# Classification and Regression Tree to Predict the Precipitation Labels of North-West Region in Bangladesh

Md. Habibur Rahman[1] & Md. Moyazzem Hossain[1]

[1] Department of Statistics, Jahangirnagar University, Bangladesh

Correspondence: Md. Habibur Rahman, Department of Statistics, Jahangirnagar University, Bangladesh. E-mail: habib.drj@juniv.edu; habib.drj@gmail.com

## Abstract

The north-west region of Bangladesh is completely an agricultural region and prediction of precipitation labels is extremely important to people of this region as well as for the policymaker also. The assumptions free model like the CART model is very helpful to predict the precipitation categories of the north-west part of Bangladesh. The fitted CART model and the pruned CART model provide approximately eighty percent accurate prediction to predict the precipitation labels of the north-west grid in Bangladesh in this study.

**Keywords:** CART, Training Data, Gini Index, Confusion Matrix, Accuracy

## 1. Introduction

Bangladesh is the country where the economy mainly depends on agriculture. The impact of low and high precipitation is very significant for agricultural production in the north-west region of Bangladesh. Livestock and human ecology depend on the water system especially precipitation. The drought of different intensities occurs across Bangladesh. Rice and other crops productions are severely affected by dryness. To make a better understanding of precipitation system this can help the formulation of policies that might include preserving and efficient use of precipitation water in near future. In order to implement a better policy the policymakers especially farmers need to know the future state of the amount of precipitation. Since in all weather happenings, precipitation plays the most imperative role in human life especially human civilization to a great extent depends upon its frequency and amount to various scales. Data mining is a process of analyzing data from different perspectives and gathering knowledge from it. Various studies have been carried out that focus on data mining especially classification algorithms. One of the most efficient, easy to implement, and effective classification method to mine the data from a large database is decision tree construction method. Here, different decision tree algorithms applied for various datasets are considered and explained. There are many studies that support the applicability of data mining techniques for prediction. Being depended on time and space, prediction of precipitation labels or any kinds of weather forecasting is a chaotic system. The weather forecasting is one of the most imperative and demanding operational responsibilities carried out by meteorological services all over the world. It is a complicated procedure that includes numerous specialized fields of know-how. The task is complicated because in the field of meteorology all decisions are to be taken in the visage of uncertainty. A number of authors Brown and Murphy (1988), Wilks (1991), Elsner and Tsonis (1992), Cartalis and Varotsos (1994), Jacovides et al. (1994), Kondratyev and Varotsos (2001), Varotsos et al. (2001) have discussed the uncertainty associated with the weather systems. Chaotic features associated with the atmospheric phenomena also have been attracted the attention of the modern scientists Sivakumar et al. (1999), Sivakumar (2000), Men et al. (2004), Varotsos (2005), Varotsos and Krik-Davidoff (2006). Saxena et al. (2013) presented the review of weather prediction using artificial neural networks and studied the benefit of using it. CART, an abbreviation of Classification and Regression Trees, was first introduced by Bremain et al. (1983) and Bremain et al. (1984) which is a binary tree using GINI Index as its splitting criteria. CART can handle both nominal and continuous attributes to construct a decision tree. Also, CART can handle missing values by surrogating tests to fairly accurate outcomes. CART is also a cross-validated method in its pruning technique. Petre (2009) presented a small application of CART decision tree algorithm for weather prediction of Hong Kong based on year, month, average pressure, relative humidity, clouds quantity, precipitation, and average temperature. Kalyankar and Alaspurkar (2013) used data mining techniques to acquire weather data and find the hidden patterns inside the large dataset so as to transfer the retrieved information into

usable knowledge for classification and prediction of weather condition of Gaza city and this knowledge can be used to obtain useful predictions and support the decision making the process. Kumar (2013) compared the different classification methods namely Decision Trees, Rule-based Methods, Neural Networks, Naïve Bayes, Bayesian Belief Networks, and Support Vector Machines. The task weather prediction also requires sophisticated statistical models, which do not need making assumptions regarding the system. Many studies around the globe have developed stochastic weather models which are basically statistical models that can be used as a random number generators whose output resembles the weather data to which they have been fit (Wilks, 1999). Hu (1964) initiated the implementation of the artificial neural network (ANN), an important soft computing methodology in weather forecasting. Surajit and Manojit (2007) used a soft computing technique in rainfall forecasting. Kalogirou et al. (1997) used ANN in rainfall prediction by splitting the available data into homogeneous subpopulations. The self-organizing map, backpropagation neural networks, and fuzzy rule systems used as a soft computing technique to predict rainfall (Wong et al., 2003). Michaelides et al. (1995) compared the performance of ANN with multiple linear regressions in estimating missing rainfall data over Cyprus. The main objectives of this research work are – to reveals the nature of atmospheric data, to fit the classification and regression tree (CART) to predict the precipitation labels, to test the quality-of-fit of CART by the goodness-of-fit statistic and the accuracy measures, and to select the best-fitted CART. In this research work especially develops and present cluster following steps will be implemented – Collecting of Data of interest and exploratory data analysis is such as visualization, detection of outlier, trimming etc., the data must be taken into streamline, also as well as summarized, Reveals the nature of atmospheric data, Fit and select the best-fitted CART, and test the quality-of-fit of the CART by the goodness-of-fit statistic. The data for this study have been collected from Bangladesh Meteorological department on temperature (TEM), due point temperature (DPT), precipitation (PRE), wind speed (WIS), humidity (HUM) and sea level pressure (SLP) for this study of north-west region for the period of 1960-2015. This study is organized into different sections including this section. The second section reviews the theory, methods and the methodology of different techniques related to Classification and Regression Tree (CART). In this section, the basic concepts of the CART, Gini index, and accuracy measures are discussed. Also in this section, the different goodness-of-fit tests for fitted CART model are talked about. The third section is center of attention on the computational issues and results of the fitted CART model. As well as the discussions about the results are presented in this segment. A summary with some concluding explanations and some suggestions for further research is contained in the final section.

## 2. Method and Methodology

This research work is devoted to Classification and Regression Trees (CART). CART is a classification method which uses historical data to construct decision trees. Depending on available information about the dataset, classification tree or regression tree can be constructed. The constructed tree can be then used for classification of new observations. The CART is the fundamental principles of tree construction, different splitting algorithms, and pruning procedures. During tree construction, attribute selection measures are used to select the attribute that best partitions the tuples into distinct classes. When decision trees are built, many of the branches may reflect noise or outliers in the training data. An attribute selection measure is a heuristic for selecting the splitting criterion that "best" separates a given data partition, $D$, of class-labeled training tuples into individual classes. Attribute selection measures are also identified as splitting rules because they determine how the tuples at a given node are to be split.

The attribute selection measure provides a ranking for each attribute describing the given training tuples. The attribute having the best score for the measure is chosen as the splitting attribute for the given tuples.In this section, three popular attribute selection measures information gain, gain ratio, and Gini index is discussed. The notation used here is as follows. Let $D$, the data partition, is a training set of class-labeled tuples. Suppose the class label attribute has m distinct values defining m distinct classes, $C_i$ (for $i = 1... m$). Let be the set of tuples of class $C_i$ in $D$. Let $|D|$ and $|\ |$ denote the number of tuples in $D$ and, respectively.

The Gini index is used in CART as attribute selection measure.

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2 \tag{1}$$

Where $p_i$ is the probability that a tuple in $D$ belongs to class $C_i$ and is estimated by $|C_{i,D}|/|D|$. The sum is computed over $m$ classes. The Gini index considers a binary split for each attribute. Let's first consider the case where $A$ is a

discrete-valued attribute having $v$ distinct values, $\{a_1, a_2...a_v\}$ occurring in $D$. To determine the best binary split on $A$. There are $2^v - 2$ possible ways to form two partitions of the data, $D$, based on a binary split on $A$. When considering a binary split, by computing a weighted sum of the impurity of each resulting partition. For example, if a binary split on $A$ partitions $D$ into $D_1$ and $D_2$, the Gini index of $D$ given that partitioning is –

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \tag{2}$$

For each attribute, each of the possible binary splits is considered. For a discrete-valued attribute, the subset that gives the minimum Gini index for that attribute is selected as its splitting subset. The reduction in impurity that would be incurred by a binary split on a discrete- or continuous-valued attribute $A$ is –

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \tag{3}$$

The attribute that maximizes the reduction in impurity (or, equivalently, has the minimum Gini index) is selected as the splitting attribute. When a decision tree is built, many of the branches will reflect anomalies in the training data due to noise or outliers. Tree pruning methods address this problem of over fitting the data. Such methods typically use statistical measures to remove the least-reliable branches.

*Model Evolution and Selection:* To build a classification model, then like an estimate of how accurately the classifier can predict for future purpose, that is future atmospheric data on which the classifier has not been trained. This may even have tried different methods to build more than one classifier and now wish to compare their accuracy. In this section, it is described various evaluation metrics for the predictive accuracy of a classifier. Holdout and random sub-sampling, cross-validation, and bootstrap methods are common techniques for assessing accuracy, based on randomly sampled partitions of the given data. Tests of statistical significance to assess whether the difference in accuracy between two classifiers is due to chance. Presents how to compare classifiers based on cost-benefit and receiver operating characteristic (ROC) curves. Accuracy matrix is what that usually mean, when using the term accuracy. It is the ratio of number of correct predictions to the total number of input samples (Table 1). Calculating an accuracy matrix can give a better idea of what classification model is getting right and what types of errors it is making. Accuracy can be calculated by –

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made} \tag{4}$$

It works well only if there are equal number of samples belonging to each class. It is also a technique for summarizing the performance of a classification algorithm.

Table 1. The Confusion Matrix to find the accuracy in Classification.

| Confusion Matrix | | Target | | | |
|---|---|---|---|---|---|
| | | Positive | Negative | | |
| Model | Positive | $a$ | $b$ | Positive Predictive Value | $\frac{a}{(a+b)}$ |
| | Negative | $c$ | $d$ | Negative Predictive Value | $\frac{d}{(c+d)}$ |
| | | Sensitivity $\frac{a}{(a+c)}$ | Specify $\frac{d}{(b+d)}$ | $Accuracy = \frac{(a+d)}{(a+b+c+d)}$ | |

Accuracy matrix is great, but gives us the false sense of achieving high accuracy. The real problem arises, when the cost of misclassification of the minor class samples are very high.

## 3. Results and Discussions

The atmospheric parameters temperature (TEM), dew point temperature (DPT), wind speed (WIS), sea level pressure (SLP), and humidity (HUM) of three different locations of North-West region collected from the Bangladesh Meteorological Department used to predict the precipitation (PRE) labels. The data on the atmospheric variables precipitation (PRE), temperature (TEM), dew point temperature (DPT), wind speed (WIS), sea level pressure (SLP), and humidity (HUM) of three different locations of North-West region have been collected from the Bangladesh Meteorological Department. The locations (Latitude, Longitude) are Dinajpur (25.65, 88.68), Rangpur (25.73, 89.27), and Sydpur (25.75, 89.92). The monthly data for this study are generated from the available atmospheric data record stations in Bangladesh for the North-West region (88E-90E, 25N-27N) for the period 1960-2015. The World Meteorological Organization (WMO) Classification for rainfall within twenty-four hours categories are light rain (1-10mm), moderate rain (11-22mm), moderately heavy rain (23-43mm), heavy rain (44-88mm), very heavy rain with greater than 88mm (Mannan, 2008; Khatun et al., 2016, Mannan et al., 2016). In this study, the monthly data for precipitation, temperature, dew point temperature, wind speed, sea level pressure, and humidity are generated from the available data of three locations. The precipitation series are generated based on the average of total precipitation in a month for each location and then take the average of available three locations precipitations. The summary statistics for different six atmospheric parameters are presented in Table 2. The maximum statistic of atmospheric parameters found 30.6 $^0$C, 26.5 $^0$C, 40.48mm, 6.13 kph, 91.93 %, and 1019 hPa for Temperature, Dew Point Temperature, Precipitation, Wind Speed, Humidity, and Sea Level Pressure respectively. The correlations among atmospheric parameters are presented in Table 3 and the correlation among precipitation and other parameters temperature, dew point temperature, wind speed, humidity, sea level pressures are 0.6136, 0.7098, 0.1820, 0.4747, -0.7404 respectively. The temperature, dew point temperature, and humidity are negatively skewed; precipitation is positively skewed; and the wind speed and sea level pressure are approximately symmetric (Figure 1).

Table 2. Summary Statistics

| Statistics | Temperature | Dew Point Temperature | Precipitation | Wind Speed | Humidity | Sea Level Pressure |
|---|---|---|---|---|---|---|
| Mean | 24.44 | 19.17 | 5.50 | 2.78 | 78.09 | 1008.01 |
| Standard Error | 0.17 | 0.22 | 0.25 | 0.03 | 0.35 | 0.21 |
| Median | 26.00 | 19.70 | 2.80 | 2.77 | 80.61 | 1008.29 |
| Mode | 28.70 | 26.00 | 0.00 | 2.50 | 84.27 | 1005.40 |
| Standard Deviation | 4.44 | 5.58 | 6.42 | 0.88 | 8.95 | 5.50 |
| Kurtosis | -0.98 | -0.55 | 1.40 | 0.34 | 2.35 | -1.28 |
| Skewness | -0.61 | -0.49 | 1.25 | 0.17 | -1.42 | -0.03 |
| Minimum | 13.60 | 0.00 | 0.00 | 0.00 | 39.27 | 997.55 |
| Maximum | 30.60 | 26.50 | 40.48 | 6.13 | 91.93 | 1019.00 |
| Confidence Level (95.0%) | 0.34 | 0.42 | 0.49 | 0.07 | 0.68 | 0.42 |

Table 3. Correlation among Atmospheric parameters-Temperature (TEM), Dew point temperature (DPT), Precipitation (PRE), Wind Speed (WIS), Humidity (HUM), and Sea level pressure (SLP) of north-west region of Bangladesh for the period of 1960-2015

| Parameters | TEM | DPT | PRE | WIS | HUM | SLP |
|---|---|---|---|---|---|---|
| TEM | 1.0000 | 0.8136 | 0.6136 | 0.3472 | 0.1134 | -0.8658 |
| DPT | 0.8136 | 1.0000 | 0.7098 | 0.1758 | 0.5297 | -0.7648 |
| PRE | 0.6136 | 0.7098 | 1.0000 | 0.1820 | 0.4747 | -0.7404 |
| WIS | 0.3472 | 0.1758 | 0.1820 | 1.0000 | -0.2687 | -0.3084 |
| HUM | 0.1134 | 0.5297 | 0.4747 | -0.2687 | 1.0000 | -0.2407 |
| SLP | -0.8658 | -0.7648 | -0.7404 | -0.3084 | -0.2407 | 1.0000 |

To fit CART predictive model to predict the precipitation labels the following three categories are used - (i) No rain (dry day) and Trace (NT), (ii) Light rain (1-10mm) (L), (iii) moderate rain (11-22mm), moderately heavy rain (23-43mm), heavy rain (44-88mm), and very heavy rain with greater than 88mm (MH) (Figure 2). The portion of different categories of precipitation labels are 40.48%, 35.71%, and 23.81% for Category I (NT- No rain and Trace), Category II (L-Light rain (1-10mm)) and Category III (MH - moderate rain (11-22mm), moderately heavy rain (23-43mm), heavy rain (44-88mm), and very heavy rain with greater than 88mm) respectively (Figure 2).

**(a)**        **(b)**
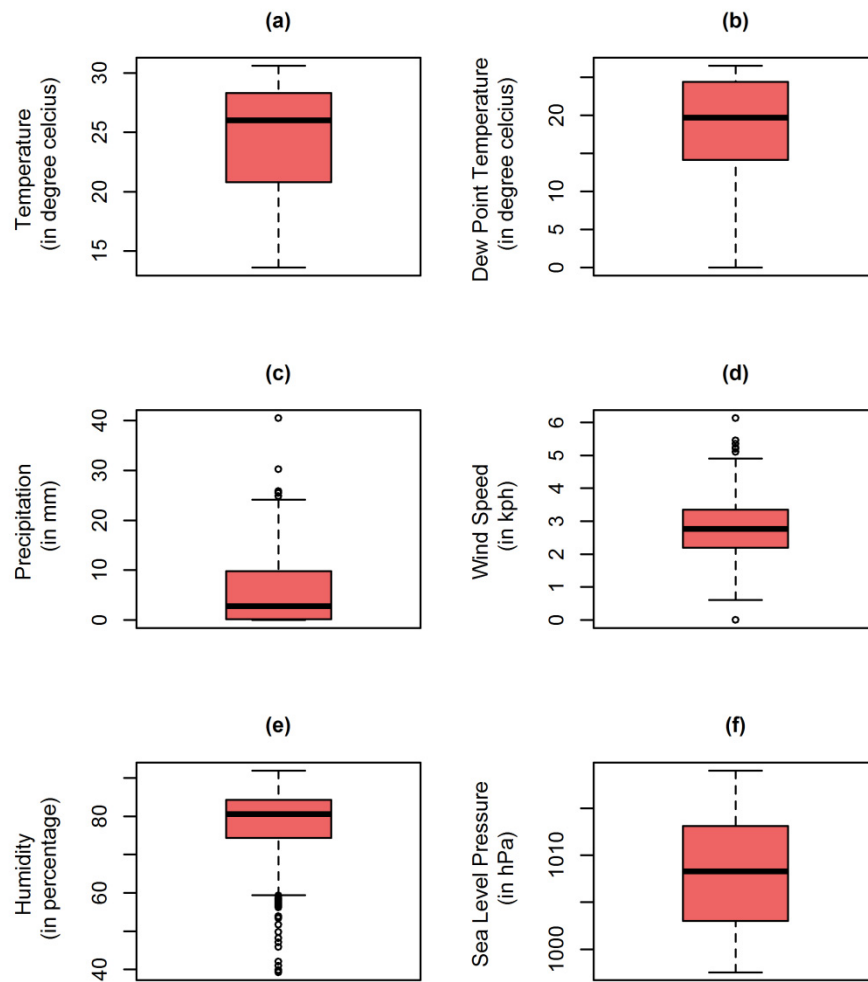
**(c)**        **(d)**

**(e)**        **(f)**

Figure 1. Box-and-Whisker plot for (a) Temperature (b) Dew point temperature (c) Precipitation (d) Wind Speed (e) Humidity and (f) Sea level pressure of the north-west region in Bangladesh for the period of 1960-2015
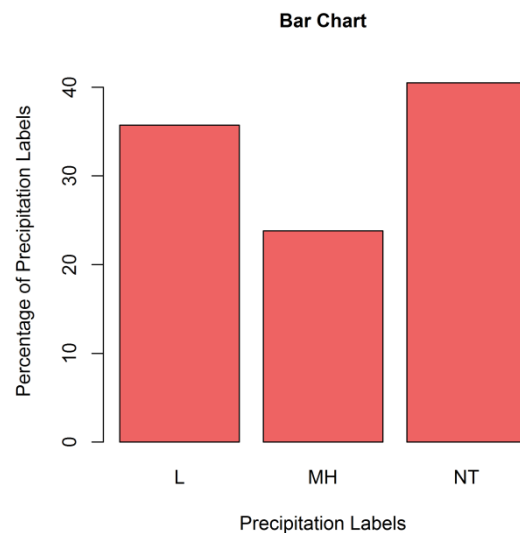
**Bar Chart**

Figure 2. Schematic bar chart for the precipitation categories – (i) No rain (dry day) and Trace (NT), (ii) Light rain (1-10mm) (L), (iii) moderate rain (11-22mm), moderately heavy rain (23-43mm), heavy rain (44-88mm), and very heavy rain with greater than 88mm (MH)
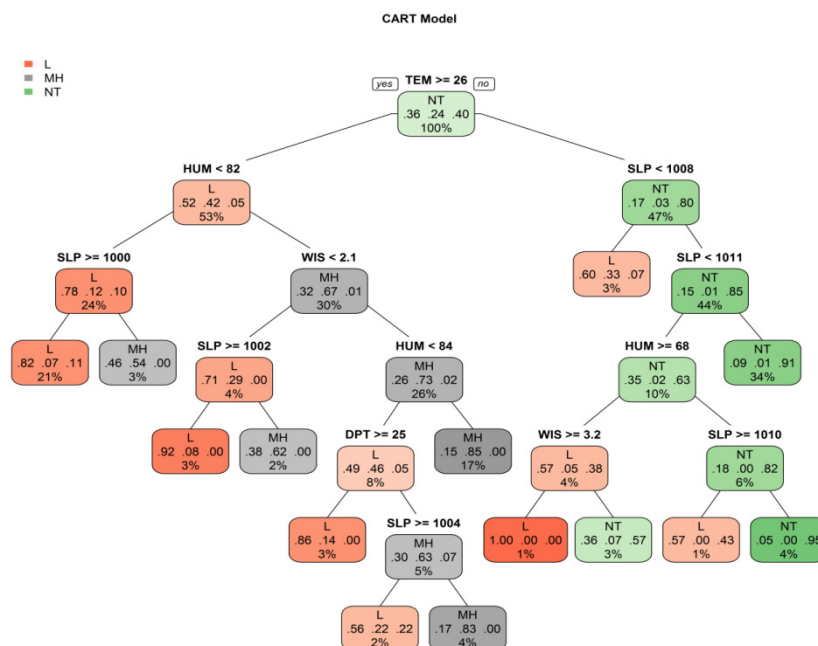
Figure 3. CART model to predict the precipitation labels of north-west region in Bangladesh

Table 4. CP, relative error, xerror, and xstandard deviation table

| CP | nsplit | rel error | xerror | xstd |
|---|---|---|---|---|
| 0.4191 | 0 | 1.000 | 1.000 | 0.0363 |
| 0.1716 | 1 | 0.581 | 0.617 | 0.0358 |
| 0.0297 | 2 | 0.409 | 0.498 | 0.0339 |
| 0.0264 | 3 | 0.380 | 0.488 | 0.0337 |
| 0.0165 | 4 | 0.353 | 0.465 | 0.0333 |
| 0.0099 | 6 | 0.320 | 0.426 | 0.0323 |
| *0.0066* | *7* | *0.310* | *0.413* | *0.0320* |
| 0.0033 | 11 | 0.281 | 0.429 | 0.0324 |
| 0.0010 | 13 | 0.274 | 0.419 | 0.0322 |

The seventy five percent randomly data of total 672 observations (months) are used as the training data to fit the classification and regression tree model to predict the precipitation labels of north-west region of Bangladesh, and rest of the data (168 observations ) are used as the test data. The fitted CART model is presented in Figure 3. The atmospheric variable temperature is used as a first splitting variable based on Gini index. The median value of temperature is used as the splitting criteria; and fifty three percent observations are found for the temperature greater than or equal to 26 degree centigrade, and rest of the forty seven observations are found for less than 26 degree centigrade (Figure 3). At the terminal Node-8 (at left side) 106 observations (21%) of training data (504 observations or 75% of total data) are predicted as light rain where 87 (82%) observations are correctly predicted as light rain, 7 (7%) observations (from moderate, moderately heavy, and heavy rain category) and 12 (11%) observations (from no rain and trace rain category) are wrongly predicted as light rain. The variables importance's in percentage to fit the classification and regression tree model are 25%, 25%, 23%, 19%, and 8% for sea level pressure, temperature, dew point temperature, humidity, and wind speed respectively (Figure 4).
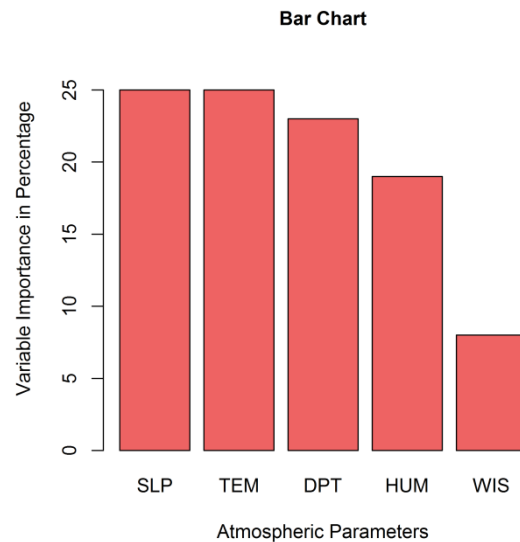
**Bar Chart**



Figure 4. Chart for the importance in percentage of atmospheric variables – Sea Level Pressure (SLP), Temperature (TEM), Dew Point Temperature (DPT), Humidity (HUM), and Wind Speed (WIN)
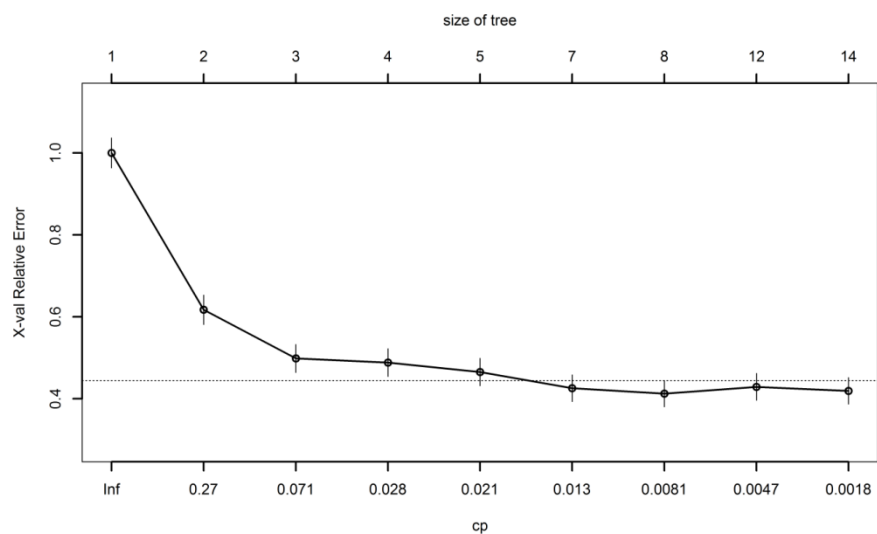


Figure 5. Schematic plots for CP (Complexity Parameter) and relative error

The complexity parameter (cp) is used to control the size of the decision tree and to select the optimal tree size. If the cost of adding another variable to the decision tree from the current node is above the value of cp, then tree building does not continue. It could also say that tree construction does not continue unless it would decrease the overall lack of fit by a factor of cp. The values of the complexity parameter (cp), relative error, error present in Table 4. Also the best cp (0.00660066) found from the graphic representation for the relative error and value of complexity parameters and table for complexity parameter (cp), relative error, and error (Table 4 and Figure 5). The prediction accuracy rates based on CART model for training data, test data, and overall data come across as 83.53%, 73.81%, and 81.10% respectively (Table 5).

Table 5. Confusion Matrix for training data, test data, and over all data based on CART model

| Confusion Matrix | | | Predicted | | | Accuracy Rate |
|---|---|---|---|---|---|---|
| | | Labels | NT | L | MH | |
| Training Data | Actual | NT | 183 | 18 | 0 | |
| | | L | 21 | 136 | 25 | Accuracy rate 83.53% |
| | | MH | 2 | 17 | 102 | |
| Test Data | Actual | NT | 60 | 10 | 1 | |
| | | L | 6 | 41 | 11 | Accuracy rate 73.81% |
| | | MH | 0 | 16 | 23 | |
| Overall data | Actual | NT | 243 | 28 | 1 | |
| | | L | 27 | 177 | 36 | Accuracy rate 81.10% |
| | | MH | 2 | 33 | 125 | |

The prediction error rate based on CART model for training, test, and over all data lies within 16% to 26%. The tree is pruned based on the best complexity parameter (0.00660066) and obtained a pruned classification and regression tree (Figure 6).
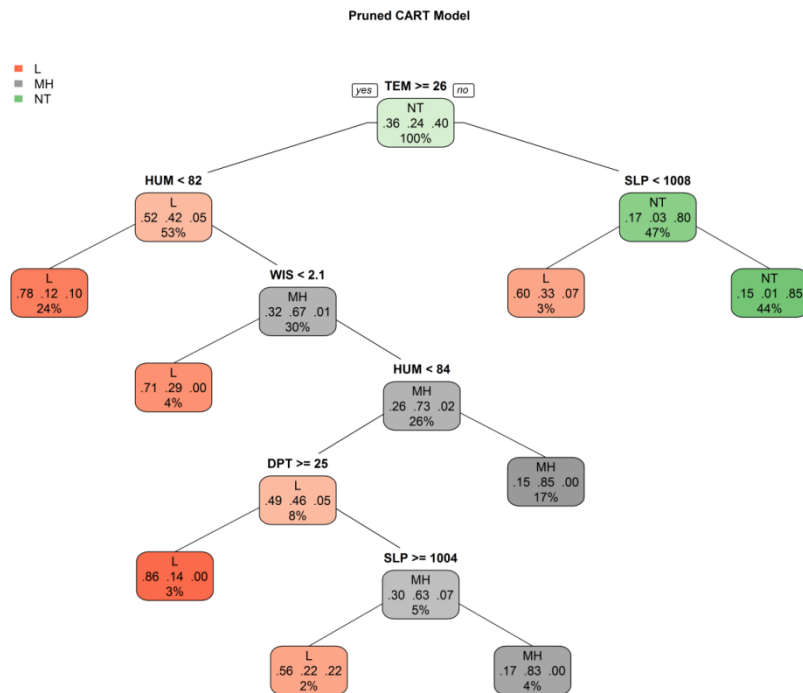


Figure 6. Pruned Classification and Regression model to predict the precipitation labels of north-west region in Bangladesh

Table 6. Confusion Matrix for training data, test data, and over all data based on pruned CART model

| Confusion Matrix | | | Predicted | | | Accuracy Rate |
|---|---|---|---|---|---|---|
| | | Labels | NT | L | MH | |
| Training Data | Actual | NT | 186 | 15 | 0 | |
| | | L | 32 | 134 | 16 | Accuracy rate 81.35% |
| | | MH | 2 | 29 | 90 | |
| Test Data | Actual | NT | 62 | 8 | 1 | |
| | | L | 6 | 46 | 6 | Accuracy rate 76.79% |
| | | MH | 0 | 18 | 21 | |
| Overall data | Actual | NT | 248 | 23 | 1 | |
| | | L | 38 | 180 | 22 | Accuracy rate 80.21% |
| | | MH | 2 | 47 | 111 | |

The pruned classification and regression tree predict accurately the different precipitation labels 81.35 percent for training data, 76.79 percent for test data, and 80.21 percent for the entire data (Table 6).

## 4. Summary and Conclusion

The advantages of classification and regression tree model are pre-assumptions free. The chaotic system like atmospheric condition prediction should not be bounded by any assumptions. The fitted CART model provides approximately eighty percent accurate predictions for training, test, and entire data in this study. The fitted classification and regression tree model is pruned to avoid the complexity of the model based on the best complexity parameter to obtain a more easily understandable pruned CART model. The pruned CART model also provides approximately eighty percent accurate prediction to predict the different categories of precipitation labels of the north-west region in Bangladesh. The research highly suggests using the assumptions free pruned CART model to predict the precipitation categories of the north-west region in Bangladesh because of its simplicity.

### Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### References

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth Inc.

Breiman, L., Friedman, J., Olshen, R., Stone, C., Steinberg, D., & Colla, P. (1983). *CART: Classification and regression trees* (p. 156). Wadsworth: Belmont, CA.

Brown, B. G., & Murphy, A. H. (1988). The economic value of weather forecasts in wildfire suppression mobilization decisions. *Canadian J. Forest Res., 18*, 1641-1649.

Cartalis, C., & Varotsos, C. (1994). Surface ozone in Athens, Greece, at the beginning and at the end of the 20th-century. *Atmos. Environ., 28*, 3-8.

Elsner, J. B., & Tsonis, A. A. (1992). Non-linear prediction, chaos, and noise. *Bull. Am. Me-teor. Soc., 73*, 49-60.

Hu, M. J. C. (1964). *Application of ADALINE system to weather forecasting*. Technical Report, Stanford Electron, Stanford, CA.

Jacovides, C. P., Varotsos, C., Kaltsounides, N. A., Petrakis, M., & Lalas, D. P. (1994). Atmospheric turbidity parameters in the highly polluted site of Athens basin. *Renewable Energy, 4*(5), 465-470.

Kalogirou, S. A., Constantinos, C. N., Michaelides, S. C., & Schizas, C. N. (1997). *A time series construction of precipitation records using Artificial Neural Networks. EUFIT '97*. September 8-11, 2409-2413.

Kalyankar, M. A., & Alaspurkar, S. J. (2013). Data Mining Technique to Analyse the Metrological Data. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(2), 114-118.

Khatun, M. A., Rashid, M. B., & Hygen, H. O. (2016). *MET report, Climate of Bangladesh, 08/2016*.

Kondratyev, K. Y., & Varotsos, C. A. (2001). Global tropospheric ozone dynamics – Part I: Tropospheric ozone precursors. *Environ. Sci. Pollution Res, 8*(1), 57-62.

Kumar, R. (2013). Decision Tree for the Weather Forecasting. *International Journal of Computer Applications, 76*(2), 31-34.

Mannan, M. A., & Karmakar, S. (2008). Climatic features of Heavy Rainfall activities in monsoon season and its socio-economic impact in Bangladesh. *Proceedings of SAARC Seminar on Application of Weather and Climate Forecasts in the Socio-economic Development and Disaster Mitigation, 05-07 August, 2007, Dhaka, Bangladesh* (pp. 95-115).

Mannan, M. A., Chowdhury, M. A. M., & Karmakar, S. (2016). Prediction of Rainfall over Southeastern part of Bangladesh during Monsoon Season. *International Journal of Integrated Sciences & Technology, 2*, 73-82.

Men, B., Xiejing, Z., & Liang, C. (2004). Chaotic analysis on monthly precipitation on Hills Region in Middle Sichuan of China. *Nature and Science, 2*, 45-51.

Michaelides, S. C., Neocleous, C. C., & Schizas, C. N. (1995). Artificial Neural Networks and multiple linear regression in estimating missing rainfall data. *Proc. DSP95 Intern. Confer. "Digital Signal Processing", Limassol, Cyprus* (pp. 668-673).

Petre, E. G. (2009). A Decision Tree for Weather Prediction. *Buletinul, LXI*(1), 77-82.

Saxena, A., Verma, N., & Tripathi, K. C. (2013). A Review Study of Weather Forecasting Using Artificial Neural Network Approach. *International Journal of Engineering Research & Technology (IJERT), 2*(11), 2029-2035.

Sivakumar, B. (2000). Chaos theory in hydrology: Important issues and interpretations. *J. Hy-drology, 227*(1-4), 1-20.

Sivakumar, B., Liong, S. Y., Liaw, C. Y., & Phoon, K. K. (1999). Singapore rainfall behavior: Chaotic. *J. Hydrol. Eng., ASCE 4*, 38-48.

Surajit, C., & Manojit, C. (2007). *A Soft Computing technique in rainfall forecasting*. International Conference on IT, HIT, March 19-21, 2007.

Varotsos, C. (2005). Power-law correlations in column ozone over Antarctica. *Intern. J. Remote Sensing,* (3333-3342), 26.

Varotsos, C., & Krik-Davidoff, D. (2006). Long-memory processes in ozone and temperature variations at the region 60 degrees S-60 degrees N. *Atmos. Chem. Phys., 6*, 4093-4100.

Varotsos, C., Kondratyev, K. Y., & Efstathiou, M. (2001). On the seasonal variation of the surface ozone in Athens, Greece. *Atmos. Environ., 35*(2), 315-320

Wilks, D. S. (1991). Representing serial correlation of meteorological events and forecasts in dynamic decision-analytic models. *Month. Weather Rev., 119*(1640-1662).

Wong, K. W., Wong, P. M., Gedeon, T. D., & Fung, C. C. (2003). Rainfall prediction model using soft computing technique. *Soft Computing, 7*(6), 434-438.