International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 9 ǁ 2021 ǁ PP. 110-113

Machine Learning Models for Rainfall Data

B. Hari Mallikarjuna Reddy[1], S. Venkatramana Reddy[2], B. Sarojamma[3*]
1&3: Department of Statistics, Sri Venkateswara University, Tirupati – 517 502, A.P., India.
2: Department of Physics, Sri Venkateswara University, Tirupati – 517 502, A.P., India.

Abstract
Rainfall plays a vital in India for cultivation and drinking purpose. 70% population in India is directly or indirectly depends on cultivation. In this paper, we are used three Machine Learning models such as Decision Stump, M5p and Multilayer Perceptron for Rainfall Data of Coastal Andhra Pradesh, Rayalaseema and Telangana from the year 1988 to 2017. By Measuring Accuracy using Accuracy models such as Mean Absolute Error, Root Mean Square Error, Relative Absolute Error and Root Relative Absolute Error. We choose the best model among these three models

Keywords: *Rainfall, Decision Stump, M5p, Multilayer Perceptron, RMSE.*

Date of Submission: 05-09-2021 Date of acceptance: 18-09-2021

I. Introduction

In India, Cultivation is mainly based on rainfall and there are people nearly 70% are directly or indirectly depend on Agricultural Rainfall-Sector. The average annual rainfall is 300-650 millimeters i.e. 11.8 – 25.6 inches, but it is very difficult to get maximum rainfall in many areas. Rainfall plays major role for plants and soil. In time of drought, it reduces pollution in air and rain water is eco-friendly. Water fall in many ways to earth i.e. hail, rain, snow or sleet. The main source of water for human being is digestive purpose.

Dimitri P. Solomatire & Khada N. Dulal [1] gave Model Trees as an alternative to Neural Networks in Rainfall – Runoff modeling explains about comparative study of two data modeling techniques Artificial Neural Networks(ANN) and model trees(Mts) for rainfall runoff transformations. By taking rainfall data, they conclude that one hour ahead prediction is better than three hours ahead prediction and is better than six hour prediction, ANN is slightly better performance than MTs.

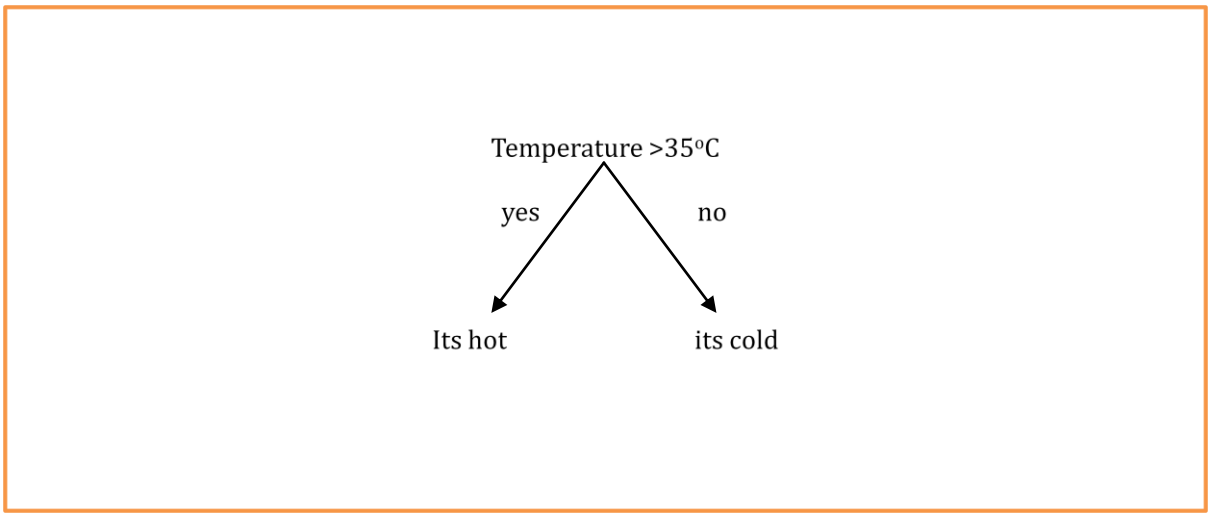
Ramesh et al [2] gave A Comparative analysis of Classification Algorithms on Weather Dataset using Data Mining tool describes Temperature data classified using Decision Trees(J48), REPTREE and Random Tree. These are compared using Root Mean Square Error criteria. Zealand et al [3] gave article on short term stream flow forecasting Artificial Neural Networks.

Eda Cinarogcu et al [4] explains a data mining Application of Local Weather Forecast for Kayseri Erkild Airport describes Meteorological Weather parameters such as fog, rain, snow, hail and thunderstorm using Data Mining Techniques. Chinnaiyam Ponnuraja et al [5] published a paper on Performance Accuracy between classifiers in sustain of disease conversion for clinical trials Tuberculosis Data using a Data Mining Approach regarding clinical Trials of Tuberculosis (TB) data classification using C4.5 Tree Classifier, ID_3 , Multilayer Perceptron and Naïve Bayes. Begum Cigsar and Deniz unal in their paper discussed about Comparison of Data Mining Classification Algorithms determining the Default Risk.

II. Methodology:

The atmospheric variable such as rainfall annual average took from the year 1988 to 2017 for Coastal Andhra Pradesh region, Rayalaseema region and Telangana region[7]. For Decision, we are using decision trees using WEKA. The Decision Trees we are using is Decision Stump, M5p – M4.0 and Multilayer Perceptron for each region.

Decision Stump: This is an Machine Learning model which consists of only one layer decision tree. This decision tree with one internal node which is connected with only one terminal nodes. The decision tree splits into binary features like ‘Yes’ or ‘No’.



M5p – M4.0: M5p is also called as M5 tree is a Learner for Regression problem. For terminal nodes, it assigns Linear Regression and fits a Multivariates Linear Regression model to each subspace by dividing the whole data into subdivisions. M5 Tree, generally deals with continuous class problems, then the discrete class problems. For each sub division, it gives linear models constructed to approximate non-linear relationships of data set.
Multilayer Perceptron(MLP): Feed Forward Artificial Neural Network(ANN) branch is Multilayer Perceptron. Sometimes Multilayer Perceptron is used in place of Feed Forward Artificial Neural Networks and sometimes they are strictly refers to networks. MLP contains more than two layers and also hidden layers.
Correlation Coefficient: It gives the numerical measure value of independent and dependent variables.
Mean Absolute Error: Difference between original and expected values is Error. Average of all Absolute errors is Mean Absolute Error.
Root Mean Square Error: Arithmetic Mean of the square of a set of numbers to errors.
Relative Absolute Error: Relative Absolute error is expressed as a ratio of mean error to errors produced by Naïve model.

III. Empirical Investigations:

For rainfall annual average data from 1988 to 2017 to Rayalaseema, Coastal Andhra Pradesh and Telangana, we are fitted Decision Stump, M5p and Multilayer Perceptron Machine Learning models using WEKA software as explained below. Figures shows that the outputs of Coastal Andhra Pradesh using Machine Learning Methods such as Decision Stump, M5p and Multilayer Perceptron. These figures contains measures of accuracy like MAE, RMSE, RAE, RRSE and also explains Correlation Coefficient between Rainfall annual average with Year wise data from 1988 to 2017.

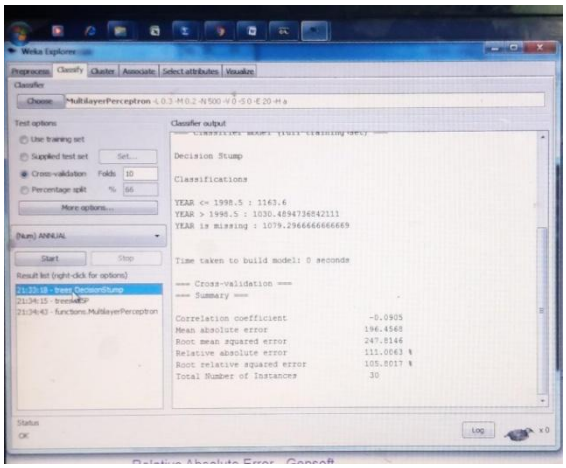


Fig.1: Decision Stump

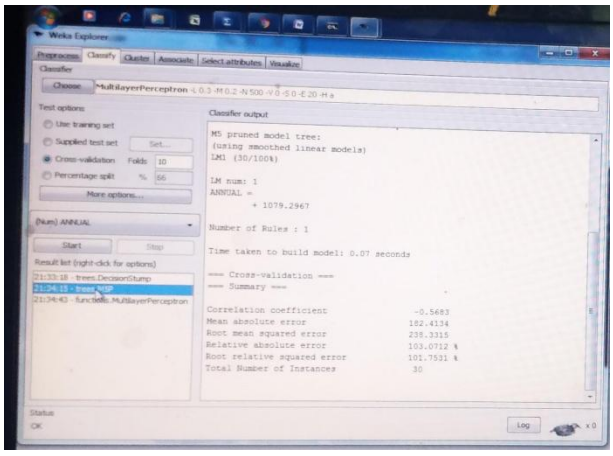


Fig.2: M5p

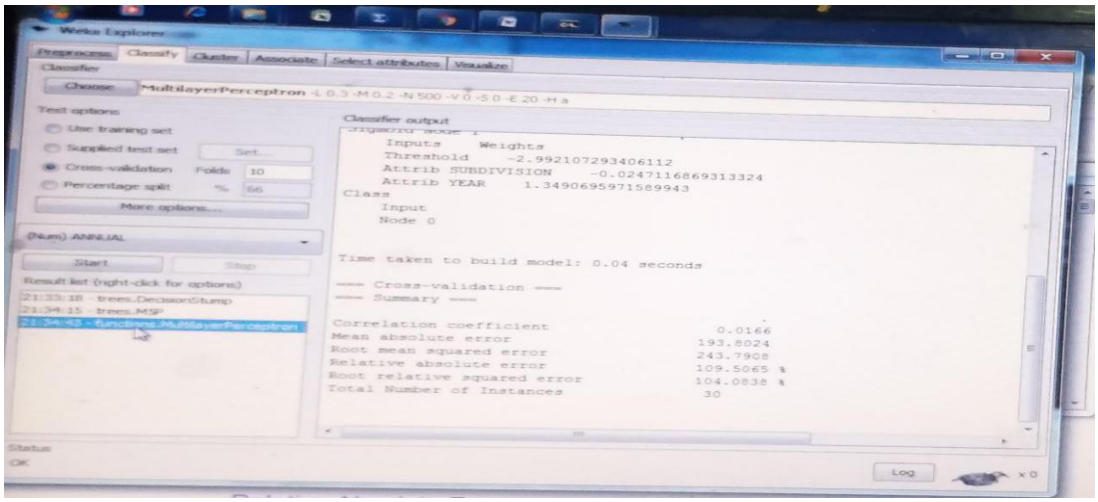


Fig.3: Multilayer Perceptron

Figures 4, 5 and 6 gives the output of Rayalaseema region by taking independent variables as year and region and dependent variables as average annual rainfall for Decision Stump, M5p and Multilayer Perceptron Machine Learning models. Best among three models is picked up using measures of accuracy MAE, RMSE, RAE, RRSE.

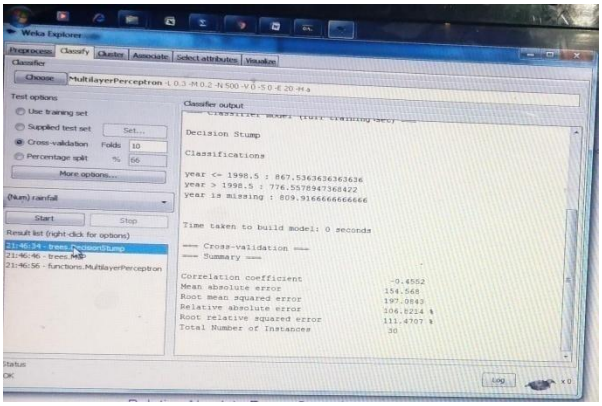


Fig.4: Decision Stump

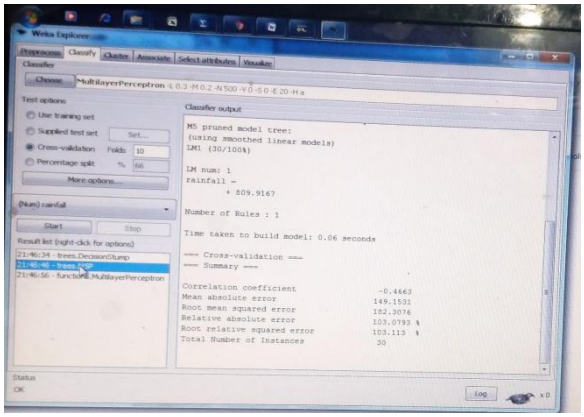


Fig.5: M5p

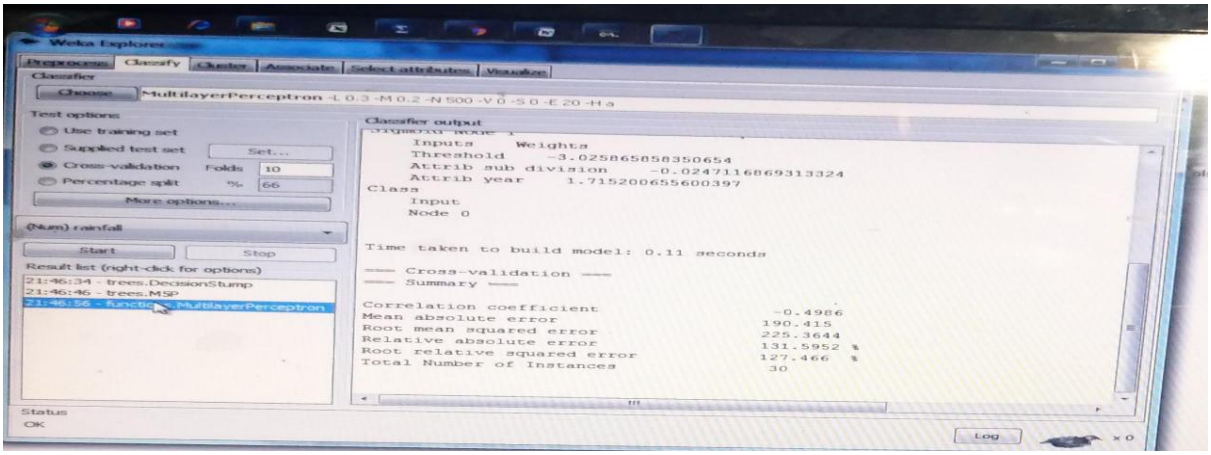


Fig.6: Multilayer Perceptron

Figures 7, 8 and 9explains Decision Stump, M5p and Multilayer Perceptron Machine Learning methods of average Rainfall to year wise data in Telangana Region. Best among three is selected using MAE, RMSE, RAE, RRSE measures.

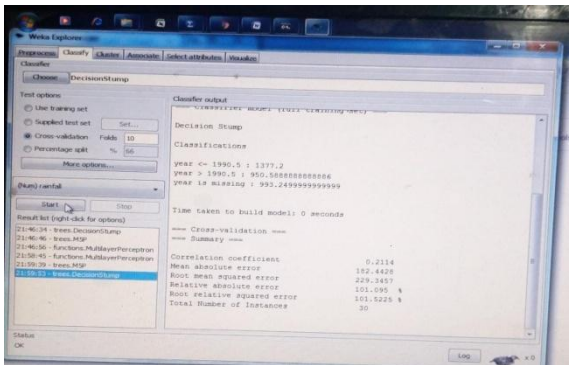


Fig.7: Decision Stump

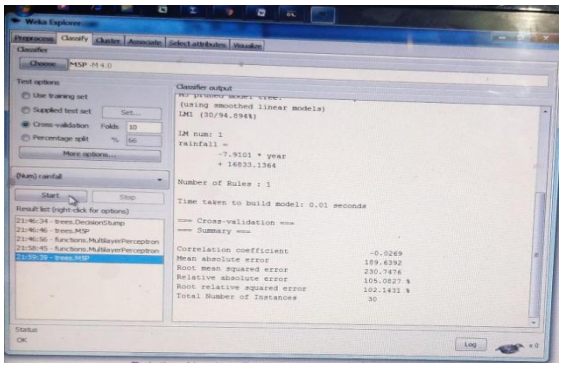


Fig.8 : M5p

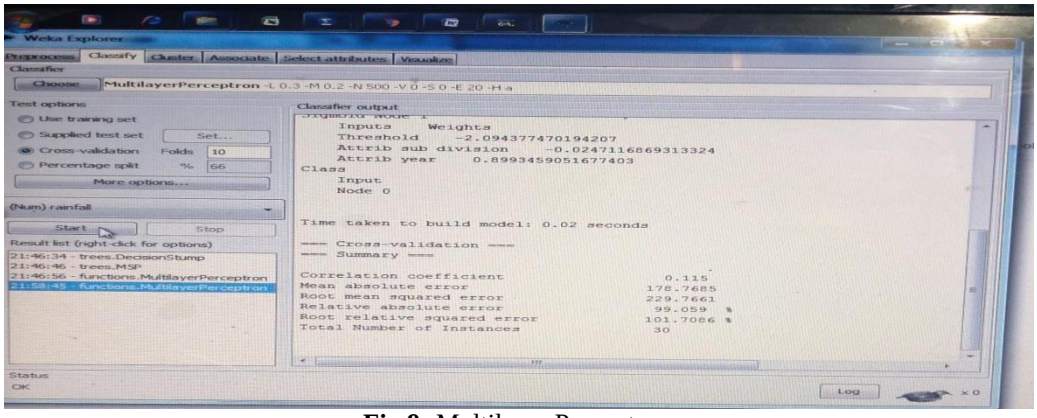


Fig.9: Multilayer Perceptron

IV. Summary and Conclusions:

There are many Machine Learning models available in the Literature but we chosen Decision Stump, M5p and Multilayer Perceptron for Sub divisions of Coastal Andhra Pradesh, Telangana and Rayalaseema regions annual rainfall from 1988 to 2017. The best Machine Learning model is chosen by means of Mean Absolute Error, Root Mean Squares Error, Relative Absolute Error, Root Relative Square Error as Measures of Accuracy.

Model RMSE values

Sub division	Decision Stump	M5p	MLP
Coastal AP	247.8146	238.3315	243.7908
Rayalaseema	197.0843	182.3076	225.3644
Telangana	229.3457	229.7661	230.7476

For estimation of Coastal Andhra Pradesh and Rayalaseema, M5p tree is the best, where as in Telangana all most all models i.e Decision Stump, M5p and Multilayer Perceptron are showing similar approximate results but a little bit Decision Stump is better.

References:

[1]. Dimitri P. Solomatire & Khada N. Dulal, Model trees as an alternative to Neural networks in Rainfall Runoff modeling, Hydrological Sciences Journal,48(3) (2010) 399-411.

[2]. D. Ramesh, Syed Nawaz Pasha and G. Roopa, A Comparative Analysis of Classification Algorithms on Weather Dataset using Data Mining tool, Oriental Journal of Computer Science and Technology, 10(4)(2017)1-6.

[3]. Zealand, C.M., Burn, D.H & Simonovic, S.P, Short Term Stream flow Forecasting using Artificial Neural Networks, Journal of Hydrol, 216(1999)32-55.

[4]. Eda Cinarogcu, Osman Unutulmoz, A Data Mining Application of Local Weather Forecast for Kayseri Erkild Airport, Journal of Polytechnic, 22(1)(2019) 103-113.

[5]. Chinnaiyam Ponnuraja, Babu C. Lakshmanan, Valarmathi Srinivasan, Performance Accuracy between classifiers in sustain of disease conversion for clinical trials Tuberculosis Data: Data Mining Approach, IOSR journal of Dental and Medical Sciences(IOSR-JDMS), 15(1)(2016)105-111.

[6]. Begum Cigsar and Deniz unal, Comparison of Data Mining Classification Algorithms determining the Default Risk, Hindowi, Scientific programming,12(2019)1-8.

[7]. Data website: URL:http://www.imd.gov.in/Welcome%20To%20IMD/Welcome.php