

# Soft Computing Research For Weather Prediction Using Multilayer Architecture

S. Sheik Mohideen Shah, S. Meganathan, A. Kamali



**ABSTRACT---** *Rainfall prediction is helpful for the agriculture sector. Early prediction of drought and torrent situations is achieved through time series data. For the precise prediction, Artificial Neural Network(ANN) technique is used. The rainy dataset is tested using Feed Forward Neural Network(FFNN). The performance of this model is evaluated using Mean Square Error(MSE) and Magnitude of Relative Error(MRE). Better performance achieved when compared with other data mining techniques.*

**Keywords—** *Artificial Neural Network, Multi-Layer Perceptron, Time-Series Data*

## I. INTRODUCTION

Water is salient for all life on the grange. Farmers obtain water from rain and it is focal for farmers to farm. In India, most of the people depending on farming they rely on the season for their crop, and actual rainfall. Most of the farmers in India agonized from the torrent and drought. In these circumstances, we need precise rainfall prediction model to avoid this kind of complications. Weather forecasting is the process of recognizing and predicting a determined accuracy of climatic conditions using multiple technologies[1]. Forecasting will be helpful to take requisite measures to intercept damage of life. Many of the live systems rely on weather conditions to make necessary adjustments in their systems. This kind of model is helpful to protect our environment and save lives of people to the arrangement of adequate facilities[2].

In Data mining, a lot of techniques used to extract unknown information from the large database. It is used to find patterns from the enormous dataset and it is useful for predictions. It is the interactive process of starting at requirement gathering, data exploration, data preparations, modeling, evaluation and ends with deployment process.

It is an easy technique to extract useful information from a large amount of database. It differs from statistical methods, in order to search for interesting and valuable information. In general data mining contain two types of tasks: descriptive and predictive. Both data mining tasks perform their corresponding tasks accurately[3]. The architecture of this contains 4 types of coupling namely no - coupling, loose, semi-tight, and tight coupling. These techniques are very powerful and efficient. The commonly used techniques are Classification, Clustering, Regression, and ANN. These techniques contain corresponding methods to analyze the set of data. Time series data is the important aspects of modern research in the domain of knowledge discovery[4]. Spatial, temporal and spatio-temporal data is censorious for climate change and climate effect on the environment [5]. Time-series data is part of temporal data and it is a tremendously dimensional, bulky in volume and needs meticulous updating[6]. Time series data values are recorded based on the regular interval of the time period. Time series data perform their operation based on pattern identification[7][8]. This type of data is mostly used in prediction problems based on previously obtained values. Data mining techniques perform well in order to process this type of data.

The rainfall prediction department has to prolong upcoming rainfall data for giving alert to people.so a time series data have been maintained. The former researchers proposed a different type of prediction techniques using various algorithms such as k-means clustering, artificial neural network, etc. Neural Network is the easiest technique and it has the advantage of adaptive learning. It can be easily comprehensible compared statistical methods because this is a very easy technique for the prophecy problems[22]. Therefore the most accurate method is the appraisal using an artificial neural network in which Multi-Layer Perceptron(MLP) is used for the purpose of reckoning the effectiveness. A single layer is used only for simple problems but for very complex problems multi-layer perceptron is used. Multi-layer perceptron is really important for real-life problems[9].In ANN the bias and activation functions are used for the purpose of getting a result. ZeroR classifier is the baseline algorithm with the combination of MLP the rainfall rate is predicted and then the rainfall rate is classified as three classes. It is also a real-time problem for this reason MLP is used to provide an accurate result.The predicted result is classified as excess, deficit and normal classes based on the predicted rainfall rate.

Revised Manuscript Received on August 30, 2019.

\* Correspondence Author

**S. Sheik Mohideen Shah\***, Dept. Of Computer Science & Engineering, SRC, Sastra Deemed University, Thanjavur, Tamil Nadu, India.  
(Email: sheikme2005@src.sastra.edu)

**S. Meganathan**, Dept. Of Computer Science & Engineering, SRC, Sastra Deemed University, Thanjavur, Tamil Nadu, India.  
(Email: meganathan@src.sastra.edu)

**A. Kamali**, Dept. Of Computer Science & Engineering, SRC, Sastra Deemed University, Thanjavur, Tamil Nadu, India.  
(Email: kamalisc2017@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The performance is measured based on Mean Square Error and accuracy, recall, precision and F-Measure values[10]. These are all the metrics to measure the performance. The confusion matrix is also used to visualize the performance of the algorithm.

## II. LITERATURE REVIEW

There are various data mining techniques used for rainfall rate prediction and processing the time series dataset. M.Kannan, S.Prabhakaran, P.Ramachandran proposed empirical method technique that belongs to regression analysis which tries to predict short term rainfall for a specified region. Finally, the predicted values are not an accurate result but it shows an approximate result[11]. From which neural network is the best technique for forecasting problems. Most of the neural network techniques focused mainly on weekly, monthly and yearly rainfall prediction[12]. Hiren Kumar Deva Sarma used Association Rule Mining to analyze the time series data, this will not provide an accurate result when the dataset is small [8] because it finds the result based on frequent itemset. If the dataset is small this provides an inappropriate result. Mary N. Ahuna used an ANN technique to forecast the short span(30 seconds) of time when given three adjacent historical rain rates. The technique used in this method is pattern recall that considers historical rainfall rate patterns over Durban. For this model verification, RMSE method is used[13]. Mr. Chetan C. Janbandhu, Prof. Praful D. Meshram, Prof. Madhuri N. Gedam use a Bayesian approach to predict the rainfall and accuracy by comparing the different size of the training dataset. This model provides a good result when a large dataset is used[14]. It takes a lot of time when the large dataset has been providing but it will give an accurate result. Mislan, Haviluddin, Sigit Hardwinarto, Sumaryono, Marlon Aipassa use BPNN to test rainfall data by using three different epochs (500,1000,1500). The final result showed that 1000 epochs giving minimum error rate when compared to 500 and 1500 epochs[15]. Shabib Aftab, Munir Ahmad, Noureen Hameed, Muhammad Salman Bashir, Iftikhar Ali, Zahid Nawaz proposed a systematic literature survey for various rainfall prediction techniques[4]. Sankhadeep Chatterjee, Bimal Datta, Soumya Sen, Nilanjan Dey use Hybrid Neural Network to predict the rainfall with the accuracy of 89.54%[10]. They also compare the precision, recall, and F-Measure. Joko Azhari Suyatno, Fhira Nhita, Aniq Atiq Rohmawati use C4.5 algorithm to forecast rainfall. It is a decision tree algorithm that classifies the rainfall rate based on no, low, medium, heavy, and very heavy rainfall. This provides the accuracy result of 93.33%[16]. Afolayan Abimbola Helen, Ojokoh Bolanle A., Falaki Samuel O.their research is establishing the best model to predict the rainfall with a minimum error rate. The two models are ANN and Fuzzy logic models. From which the neural network is the best model than the fuzzy logic[17]. When compared to these proposed system used to found 1 year ahead of rainfall rate and artificial neural network perform well and provide a 96.63% of accurate result with minimum error rate by using 500 epochs in the neural network. This is also processed with a large dataset and it takes less time to predict the result. Former researchers use Generalized regression,

Fuzzy inference system, Pearson coefficient technique, classification, clustering, etc when compared to these techniques ANN with MLP provides an accurate result.

## III. PROPOSED SYSTEM

### A. Artificial Neural Network

It is a reckoning model based on the anatomy of the human intelligence. The Artificial Neural Network formation will be pretended by information because neural network changes or grasp new knowledge, based on the flowing of information[18]. It can handle a tremendous portion of the dataset and it has the aptness to tacitly identify complex non-linear relationships between contingent and individualistic attributes. It can be easily comprehensible compared to mathematical methods because this is the simplified technique for the prophecy problems. The output rate depends upon the training parameters and the testing dataset.

### B. Multi-layer Perceptron

In this paper, the FFNN is used to predict the rainfall in advance. Feed Forward Neural Network is also called Multilayer Perceptron(MLP)[13].In which the information flows in only forward, this type of network creates a multilayer network. The neurons in the layer perform independent computation in input and passthrough hidden layer to the output layer[3].

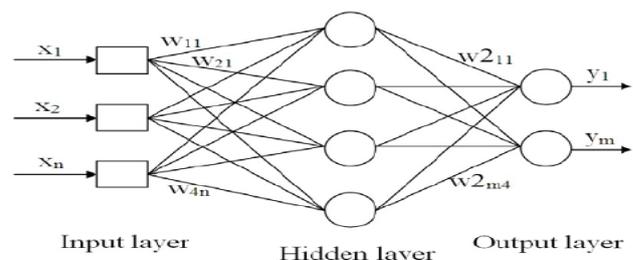


Fig. 1. Feed Forward Neural Network

### C. Hidden Layers

This layer has n number of nodes and it is calculated based on trial-and-error.

$$\text{Hidden layer} = (\text{number of attributes} + \text{number of classes}) / 2$$

### D. ZeroR classifier

ZeroR is the uncomplicated classification method which based on the target and disregard all predictors. ZeroR classifier predicts only the majority of the class. ZeroR, it is most applicable to regulating a margin performance as a paradigm for classification methods.

MLP used with the combination of ZeroR classifier to predict the one year ahead of data. The predicted rainfall data is further processed to find the annual rainfall is excess, deficit or normal rainfall.

For this, some conditions are used

Excess  $\rightarrow m+sd \leq r$

Deficit  $\rightarrow m-sd \geq r$

Normal  $\rightarrow m-sd < r < m+sd$

Where,

$r \rightarrow$  Annual Rainfall rate

$m \rightarrow$  Mean value

$sd \rightarrow$  Standard Deviation

**E. Evaluation Metrics Correlation Coefficient**

The variation has been calculated between predicted and actual values by the factor correlation coefficient. If the obtained results are high the proposed model is the best model with perfect statistical correlation otherwise there is no correlation between predicted and actual values.

**Root Mean Squared Error**

This is the average amount of error in the output variable. This will help to get an idea on the amount a prediction wrong on average.

**The magnitude of Relative Error**

This is the difference between the actual and estimated effort relative to the actual effort.

The other performance metrics are accuracy, precision, recall, F-Measure values.

**IV. MATERIALS AND METHODOLOGY**

**A. Data Collection**

In this proposed method input data has been collected from the Open Government Data(OGD) Platform supported by the government of India. We extracted month wise rainfall dataset for TamilNadu region covered the period of 1392 months from Jan 1901 to Dec 2017 and the attributes are shown in Table 1[8]. The meteorological data is collected in millimeter values and the dataset statistical values are shown in Table 2.

**Table 1: Attributes of Dataset**

Attribute	Data Type	Description of Attribute
YEAR	Numerical	A year has been considered
JAN-DEC	Numerical	Monthly rainfall rate
RESULT	Nominal	Descriptive term used

**B. Data Pre-Processing**

Pre-processed data play a major role for further work. The dataset imported in WEKA tool for removing noisy, incomplete, inconsistent and redundant data. Some data was selected to train the neurons and the remaining was used to testing the model.

**C. Data Transformation**

In this stage, the collected data is transformed into a CSV(Comma Separated Value) by saving as Excel formatted dataset into a .csv format because weka tool supports ARFF, CSV, C4.5, and binary format so we have to convert the file format. For the prediction of one year ahead of data the ARFF file is loaded as input data for MLP and then the predicted data is given as CSV format input for neurons in the network.

**Table 2: Analysis of Numeric Data Values**

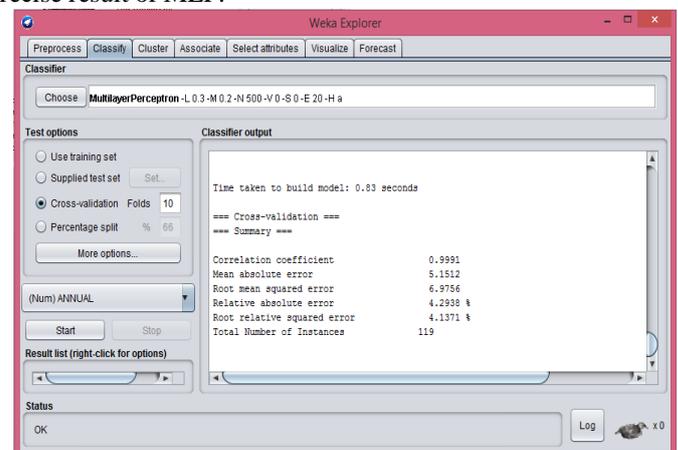
NO	Variable	Statistics			
		Min	Max	Mean	SD
1	YEAR	1901	2017	-	-
2	JAN	0.1	141.2	23.75	32.06
3	FEB	0	131.3	13.209	19.4
4	MAR	0	164.7	19.47	22.30
5	APR	5.5	132.1	44.43	28.02
6	JUN	21.2	128.1	52.12	19
7	JUL	11.8	174.7	71.2	30.94
8	AUG	18.7	206.6	96.08	37.35
9	SEP	26.1	210.4	111.52	39.31
10	OCT	62.1	436.1	181.96	67.5
11	NOV	23.6	416.7	175.38	90.66
12	DEC	0.6	295.8	81.15	65.45
13	ANNUAL	318	1365.3	940.47	168.89

**V. EXPERIMENTAL RESULT**

This work has been done in windows platform with WEKA tool for entire implementation.

**A. Training and testing of Feed Forward Neural Network**

Multilayer Perceptron in weka that connects a set of input data into an appropriate output. Momentum term is used to increase the performance of the algorithm[20]. The ZeroR classified model output is loaded as ARFF file format and it is given a test set for Multilayer Perceptron function. By using this function the result was predicted for one year ahead. The cross-validation test case is used to achieve the precise result of MLP.



**Fig. 1. Classifier Output for Predicted Rainfall Data**

Figure 1 shows the accurate result for the given rainfall dataset. In which the Correlation coefficient is 0.9991, it is a perfect statistical correlation and the Mean squared error is 6.9756 and root relative squared error is 4.1371%.

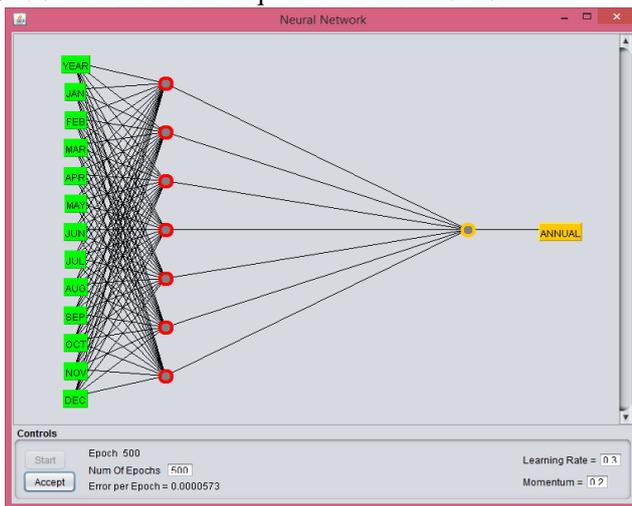


Fig. 2. Multilayer perceptron after prediction

Figure 2 shows the multilayer perceptron architecture for the predicted dataset. The number of epochs given is 500 and the error per epoch is 0.0000573 and the learning rate and momentum rate is set to default values. The purpose of the learning rate is to update the weights of neural each iteration and momentum term used to increase the accuracy and speed of the neural network.

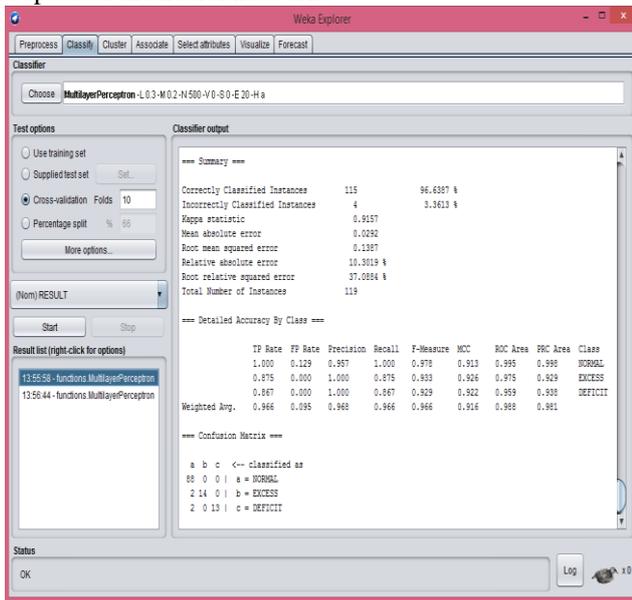


Fig. 3. Accuracy Result for Classified Output

Figure 3 shown here summarize the dataset and give the three classes of the result That is Excess, Deficit and Normal classes of rainfall. The dataset contains 96.6387% of correctly classified instances and kappa statistic is 0.9157. This provides an accurate result for time series data. Cross-validation is chosen for training dataset. It divides the dataset into 10 folds and it takes 9 folds as a training set and 1 fold is taken as a testing set.

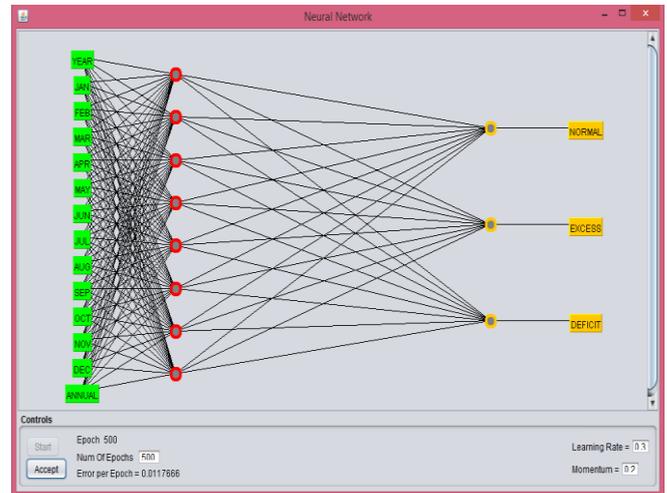


Fig. 4. Neural Network for classified instances

Figure 4 shows the neural network architecture for classified instances of rainfall dataset. For this also the number of epochs is 500 and learning and momentum rate is also fixed as default values. The error per epoch is 0.011766.

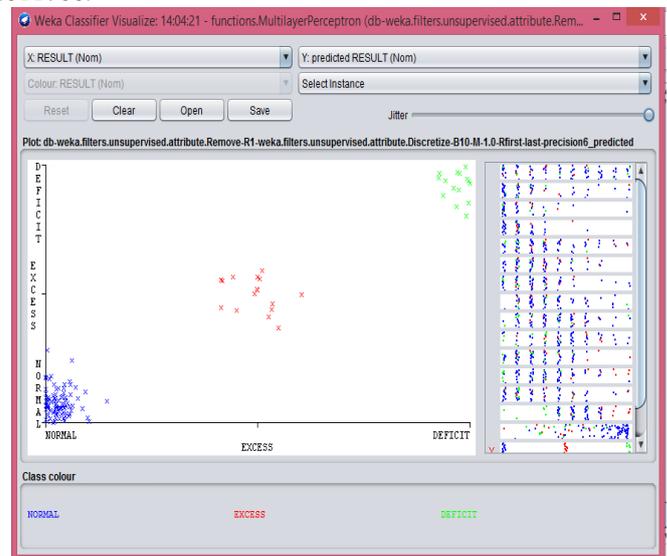


Fig. 5. Visualize Classification Errors

Visualize classify error is shown in Figure 5. The weka tool works best with numerical data like this time-series data. The above graph contains little crosses that denote the instances and the color of this graph depends on the class of the instance. The x-axis shows sepal length attribute, and the y-axis shows petal width.

VI. CONCLUSION

Time series data is a temporal data and it can be accessed well by Artificial Neural Network technique. Neural network provides the best result for classification and prediction problems. By combining these time series data and Artificial Neural Network is used for prediction provide the best result than other data mining techniques. In these models, multi-layer perceptron is used to provide an accurate result. In future work, we can develop for the entire nation.

## REFERENCES

1. Zhan Jie Wang and A. B. M. Mazharul Mujib, "The Weather Forecast Using Data Mining Research Based on Cloud Computing", IOP Conf. Series: Journal of Physics: Conf. Series 910 012020, 2017.
2. Suvridha Jambekar, Shikha Nema, Zia Saquib, "Application of Data Mining Techniques for Prediction of Crop Production in India", International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 7, Issue 4 April 2018.
3. Folorunsho Olaiya, Adesesan Barnabas Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies" I.J. Information Engineering and Electronic Business, 1, 51-59, 2012.
4. Shabib Aftab, Munir Ahmad, Noureen Hameed, Muhammad Salman Bashir, Ifikhar Ali, Zahid Nawaz, "Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 5, 2018.
5. Aishwarya Dhore1, Anagha Byakude2, Bhagyashri Sonar3, Mansi Waste4, "Weather prediction using the data mining Techniques", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 04 Issue: 05 | May -2017.
6. Neelam Mishra, Hemant Kumar Soni, Sanjiv Sharma, A K Upadhyay, "Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by using Time-series Data", I.J. Intelligent Systems and Applications, 2018, 1, 16-23.
7. Fu Tak-Chung. "A review on time series data mining". Engineering Applications of Artificial Intelligence. Vol. 24, pp 164-181, 2011.
8. Hiren Kumar Deva Sarma, Swapnil Mishra, "Mining Time Series Data with Apriori/Tid Algorithm" 2016 International Conference on Information Technology.
9. Folorunsho Olaiya, Adesesan Barnabas Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies" I.J. Information Engineering and Electronic Business, 2012, 1, 51-59.
10. Sankhadeep Chatterjee, Bimal Datta, Soumya Sen, Nilanjan Dey, "Rainfall Prediction using Hybrid Neural Network approach", 2018 2<sup>nd</sup> International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom).
11. M.Kannan, 2 S.Prabhakaran, 3 P.Ramachandran, "Rainfall Forecasting Using Data Mining Technique", M. Kannan et. al. / International Journal of Engineering and Technology Vol.2 (6), 397-401, 2010
12. K.Abhishek, A.Kumar, R.Ranjan and S.Kumar, "A rainfall Prediction Model using Artificial Neural Network", IEEE Control and System Graduate Research Colloquium, 2012
13. Mary N. Ahuna, Thomas J. Afullo and Akintunde A. Alonge, "Rainfall Rate Prediction based on Artificial Neural Networks For Rain fade Mitigation over Earth-Satellite link", IEEE African 2017 Proceedings.
14. Mr. Chetan C. Janbandhu, Prof. Praful D. Meshram, Prof. Madhuri N. Gedam, "Modelling Rainfall Prediction Using Data Mining Method - A Bayesian Approach", International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248 Volume: 3 Issue: 11 472 – 474.
15. Mislan, Haviluddin, Sigit Hardwinarto, Sumaryono, Marlon Aipassa, "Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggara Station, East Kalimantan - Indonesia", International Conference on Computer Science and Computational Intelligence (ICCSICI 2015)
16. Joko Azhari Suyatno, Fhira Nhita, Anq Atiqi Rohmawati, "Rainfall Forecasting in Bandung Regency using C4.5 Algorithm", 2018 6<sup>th</sup> International Conference on Information and Communication Technology (ICoICT).
17. Afolayan Abimbola Helen, Ojokoh Bolanle A., Falaki Samuel O, "Comparative Analysis of Rainfall Prediction Models Using Neural Network and Fuzzy Logic", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-5 Issue-6, January 2016.
18. Aishwarya Dhore1, Anagha Byakude2, Bhagyashri Sonar3, Mansi Waste4, "Weather prediction using the data mining Techniques", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 04 Issue: 05 | May -2017
19. Dataset collected from Open Government Data (OMD) platform, India- <https://data.gov.in/catalog/rainfall-india>
20. Koushal Kumar, GourSundarMitra Thakur, "Extracting Explanation from Artificial Neural Networks", International Journal of Computer Science and Information Technologies, Vol. 3(2), 2012
21. Mrs. Bharati M. Ramageri, "Data Mining Techniques and Applications", Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305.
22. Ms. Sonali. B. Maind, Ms. Priyanka Wankar, "Research Paper on Basic of Artificial Neural Network", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 1 96 – 100