



NUMERICAL AND EXPERIMENTAL INVESTIGATION OF METEOROLOGICAL DATA USING ADAPTIVE LINEAR M5 MODEL TREE FOR THE PREDICTION OF RAINFALL

 **Sheikh Amir Fayaz**¹
 **Majid Zaman**²⁺
 **Muheet Ahmed Butt**³

^{1,2}Department of Computer Science, University of Kashmir, J&K, India.

¹Email: skh.amir88@gmail.com

³Email: ermuheet@gmail.com

²Directorate of IT & SS, University of Kashmir, J&K, India.

³Email: zamanmajid@gmail.com



(+ Corresponding author)

ABSTRACT

Article History

Received: 7 February 2022

Revised: 10 March 2022

Accepted: 25 March 2022

Published: 13 April 2022

Keywords

Linear regression
Meteorological data
M5 model tree
Smoothing
Splitting nodes
Linear model functions.

Predicting a class with a continuous numeric value encounters many problems when applying machine learning to the data. Only a few machine-learning techniques can do this, but it is still considered one of the most complex tasks to perform. In this study, we demonstrate one of the techniques called the M5 Model Tree, which can handle continuous numeric data. This technique is a stepwise algorithm and uses linear functions at the leaf nodes of any decision tree inducer (like CART) constructed. These M5 model trees generate simple practical formulas like standard deviation (SD), standard deviation reduction (SDR), cost-complexity pruning (CCP), etc., which can be easily applied by another user to some other benchmark data. This work assesses the abilities of the M5 Model Tree algorithm for the assessment of rainfall data across the Kashmir province of the Union Territory of Jammu & Kashmir, India. The construction of the M5 model tree developed using (70–30) % training and test ratio, respectively, was considered one of the best fit models, predicting an RMSE of 2.593, an MAE of 1.68, and a correlation coefficient (R^2) of 0.478. Moreover, M5 model trees use a small number of trails to develop the models and thus need less computational time and are therefore more convenient to use.

Contribution/Originality: This research focuses on generating rules from continuous data numeric predictions using a heuristic approach technique (M5 Model Tree) to generate a sequence of linear model functions at each leaf node of the tree.

1. INTRODUCTION

Real-time predictions are always important for accurate and systematic thinking in planning future processes. The failure in the availability of current machine learning approaches is a matter of concern for making accurate and timely predictions. These early predictions, in any field like agriculture, meteorology, medical field, etc., play a very important role for farmers, doctors, etc. [1]. Farmers fear that too much or too little rainfall could ruin agriculture. It has hampered agricultural productivity in several arid and semiarid locations around the world, and it has become one of the most critical concerns facing human life. For determining future predictions of rainfall, historical-geographical parameters are used, which include vapor pressure, wind speed, humidity, temperature, density, precipitation, etc. Various soft computing models have been used for the prediction of rainfall [2]. In particular, efficient techniques using Artificial Neural Networks (ANN) and hybrid models have been recently used in the prediction of rainfall. These methods perform accurate and timely results, but one of the biggest disadvantages of

using these models is their complex architectures. These models provide their results using a "black box" approach, where the user can only analyze the input and output values without knowing the internal workings of the model. In this study, we have assessed the performance of an efficient M5 model tree algorithm on the rainfall data of Kashmir province, which is a stepwise algorithm and uses linear functions at the leaf nodes of any decision tree inducer (CART) constructed [3]. These model trees generate simple practical formulas like standard deviation (SD), standard deviation reduction (SDR), Cost-Complexity Pruning (CCP), etc., which can be easily applied by another user to some other benchmark data. Moreover, model trees use a small number of trails to develop the models and thus need less computational time and are therefore more convenient to use [4, 5]. This paper is structured as: A brief overview of Model tree, M5 Model trees is provided in section 1.A. Brief review of literature is described in section 2. A. Description of the material and dataset is provided in section 3, while in section 4 Implementation of M5 Model trees and brief analysis of the results are discussed along its individual prediction of each attribute. Finally, section 5 concludes the paper.

A. M5 Model Tree

A Model Tree is a machine learning strategy that works with continuous goal values that are numeric in nature, and the M5 Model Tree is the learning algorithm that can handle such values. Quinlan [6] first proposed this method Quinlan [6] and the M5 model tree algorithm's main role is to combine any classic decision tree inducer with the likelihood of linear regression at the created decision tree's leaf nodes. Because the decision tree appears to be a straightforward technique, the regression function only involves a few variables.

Construction of the M5 model tree involves two steps: Conventional Decision Tree and Linear Regression Function. Firstly, the decision tree induction algorithm is applied for building the regression tree. For the splitting criterion, the standard deviation at each node will be calculated to check the expected reduction in error. This splitting of nodes in M5 will continue until the numbers of instances left are very few. Secondly, after the construction of a normal regression tree, pruning of internal sub-nodes is done and is replaced with the regression plane instead of constant values, with the pruning expected error at each node estimated [7].

In splitting, a set say T is taken as input and is split into a simpler subset say T1, T2, T3 ... Tn. This approach is recursive in which each sub-split is sub further split into children and the process continues until the number of left instances is very less. This M5 follows a greedy approach for minimization of errors at each internal node in which Standard Deviation Reduction is calculated one node at a time and is given by (1). In the case of the M5 Model Tree, a Standard Deviation Reduction (SDR) is calculated at each node for splitting, and then Cost Complexity Pruning (CCP), shown below (2), is used at each leaf node to remove sections that may not be useful for the final tree, reducing the number of rules [8, 9].

$$SDR = \frac{SD(T) - \sum SD(T_i)|T_i|}{|T|} \quad (1)$$

$$CCP = \frac{err(prune(T, t), S) - err(T, S)}{||leaves(T)| - |leaves(prune(T, t))|} \quad (2)$$

Where, $err(T, S)$ is the error rate of tree T over dataset S, and $(prune(T, t), S)$ is the tree obtained by pruning the sub trees 't' after the regression is applied on the tree T.

Furthermore, after applying pruning on the generated Model Tree, smoothing is also applied which is used to avoid the sharp discontinuities of the sub trees. It is commonly used to improve prediction accuracy by flattening the sharp nodes of neighboring models [10].

The code snippet algorithm (Pseudo code) of the M5 model tree is shown below where the two main stages: split (Algorithm 1) and pruning (Algorithm 2), are defined [11, 12].

```
ConstructM5MT (Total Instances)
    {Std_Dev = s_d (Total Instances)
```

```

For each k-values nominal attribute
convert into k – 1 synthetic binary attributes
Root node. Total Instances = instances
Split (root node)
Prune (root node)
Print (root node)}
Split (node)
  {IF size of (node. Total Instances) < 4 or s_d (node. Total Instances) < 0.05 * Std_Dev
  Node. Type = LEAF_NODE
  ... .. . }

```

Algorithm 1: Code snippet for M5 model tree Algorithm with the Split function

```

ELSE
{Node. Type = Internal
For each attribute
FOR all possible split positions of the attribute, calculate the attribute's SDR
Node. Attribute = attribute with maximum SDR
Split (node. Left)
Split (node. Right)
Prune (node)
{IF node = INTERIOR then
Prune (node. Left Child)
Prune (node. Right Child)
Node. Model = linear Regression (node)
IF sub tree Error (node) > error (node) then
Node. Type = LEAF}
Sub tree Error (node)
{l = node. Left;
r = node. Right
IF node = INTERIOR then return (size of (l. instances) * sub tree Error (l) * size of (r. instances) * sub tree Error
(r)) / size of (node. Instances)
ELSE return error (node)}

```

Algorithm 2: Code snippet for M5 model tree Algorithm with prune & Error function.

2. REVIEW OF LITERATURE

Since the model tree is not a very popular method like other machine learning models. Thus, it has not been extensively used and a very less amount of work has been performed on M5 model trees in the field of geographical sciences. During the literature survey, we came across some of the research that is based on M5 model trees in geographical sciences where flood forecasting, estimating reference evapotranspiration, prediction of significant wave height in lake superior, rainfall-runoff modeling, stream flow forecasting like applications are performed.

Dimitri and Xue [13] investigated an M5 model for flood forecasting problems for the upper reaches of the Huai River in China. This study performs a head-to-head comparison between M5 Model Trees and ANN, where both models predict high floods with the same accuracy rate. Later stage, a modular model has been developed which consists of hybridization of the M5 and ANN model and it results in the best prediction performance. Since this model, works on the correlation analysis and thus it can be improved by adding the hydrological knowledge to refine the inputs from the hydrological experts. Thus, the specified inputs will produce more hydrological characteristics

using the M5 model tree for classification purposes. In Armin, et al. [14] based on the satellite images used to determine the reference Evapotranspiration [ETo] estimation. This study was performed on five weather stations in Iran. A Simple Linear Regression with an M5 model tree was developed for the estimation and it performs better on the same set of data.

In Etemad-Shahidi and Mahjoobi [15] proposed study where authors try to predict the significant wave height in Lake Superior. This study uses wind and wave data from Lake Superior from 2000 to 2001. The authors prefer M5 model trees over ANN because the M5 model tree represents some understandable rules that can be expressed by humans easily. Furthermore, the results indicate that error statistics of the M5 model tree were similar to that of ANN but it was observed that model trees were marginally more accurate.

Nourani, et al. [4] proposed a study for rainfall-runoff modeling based on the wavelet-M5 model tree. The dataset used in this study was divided into 3 different training and testing partitions (60-40%, 75-25%, 50-50%) respectively, and the model was implemented on both daily and monthly rainfall scales. This study concluded that the wavelet M5 model tree performs better in both monthly and daily rainfall scales. Fayaz, et al. [16] proposed a stepwise mathematical implementation of a Logistic Model Tree (LMT), and the data used in this study was collected from the Indian Metrological Department (IMD) Pune from the year 2012 to 2017 and it comprises around 5580 records. The data comprises five parameters including humidity and temperature as independent parameters and rainfall as the target parameter. The data used in the implementation process was discrete data and the target value of the data depicts the presence and the absence of rainfall. The accuracy of statistics and the various statistical measures were calculated based on the model and it was later compared with the other traditional and ensemble models like Random Forests (RF), Decision Tree (DT), and Distributed Decision Trees (DDT) and it was observed that the logistic model tree shows tremendous performance over these traditional and ensemble approaches. Some of the above literature studies of M5 model trees used in the geographical sciences are presented and as such, there is no study where the M5 mode tree has been used for the prediction of rainfall. Thus, this motivates us to perform an experimental analysis of the M5 model tree on the rainfall data and check how far the model is accurate and feasible [17, 18]. This paper will help us to analyze why this algorithm has not been widely used and it helps us to check the performance of the M5 model tree on other kinds of data like agricultural data, health data, academic data, etc.

3. MATERIAL AND DATASET

The description of the data gathered and the numerous parameters for all zones of Kashmir province, which are used for rainfall prediction [19-24] are shown in Table 1.

Table 1. Dataset description of various parameters.

Attribute	Station Number	Year	Station Name	Station Location	Attribute Measurement	Attribute Type
Humidity 12 A.M	42026, 42027, 42044	2012-2017	South, North & Central Kashmir	33.59°N 75.16°E, 34.05°N 74.38°E, 34.5°N 74.47°E	Percentage of relative Humidity, %	Continuous
Humidity 3 P.M	42026, 42027, 42044	2012-2017	South, North & Central Kashmir	Same Location	Percentage of relative Humidity, %	Continuous
Max_Temperature	42026, 42027, 42044	2012-2017	South, North & Central Kashmir	Same Location	°C	Continuous
Min_Temperature	42026, 42027, 42044	2012-2017	South, North & Central Kashmir	Same Location	°C	Continuous

After the data has been cleaned and preprocessed, the final instances of the dataset contain around 5580 entries. The snapshot of the final dataset is shown below in Table 2. [22, 25, 26].

Table 2 shows the processed and cleaned continuous Historical Geographical dataset of Kashmir Province.

Table 2. Historical geographical data of Kashmir province.

Max.	Min.	Hum12	Hum3	Rf
16	5.2	98	96	0
16.5	7	91	46	0
17.4	5.8	86	35	4.4
13	6	85	96	2
7.4	2.6	81	77	7.2
12.2	5	86	96	0
14	1.6	97	46	9.6
13.6	5.2	98	35	0
18.2	7.8	74	96	3.5
19.4	8	97	77	0
20.2	11	97	96	2.4
21	8	97	89	0.8
19	10	84	96	0
21	11.2	98	46	0
22.2	10.6	42	35	1.5
23.5	11.6	87	96	0.9
23.4	12.2	82	77	0
21.4	7.3	78	96	18.2
9	4.5	86	96	62
11.2	3	78	96	6.6
11	1.8	96	87	1.6
13.2	4.2	46	96	1.3
17.4	8	35	96	0.5
20.8	10	96	46	0
23	11.5	97	35	0
21.8	7	97	96	10.7

Below Table 3 shows the various statistical properties of the historical-geographical parameters used in the study.

Table 3. Statistical properties of data.

Attributes	Min.	Max.	Mean	Std. Dev	Variance	Skewness	Kurtosis
Max Temp (°C)	-7.6	35.4	18.0	8.80	77.49	-0.24	-0.86
Min Temp (°C)	-14.4	23.8	6.34	7.43	55.27	0.02	-0.84
Humid12 (%)	18	98	60.3	18.0	325.8	0.21	-0.73
Humid3 (%)	0	96	75.6	14.1	199.8	-0.76	0.40
Rf (mm)	0	206	2.76	9.07	82.25	7.75	99.4

The graphical (Figure 1, Figure 2) and tabular Table 4 representations provided below shows the various relative values of the historical-geographical parameters used in the study.

4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

In this study, we applied the M5 Model Tree algorithm to the 5 metrological parameters of the dataset. For the simulation study, an open-source data analytics tool called KNIME has been used. The experiment was carried out on the 70-30 ratio in which 70% was used as the training set and 30% was used for testing purposes. The dataset consists of 4 independent continuous parameters which include Minimum and Maximum Temperatures, Humidity at two different intervals, and one dependent variable rainfall with continuous values.

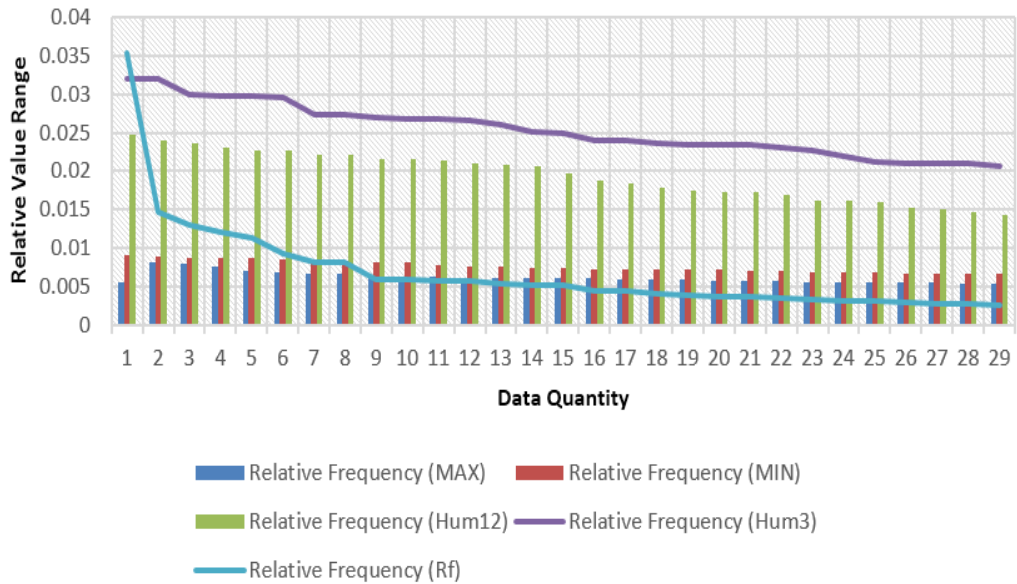


Figure 1. Graphical representations of relative values of geographical data.

Table 4. Tabular representation with relative values of the data.

Relative Frequency (Max.)	Relative Frequency (Min.)	Relative Frequency (Hum12)	Relative Frequency (Hum3)	Relative Frequency (Rf)
0.005562739	0.009	0.024	0.032	0.035
0.008195229	0.009	0.024	0.032	0.014
0.008013112	0.009	0.023	0.029	0.012
0.00764888	0.009	0.023	0.029	0.012
0.007102531	0.009	0.022	0.029	0.011
0.006920415	0.008	0.022	0.029	0.009
0.006738299	0.008	0.022	0.027	0.008
0.006738299	0.008	0.022	0.027	0.008
0.006556183	0.008	0.021	0.026	0.006
0.006374067	0.008	0.021	0.026	0.006
0.006374067	0.008	0.021	0.026	0.005
0.00619195	0.008	0.020	0.026	0.005
0.00619195	0.008	0.020	0.026	0.005
0.00619195	0.008	0.020	0.025	0.005
0.00619195	0.007	0.019	0.024	0.005
0.00619195	0.007	0.018	0.024	0.004
0.006009834	0.007	0.018	0.024	0.004
0.006009834	0.007	0.017	0.023	0.004
0.006009834	0.007	0.017	0.023	0.003
0.005827718	0.007	0.017	0.023	0.003
0.005827718	0.007	0.017	0.023	0.003
0.005827718	0.007	0.016	0.023	0.003
0.005645602	0.006	0.016	0.022	0.003
0.005645602	0.006	0.016	0.021	0.003
0.005645602	0.006	0.016	0.021	0.003
0.005645602	0.006	0.015	0.020	0.002
0.005645602	0.007	0.015	0.020	0.002
0.005463486	0.007	0.014	0.020	0.002
0.005463486	0.007	0.014	0.020	0.002

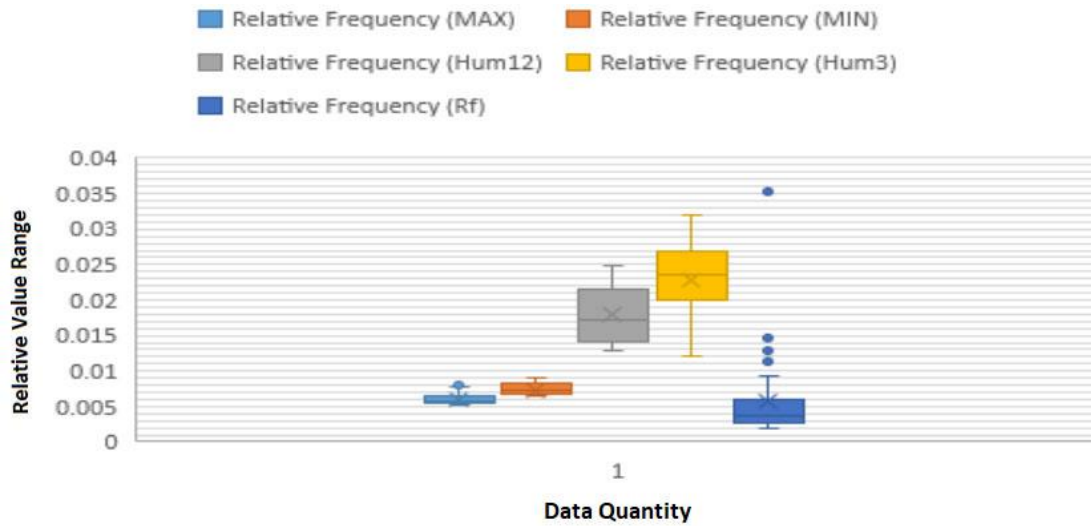


Figure 2. Q-plots of the relative values of the data.

Table 5. M5 base model assessment to check the number of rules.

M5Base Model				
M5 Pruned Model Tree			M5 Unpruned Model Tree	
Smoothed Linear Model	Un-Smoothed	Linear	Smoothed Linear Model	Un-Smoothed Linear Model
Number of Rules: 13	Number of Rules:13		Number of Rules:918	Number of Rules:918
M5 Rules with M5 Pruned Model Rules Using Smoothed Linear Model				
M5 Pruned Model Rules			M5 Unpruned Model Rules	
Smoothed Linear Model	Un-Smoothed	Linear	Smoothed Linear Model	Un-Smoothed Linear Model
Number of Rules: 10	Number of Rules: 10		Number of Rules: 111	Number of Rules: 111
M5 Rules with M5 Pruned Model Tree Using Un-Smoothed Linear Model				
M5 Pruned Model Rules			M5 Unpruned Model Rules	
Smoothed Linear Models	Un-Smoothed	Linear	Smoothed Linear Models	Un-Smoothed Linear Model
Number of Rules: 16	Number of Rules: 16		Number of Rules: 100	Number of Rules: 100

After the data was evaluated, various approaches were applied to the data to check the overall calculation of the data. It was observed that the maximum number of rules generated using the Unsmoothed Unpruned M5 model tree was around 918 and when the pruning was applied and using the smoothed linear model the overall rules drastically decreases to 13 without much effect on the performance as shown in Table 5 above. The clustered chart to compare the number of rules in pruned and un-pruned M5 linear models is shown in Figure 3.

Figure 4 shows the structure of the model tree using the M5 base algorithm with only 13 rules (LM_1 to LM_13) as their leaves for the metrological dataset of Kashmir Province.

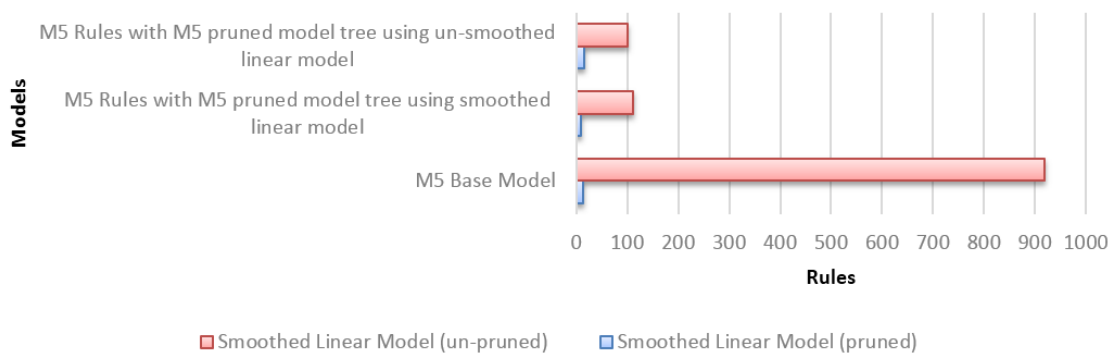


Figure 3. clustered chart to compare the number of rules.

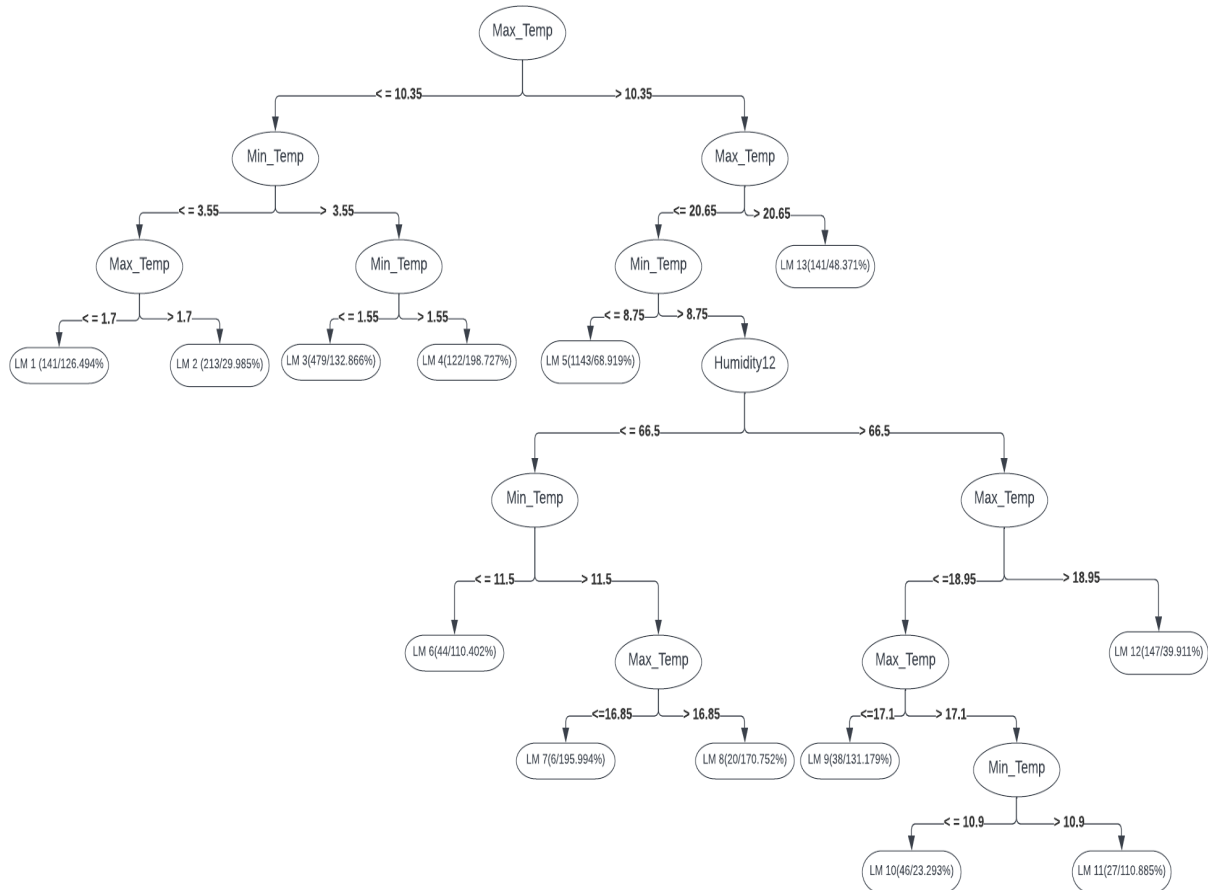


Figure 4. M5 model tree on historical geographical data of Kashmir province.

Table 6 Linear Model functions with Smoothed Linear models Generated Using M5Base Model Tree for rainfall.

Table 6. linear model functions with smoothed linear models generated using m5base model tree.

Linear Models	Rainfall
LM_1	$-1.36 * \text{Max_Temp} + 1.98 * \text{Min_Temp} + 0.02 * \text{Humidity12} - 0.03 * \text{Humidity3} + 22.1$
LM_2	$-0.31 * \text{Max_Temp} + 0.15 * \text{Min_Temp} + 0.02 * \text{Humidity12} - 0.03 * \text{Humidity3} + 4.58$
LM_3	$-1.90 * \text{Max_Temp} + 0.99 * \text{Min_Temp} + 0.19 * \text{Humidity12} - 0.47 * \text{Humidity3} + 44.8$
LM_4	$-2.02 * \text{Max_Temp} + 6.07 * \text{Min_Temp} + 1.47 * \text{Humidity12} - 2.02 * \text{Humidity3} + 74.5$
LM_5	$-0.25 * \text{Max_Temp} + 0.35 * \text{Min_Temp} + 0.01 * \text{Humidity12} - 0.01 * \text{Humidity3} + 4.68$
LM_6	$-1.40 * \text{Max_Temp} + 0.85 * \text{Min_Temp} + 0.33 * \text{Humidity12} - 0.53 * \text{Humidity3} + 45.8$
LM_7	$-2.46 * \text{Max_Temp} + 2.23 * \text{Min_Temp} + 0.46 * \text{Humidity12} - 0.70 * \text{Humidity3} + 60.4$
LM_8	$-1.86 * \text{Max_Temp} + 2.77 * \text{Min_Temp} + 0.46 * \text{Humidity12} - 0.70 * \text{Humidity3} + 38.1$
LM_9	$-2.79 * \text{Max_Temp} + 0.57 * \text{Min_Temp} + 0.13 * \text{Humidity12} - 0.37 * \text{Humidity3} + 65.9$
LM_10	$-0.54 * \text{Max_Temp} + 0.71 * \text{Min_Temp} + 0.22 * \text{Humidity12} - 0.64 * \text{Humidity3} + 44.9$
LM_11	$-0.54 * \text{Max_Temp} + 0.84 * \text{Min_Temp} + 0.27 * \text{Humidity12} - 0.81 * \text{Humidity3} + 56.4$
LM_12	$0.59 * \text{Max_Temp} + 0.19 * \text{Min_Temp} + 0.045 * \text{Humidity12} - 0.14 * \text{Humidity3} - 3.49$
LM_13	$-0.21 * \text{Max_Temp} + 0.18 * \text{Min_Temp} - 0.09 * \text{Humidity12} - 0.02 * \text{Humidity3} + 5.70$

Note: "*" indicates "multiply (x)".

The overall statistical properties of the dataset used in this study is shown Figure 5.

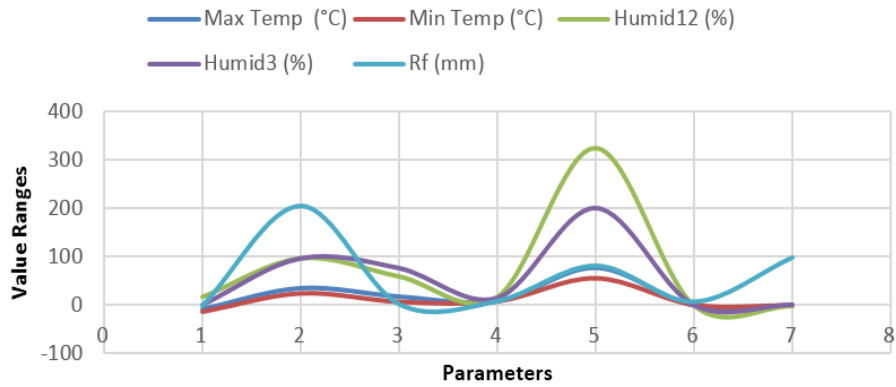


Figure 5. Statistical properties of the dataset.

The snapshot of the rules generated by the M5 Model tree algorithm is shown in Figure 6:

```

Max_Temp <= 10.35 :
|  Min_Temp <= -3.55 :
|  |  Max_Temp <= 1.7 : LM1 (141/126.494%)
|  |  Max_Temp > 1.7 : LM2 (213/29.985%)
|  Min_Temp > -3.55 :
|  |  Min_Temp <= 1.55 : LM3 (479/132.866%)
|  |  Min_Temp > 1.55 : LM4 (122/198.727%)
Max_Temp > 10.35 :
|  Max_Temp <= 20.65 :
|  |  Min_Temp <= 8.75 : LM5 (1143/68.919%)
|  |  Min_Temp > 8.75 :
|  |  |  Humidity12 <= 66.5 :
|  |  |  |  Min_Temp <= 11.5 : LM6 (44/110.402%)
|  |  |  |  Min_Temp > 11.5 :
|  |  |  |  |  Max_Temp <= 16.85 : LM7 (6/195.994%)
|  |  |  |  |  Max_Temp > 16.85 : LM8 (20/170.752%)
|  |  |  Humidity12 > 66.5 :
|  |  |  |  Max_Temp <= 18.95 :
|  |  |  |  |  Max_Temp <= 17.1 : LM9 (38/131.179%)
|  |  |  |  |  Max_Temp > 17.1 :
|  |  |  |  |  |  Min_Temp <= 10.9 : LM10 (46/23.293%)
|  |  |  |  |  |  Min_Temp > 10.9 : LM11 (27/110.885%)
|  |  |  |  Max_Temp > 18.95 : LM12 (147/39.911%)
|  Max_Temp > 20.65 : LM13 (1417/48.371%)

```

Figure 6. Rules generated using m5base model tree algorithm on historical geographical data of Kashmir province.

In this study, our main aim was to analyze the data using the M5 model tree algorithm and it was observed that the continuous target values can be analyzed effectively as compared to the normal regression. It is because the size of the M5 model generated is very less when the regression tree is taken into consideration. The number of rules in the case of the normal regression would be equal to the total number of instances of the target variable but in the case of M5 model trees number of rules are countable [27-29].

M5 model tree has been applied to the same Historical-Geographical data for the prediction purposes of the rainfall and the correlation coefficient is calculated as shown in Figure 7:

R ² :	0.478
Mean absolute error:	1.689
Mean squared error:	6.726
Root mean squared error:	2.593
Mean signed difference:	0.844

Figure 7. Accuracy statistics.

Below are the snapshots in Figure 8 which check the individual attribute actual and predicted values obtained from the M5 model.

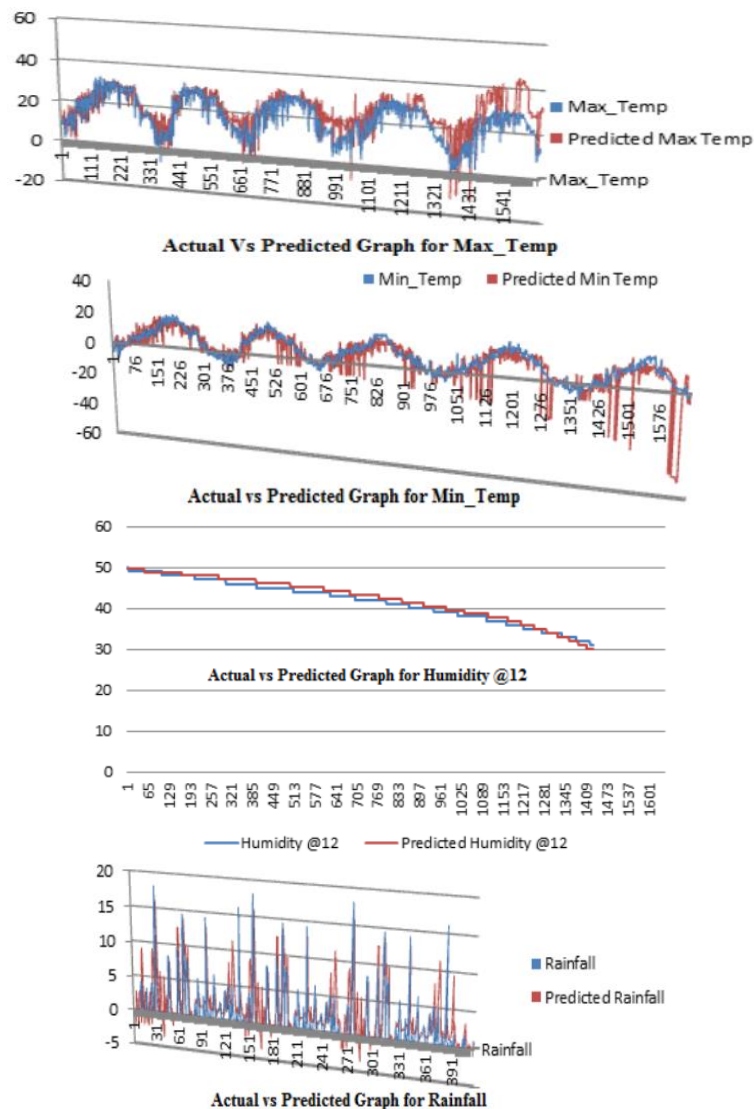


Figure 8. Actual vs Predicted graphs for the attributes of geographical data of Kashmir province using M5 model tree algorithm.

5. CONCLUSION AND FUTURE SCOPE

This study mainly focuses on the generation of rules from the numeric prediction of the continuous data by introducing the heuristic approach technique to generate the sequence of linear model functions Table 4 at each leaf node of the tree. It defines how the classification and regression problems can be transformed into linear model

function approximations in a much more standard way. In this study, our main aim was to analyze the data using the M5 model tree algorithm and it was observed that the continuous target values can be analyzed effectively as compared to the normal regression models. It is because the size of the M5 model generated is very less when the regression tree is taken into consideration. The number of rules in the case of the normal regression would be equal to the total number of instances of the target variable but in the case of M5 model trees, a number of rules are countable. M5 model tree outperforms the normal regression tree by providing almost 5 times significantly fewer rules, thus reducing the size of the tree without overall affecting the performance & provides the ability to exploit the local linearity in the data. The pruned and smoothed results of the M5 model trees indicate a subsequent increase in the accuracy of predictions. However, smoothing in many cases may increase the complexity of the linear models and thus leave difficulties analyzing the prediction accuracies. The construction of the M5 model tree developed using (70-30) % training and test ratio respectively was considered one of the best fit models which predicts RMSE of 2.593, MAE of 1.68, and correlation coefficient (R^2) of 0.478. One of the major advantages of M5 model trees over normal regression trees is that normal regression trees can never predict the values outside the range of the trained model but the same is not the case with M5 model trees as they can extrapolate. As of now, no such technique has been found that can simplify the heuristic functions at the leaf nodes which always remain a tradeoff between the pruning factor and the size of the tree and it needs to be further rectified. Furthermore, the computational complexity of the M5 un-pruned model tree needs to be investigated. In this study, we considered the Multiple-Input and Single-Output (MISO) models. In the future, we can also check the performance based on the Multiple-Inputs and Multiple-Outputs (MIMO) model.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: All authors contributed equally to the conception and design of the study.

REFERENCES

- [1] M. Rezaie-balf, S. R. Naganna, A. Ghaemi, and P. C. Deka, "Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting," *Journal of hydrology*, vol. 553, pp. 356-373, 2017. Available at: <https://doi.org/10.1016/j.jhydrol.2017.08.006>.
- [2] R. M. Adnan, A. Petrocelli, S. Haddam, C. A. G. Santos, and O. Kisi, "Comparison of different methodologies for rainfall-runoff modeling: Machine learning vs conceptual approach," *Natural Hazards*, vol. 105, pp. 2987-3011, 2021. Available at: <https://doi.org/10.1007/s11069-020-04438-2>.
- [3] R. Bahmani, A. Solgi, and T. B. Ouarda, "Groundwater level simulation using gene expression programming and M5 model tree combined with wavelet transform," *Hydrological Sciences Journal*, vol. 65, pp. 1430-1442, 2020. Available at: <https://doi.org/10.1080/02626667.2020.1749762>.
- [4] V. Nourani, A. Davanlou Tajbakhsh, A. Molajou, and H. Gokcekus, "Hybrid wavelet-M5 model tree for rainfall-runoff modeling," *Journal of Hydrologic Engineering*, vol. 24, p. 04019012, 2019. Available at: [https://doi.org/10.1061/\(ASCE\)he.1943-5584.0001777](https://doi.org/10.1061/(ASCE)he.1943-5584.0001777).
- [5] Y. Z. Kaya, F. Uneş, M. Demirci, B. Taşar, and H. Varçin, "Groundwater level prediction using artificial neural network and M5 tree models," *Conference on Air and Water Components of the Environment*, pp. 195-201, 2018.
- [6] J. R. Quinlan, "Learning with continuous classes," presented at the 5th Australian Joint Conference on Artificial Intelligence, World Scientific, 1992.
- [7] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Machine Learning*, vol. 59, pp. 161-205, 2005.
- [8] E. Frank, Y. Wang, S. Inglis, G. Holmes, and I. H. Witten, "Using model trees for classification," *Machine Learning*, vol. 32, pp. 63-76, 1998.
- [9] K. Raza, "M5 model tree and gene expression programming for the prediction of metrological parameters," presented at the In 2015 International Conference on Computers, Communications, and Systems (ICCCS). IEEE, 2015.
- [10] M. Samadi, E. Jabbari, and H. M. Azamathulla, "Assessment of M5' model tree and classification and regression trees for prediction of scour depth below free overfall spillways," *Neural Computing and applications*, vol. 24, pp. 357-366, 2014. Available at: <https://doi.org/10.1007/s00521-012-1230-9>.
- [11] D. Malerba, A. Annalisa, B. Antonia, C. Michelangelo, and P. Domenico, "Stepwise induction of model trees," presented at the In Congress of the Italian Association for Artificial Intelligence, pp. 20-32. Springer, Berlin, Heidelberg, 2001.

- [12] D. Malerba, F. Esposito., M. Ceci., and A. Appice., "Top-down induction of model trees with regression and splitting nodes," presented at the IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004.
- [13] P. S. Dimitri and Y. Xue, "M5 model trees and neural networks: Application to flood forecasting in the upper reach of the Huai River in China," *Journal of Hydrologic Engineering*, vol. 9, pp. 491-501, 2004. Available at: [https://doi.org/10.1061/\(ASCE\)1084-0699\(2004\)9:6\(491\)](https://doi.org/10.1061/(ASCE)1084-0699(2004)9:6(491)).
- [14] A. Armin, Y. Jalal, and M. Maryam, "Comparative study of M5 model tree and artificial neural network in estimating reference evapotranspiration using MODIS products," *Journal of Climatology*, pp. 1-11, 2014. Available at: <https://doi.org/10.1155/2014/839205>.
- [15] A. Etemad-Shahidi and J. Mahjoobi, "Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior," *Ocean Engineering*, vol. 36, pp. 1175-1181, 2009. Available at: <https://doi.org/10.1016/j.oceaneng.2009.08.008>.
- [16] S. A. Fayaz, M. Zaman, and M. A. Butt, "An application of logistic model tree (LMT) algorithm to ameliorate Prediction accuracy of meteorological data," *International Journal of Advanced Technology and Engineering Exploration*, vol. 8, p. 1424, 2021. Available at: <https://doi.org/10.19101/ijatee.2021.874586>.
- [17] N. M. Mir, S. Khan., M. A. Butt., and M. Zaman., "An experimental evaluation of bayesian classifiers applied to intrusion detection," *Indian Journal of Science and Technology*, vol. 9, pp. 1-7, 2016.
- [18] R. Mohd, A. B. Muheet, and Z. B. Majid, "SALM-NARX: Self Adaptive LM-based NARX model for the prediction of rainfall," presented at the In 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC), 2018 2nd International Conference on. IEEE, 2018.
- [19] R. Mohd, M. A. Butt, and M. Z. Baba, "GWLM-NARX: Grey wolf levenberg-Marquardt-based neural network for rainfall prediction," *Data Technologies and Applications*, 2020.
- [20] S. A. Fayaz, Z. Majid, and A. B. Muheet, "To ameliorate classification accuracy using ensemble distributed decision tree (DDT) vote approach: An Empirical discourse of Geographical Data Mining," *Procedia Computer Science*, vol. 184, pp. 935-940, 2021.
- [21] Z. Majid, K. Sameer, and A. Muheet, "Analytical comparison between the information gain and Gini Index using Historical geographical data," *International Journal of Advanced Computer Science and Applications*, vol. 11, pp. 429-440, 2020.
- [22] S. J. Sidiq, M. Zaman, and M. Ahmed, "How machine learning is redefining geographical science: A review of literature," *International Journal of Emerging Technologies and Innovative Research*, vol. 6., pp. 1731-1746, 2019.
- [23] S. Kaul, S. A. Fayaz, M. Zaman, and M. A. Butt, "Is decision tree obsolete in its original form? A burning debate," *Artificial Intelligence Review*, vol. 36, pp. 105-113, 2022. Available at: <https://doi.org/10.18280/ria.360112>.
- [24] S. A. Fayaz, M. Zaman, and M. A. Butt, *Performance evaluation of GINI Index and Information gain criteria on geographical data: An empirical study based on JAVA and python. In: Khanna, A., Gupta, D., Bhattacharyya, S., Hassanien, A.E., Anand, S., Jaiswal, A. (eds) International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing*. Singapore: Springer, 2022.
- [25] S. A. Fayaz, Z. Majid, and A. B. Muheet, "Knowledge discovery in geographical sciences—a systematic survey of various machine learning algorithms for rainfall prediction," presented at the In International Conference on Innovative Computing and Communications. Springer, Singapore, 2022.
- [26] M. Zaman, S. M. K. Quadri, and A. B. Muheet, "Information translation: A practitioners approach," in *In Proceedings of the World Congress on Engineering and Computer Science*, 2012.
- [27] S. A. Fayaz, I. Altaf, A. N. Khan, and Z. H. Wani, "A possible solution to grid security issue using authentication: An overview," *International Journal of Web Engineering and Technology*, vol. 5, pp. 10-14, 2019.
- [28] M. Ashraf, M. A. Syed, A. G. Nazir, A. S. Riaz, Z. Majid, A. K. Sameer, and A. S. Aftab, "Prediction of cardiovascular disease through cutting-edge deep learning technologies: An empirical study based on TensorFlow, PyTorch, and Keras," presented at the In International Conference on Innovative Computing and Communications. Springer, Singapore, 2021.
- [29] M. Ashraf, M. Zaman., and M. Ahmed., "Performance analysis and different subject combinations: An empirical and analytical discourse of educational data mining," presented at the In 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2018.

Views and opinions expressed in this article are the views and opinions of the author(s). Review of Computer Engineering Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.