | | | | WindSpeed> 33: no (4.0/1.0)
| CloudCover> 4
| | MaxTemp<= 26
| | | MinTemp<= 19: no (7.0)
| | | MinTemp> 19
| | | | MinTemp<= 20
| | | | | CloudCover<= 5: yes (3.0)
| | | | | CloudCover> 5
| | | | | | MaxTemp<= 24
| | | | | | | MinHumidity<= 78: yes (5.0)
| | | | | | | MinHumidity> 78
| | | | | | | | MeanHumidity<= 96: no (2.0)
| | | | | | | | MeanHumidity> 96: yes (4.0)
| | | | | | MaxTemp> 24
| | | | | | | MaxHumidity<= 94: yes (2.0)
| | | | | | | MaxHumidity> 94: no (8.0)
| | | | MinTemp> 20: no (4.0)
| | MaxTemp> 26
| | | CloudCover<= 5
| | | | MeanHumidity<= 83
| | | | | WindSpeed<= 23: yes (19.59)
| | | | | WindSpeed> 23
| | | | | | MeanTemp<= 23: yes (8.0)
| | | | | | MeanTemp> 23
| | | | | | | MaxTemp<= 28: no (3.0)
| | | | | | | MaxTemp> 28
| | | | | | | | MinHumidity<= 47
| | | | | | | | | MeanTemp<= 27: yes (2.48)
| | | | | | | | | MeanTemp> 27: no (3.48)
| | | | | | | | MinHumidity> 47: yes (12.0/1.0)
| | | | MeanHumidity> 83
| | | | | MinHumidity<= 59
| | | | | | WindSpeed<= 24: no (9.0)
| | | | | | WindSpeed> 24
| | | | | | | MaxTemp<= 27: yes (2.0)
| | | | | | | MaxTemp> 27: no (2.0)
| | | | | MinHumidity> 59: yes (4.0)
| | | CloudCover> 5
| | | | WindSpeed<= 33: yes (23.09/0.2)
| | | | WindSpeed> 33
| | | | | MaxTemp<= 30: no (2.0)
| | | | | MaxTemp> 30: yes (3.0)
MinTemp> 24
| WindSpeed<= 29
| | MaxHumidity<= 94
| | | MaxHumidity<= 84: yes (2.0)
| | | MaxHumidity> 84
| | | | MinHumidity<= 35: no (9.0)
| | | | MinHumidity> 35
| | | | | CloudCover<= 3: yes (6.0/1.0)
| | | | | CloudCover> 3: no (5.0)
| | MaxHumidity> 94: yes (8.0/1.0)
| WindSpeed> 29
```

| | CloudCover<= 3: no (20.0)
| | CloudCover> 3
| | | MeanTemp<= 28: no (7.0)
| | | MeanTemp> 28: yes (4.0)

Number of Leaves: 78
Size of the tree: 155
Time is taken to build model: 0.11 seconds. [1] [4] [14]

6. Conclusion

The comparative analysis was made between C4.5 Decision Tree (J48) and Random Forest algorithm with dataset comprising of nine weather parameters collected over a period of two years. Though the result of both the algorithms was found to be relatively good as they fall in the category of recommended algorithms for classification and weather prediction problems yet Random Forest proved to be better than the C4.5 Decision Tree. Where C4.5 achieved an accuracy of 82.4%, Random Forest was able to secure 87.1%accuracy proving it to be better. The accuracy was obtained using 10 fold cross validation keeping the overfitting problem in mind. This result is predictable since the concept on which Random Forest is based upon helps it to generate multiple trees by selecting parameters on random basis produce a highly accurate classifier.

Random Forest is highly recommended over other decision trees when the size of the dataset or the number of parameters in the dataset is high. Additionally, it is capable of balancing error in class population unbalanced data sets. The only disadvantage of the classifier is that it can overfit the data. The confusion matrix also supported the above-made statement of Random Forest being a better performer in case of weather dataset. The number of instances that were true positives, i.e. true instances and also was predicted true by Random Forest was higher than that of C4.5 Decision Tree and in case of number instances that were true negatives, i.e. false and were predicted as false showed a similar result. Even the precision of Random Forest was slightly yet higher in this case.
Since the current result depicts that the Random Forest is better than the J48 algorithm, it will even perform better when the amount of data is increased either in terms of instances or parameters. Therefore it can be said that the performance of Random Forest algorithm was is than that of Naïve Bayes in case of dataset dealing with weather. Further improvements can be made to improve the result of the algorithm by applying appropriate filter to the dataset in pre-processing stage.s

References

[1] MsAshwiniMandale, MrsJadhawar B.A., Weather forecast prediction: a Data Mining application, International Journal of Engineering Research and General Science Volume 3, Issue 2, March-April, 2015 ISSN 2091-2730

[2] Elia Georgiana Petre, A Decision Tree for Weather Prediction, BULETINUL UniversităŃii Petrol–Gaze din Ploiesti, Vol. LXI No. 1/2009, April 2009.

[3] Divya Chauhan, Jawahar Thakur, Data Mining Techniques for Weather Prediction: A Review, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169-2184 – 2189, August 2014.

[4] FolorunshoOlaiya, Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, I.J. Information Engineering and Electronic Business, 2012, 1, 51-59, DOI: 10.5815/ijieeb, July 2012.

[5] Sunita Joshi, Bhuwaneshwari Pandey, Nitin Joshi, Comparative analysis of Naive Bayes and J48 Classification, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 12, ISSN: 2277 128X, December 2015.

[6] Mrs. Tina R. Patil, Mrs. S.S. Sherekar, Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification, International Journal of Computer Science and Applications, Vol. 6, No.2, April 2013.

[7] AnshulGoyal, Rajni Mehta, Performance Comparison of Naïve Bayes and J48 Classification Algorithms, International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11, 2012.

[8] Meghali A. Kalyankar, Prof. S. J. Alaspurkar, Data Mining Technique to Analyse the Metrological Data, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 2, ISSN: 2277 128X, February 2013.

[9] Amruta A. Taksande, P. S. Mohod, Applications of Data Mining in Weather Forecasting Using Frequent Pattern Growth Algorithm, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, Volume 4 Issue 6, June 2015.

[10] Pinky Saikia Dutta, Hitesh Tahbilder, Prediction Of Rainfall Using Data Mining Technique Over Assam, Indian Journal of Computer Science and Engineering (IJCSE), ISSN: 0976-5166 Vol. 5 No.2 Apr-May 2014

[11] Ankita Joshi, BhagyashriKamble, Vaibhavi Joshi, KomalKajale, NutanDhange, Weather Forecasting and Climate Changing Using Data Mining Application, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, March 2015, ISSN (Online) 2278-1021.

[12] Lakshmi Devasena C, Comparative Analysis of Random Forest, REP Tree and J48 Classifier for Credit Risk Prediction, International Journal of Computer Applications (0975 – 8887).

[13] Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood, Random Forests and Decision Trees, International Journal of Computer Science Issues(IJCSI), Vol. 9, Issue 5, No 3, September 2012, ISSN (Online) 1694-0814.

[14] Weka Software Documentation, http://www.cs. waikato.ac.nz/ml/weka/, accessed on 23 February 2016.

[15] Rajesh, M., and J. M. Gnanasekar. "An optimized congestion control and error management system for OCCEM." International Journal of Advanced Research in IT and Engineering 4.4 (2015): 1-10.

[16] S.V.Manikanthan and K.Baskaran “Low Cost VLSI Design Implementation of Sorting Network for ACSFD in Wireless Sensor Network”, CiiT International Journal of Programmable Device Circuits and Systems,Print: ISSN 0974 – 973X & Online: ISSN 0974 – 9624, Issue :November 2011, PDCS112011008.

[17] T.Padmapriya, Ms. N. Dhivya, Ms U. Udhayamathi, “Minimizing Communication Cost In Wireless Sensor Networks To Avoid Packet Retransmission”, International Innovative Research Journal of Engineering and Technology, Vol. 2, Special Issue, pp. 38-42.

[18] Meka Bharadwaj, Hari Kishore "Enhanced Launch-Off-Capture Testing Using BIST Designs” Journal of Engineering and Applied Sciences, ISSN No: 1816-949X, Vol No.12, Issue No.3, page: 636-643, April 2017.