

# Rainfall Prediction Using Machine Learning And Tableau Visualisation

Mr. Muppuru Bhargavaram, Ms. M. Meghana and Mr. Vamsi Sannareddy

School of Computer Science and Engineering,

Vellore Institute Of Technology, Vellore, Tamil Nadu – 632014 (India).

**Abstract**— Rainfall is the most crucial process of nature. All the living being rely on water and rainfall is a process that is responsible for the continual process of water cycle. Many farmers will depend upon the rainfall for cultivating many essential crops for us, government also looking ahead of prediction of rainfall which will be useful for the future purposes, many human activities like agriculture are dependent on rainfall, especially in a country like India. Thus, it is very important and necessary to predict the rainfall patterns to estimate the flooding and drowning events. Application of algorithms is the best way to forecast rainfall. These algorithms predict rainfall numerically.

**Keywords**— Linear Regression, Support Vector Regression, Multi-Layer Perceptron, Rainfall Prediction, South-East Coast, Machine Learning, Visualization, Mean Absolute Error.

## 1. INTRODUCTION

Knowledge of rainfall patterns help in drought and flood management. Considering suitable attributes for prediction of rainfall is not an easy task. The intensity and frequency estimation are very useful for an agriculture-based country like India, especially in the coastal regions in the states of Karnataka, Kerala, Andhra Pradesh, Orissa, and West Bengal. So, to correctly estimate the rainfall pattern

why not use the power of computer science and data mining.

Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It's used to predict values within a continuous range, (e.g., sales, price) rather than trying to classify them into categories.

Decision Tree is a process of obtaining data from a database with certain techniques and process. It consists of identifying the pattern in the given data, called training sample, or analysing a set of already classified objects, whose results can be used to predict the results of other data with missing attributes, called test sample.

It aims at accurate analysis of data and generation of precise results. This analysis and prediction of data in machine learning can be done by many methods like classification, clustering, regression. Rainfall prediction can be done by algorithms like, SVM, Logistic regression, Linear regression, decision tree as they provide a methodology to classify and predict data better than the traditional statistical techniques. These algorithms observe the relationship between the attributes in the data set. In this paper values for atmospheric variables like pressure, temperature, wind speed, wind direction, rainfall percentage and humidity were considered for 10 years.

Rainfall prediction was done by a classification method of Decision Tree

and Regression method of linear Regression, SVM and also the multi-layer perceptron algorithm for the prediction of the rainfall for the present and for the future days of the rainfall in the percentage format and had collected the API from the open weather website which yields to give the humidity and the windspeed and collects the present rainfall percentage in the particular place around the globe.

## **2.BACKGROUND: ALGORITHMS**

### **2.1 ALGORITHMS USED**

Rainfall Prediction is a very important task in which machine learning algorithms are used to predict the rainfall prediction in various algorithms like linear regression, SVM, multi-layer perceptron etc.

Rainfall Prediction is a navigation link where you can easily see the rainfall amount or weather of upcoming week of your region or of globally.

### **2.2 ALGORITHMS DESCRIPTION**

#### **2.2.1 LINEAR REGRESSION MODEL REPRESENTATION**

Linear Regression is an attractive model because the representation is so simple. The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

For example, in a simple linear regression problem (a single x and a single y), the form of the model would be

$$y = A_0 + A_1 * x$$

In higher dimensions when we have more than one input (x), the line is called a plane or a hyper-plane. The

representation therefore is the form of the equation and the specific values used for the coefficients (e.g., A<sub>0</sub> and A<sub>1</sub> in the above example).

#### **2.2.2 SVM MACHINE**

SVM is an attractive model because the representation is so simple.

The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

For example, in a simple SVM machine of w instance of the problem (a single x and a single y), the form of the model would be:

$$f(x) = (w \cdot x) + b \quad n \quad w \in R$$

#### **2.2.3 MULTILAYER PRECEPTRON NEURAL NETWORK**

One of the most extensively used neural network topologies is the multilayer perceptron (MLP). The back propagation approach is commonly used to train MLPs. The back propagation rule propagates faults throughout the network, allowing hidden processing parts to learn recursively.

The nonlinear processing elements (PEs) of the multilayer perceptron, which have a nonlinearity, and the multilayer perceptron's. The multilayer perceptron is taught via error corrective learning, which requires the system's desired response.

The following is how error corrective learning works:

For a given input pattern, an instantaneous error  $e_i(n)$  is defined by the system response at PE  $i$  at iteration  $n$ ,  $y_i(n)$ , and the desired response  $d_i(n)$ :

$$\text{Where } e_i(n) = d_i(n) - y_i(n)$$

Due to the large amount of data in rainfall, achieving highly efficiency is somewhat difficult for the machine learning algorithms like linear regression model, SVM model, Decision tree classifier, Self-Organising Map and multi-layer perceptron neural Algorithms there is high chance of risk in these types of machine learning algorithms

### 3. PROPOSED SYSTEM

#### 3.1 STRATEGY

➤ Collecting the data set from the Kaggle website & Arranging Data in an Accessible manner.

➤ Collecting the data from the Kaggle website for the past 10 years and we will collect the data set which includes the variables like wind speed, humidity, rainfall, temperature, etc. which helps to visualize and predict the data

➤ Visualizing the data set from the collected data set.

➤ Visualizing the data in the form of histograms, bar graphs, pie charts, density, etc. using Tableau and it will be useful for the prediction of the rainfall part.

➤ Prediction of the data set from the collected data set.

➤ Predict the data set by implementing the algorithms like linear regression, logistic regression, random forest, decision tree, entropy, etc. which helps in finding the rainfall percentage using python in Jupyter notebook

➤ Drawing conclusions from the Collected data set and Observations.

#### 3.2 DATASET USED:

The data set is from the www.kaggle.com website from the past 20 years data set which contains the

variables as wind speed, rainfall, humidity, temperature, wind direction etc which contains the data set of 4377 rows and 19 different columns

And also, the data set from the www.dataworld.com to visualize the data set in the Tableau which helps in the part of the prediction and used in the part of implementing the website of rainfall prediction.

### 4. IMPLEMENTATION

First, we implemented the visualization in the Tableau:

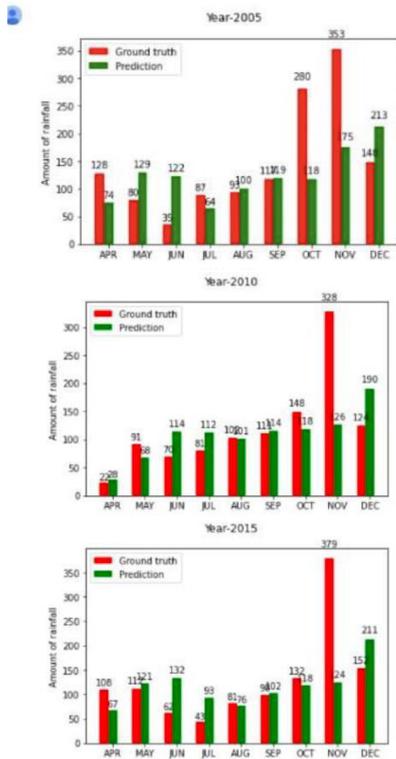
Heat Map shows the correlation(dependency) between the amounts of rainfall over months. From above it is clear that if amount of rainfall is high in the months of July, August, September then the amount of rainfall will be high annually. It is also observed that if amount of rainfall is good in the months of October, November, December then the rainfall is going to be good in the overall year.



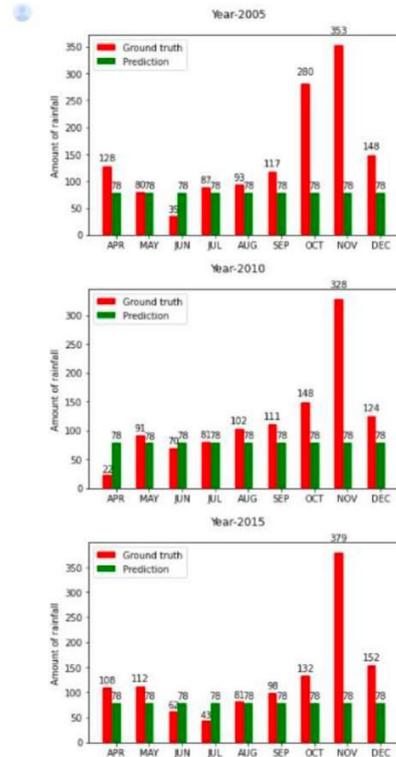
### 5. RESULTS

#### 5.1 PROPOSED SYSTEM RESULTS

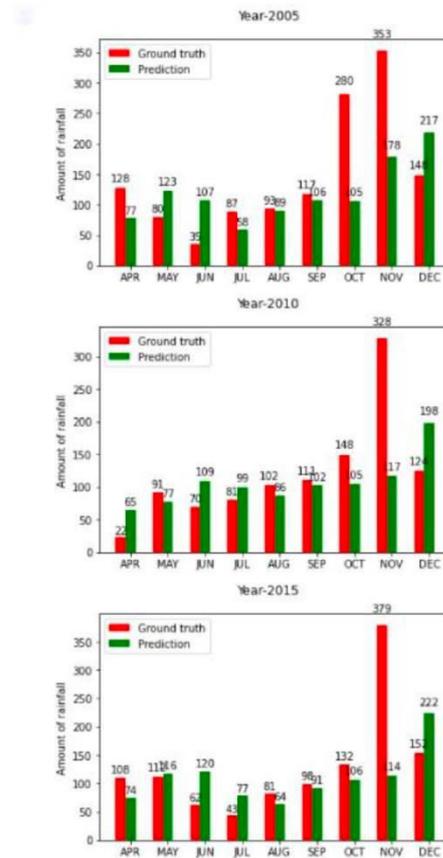
Predicted Result of Tamil Nadu using Neural Network:



Predicted Result of Tamil Nadu using SVR:



Predicted Result of Tamil Nadu using Linear Regression Model:



5.2 COMPARISON OF PROPOSED SYSTEM VS EXISTING

MAE for Whole India:

Algorithm	MAE
Linear Regression	70.61463829282977
SVR	90.30526775954294
Artificial neural nets	59.95190786532157

MAE for South-East Coast:

Algorithm	MAE
0 Linear Regression	51.751792
1 SVR	61.913961
2 Artificial_neural_net	49.234659

## 6. CONCLUSION

Machine learning algorithms like Linear regression and Arima algorithm will be helping for the Website for the rainfall prediction using the machine learning algorithms like linear regression, SVM, Self-organisation Map Machine learning algorithms hence it would be useful for the farmers and the government so that they can be predict the rainfall which are useful for the cultivation of crops and finding the floods in the coastal area ,hence the machine learning algorithms will be very useful website and important for the farmers and the government.

We can use real time data with more features to predict the upcoming Rainfall, but also weather, humidity, wind etc. And we can design and integrate all of these predictions in a web application.

## 7. REFERENCES

- [1] M. Kannan, S. Prabhakaran, P. Ramachandran, "Rainfall Forecasting Using Data Mining Technique", International Journal of Engineering and Technology Vol.2 (6), 2010, 397-401.
- [2] G. Bala Sai Tarun, J.V. Sriram, K. Sairam, K. Teja Sreenivas, M.V.B.T. Santhi, "Rainfall prediction using Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7, May 2019.
- [3] Mr. M. S. Chaudhari, Dr. N. K. Choudhari, "Study of various Rainfall Estimation & Prediction Techniques using Data Mining" American Journal of Engineering Research (AJER), Volume-6, Issue-7, pp-137-139, 2017 [e-ISSN:2320-0847, p-ISSN:2320-0936]
- [4] Mr. Dhawal Hirani, DR. Nitin Mishra, "A Survey On Rainfall Prediction Techniques", International Journal of Computer Application (2250-1797) Volume 6– No.2, March- April 2016.
- [5] Rohit R. Deshpande, "On The Rainfall Time Series Prediction Using Multilayer Perceptron Artificial Neural Network", International Journal of Emerging Technology and Advanced Engineering, (ISSN: 2250-2459) Volume 2, Issue 1, January 2012)
- [6] Deepak Kumar, Anshuman Singh, Pijush Samui & Rishi Kumar Jha, "Forecasting monthly precipitation using sequential modelling", Hydrological Sciences Journal, (DOI: 10.1080/02626667.2019.1595624)
- [7] C. Venkatesanet, S. D. Raskar, S. S. Tambe, B. D. Kulkarni, and R. N. Keshavamurty, "Prediction of all India summer monsoon rainfall using Error-Back-Propagation Neural Networks," Meteorology and Atmospheric Physics, pp. 225-240, 1997
- [8] Jörg D.Wichard, "Forecasting the NN5 time series with hybrid models", International Journal of Forecasting Volume 27, Issue 3, July–September 2011, Pages 700-707
- [9] C.I. Christodoulou, S.C. Michaelides, M. Gabella, C.S. Pattichis, "Prediction of Rainfall Rate Based on Weather Radar Measurements", International Joint Conference on Neural Networks (IJCNN), 25-28 July 2004 Budapest, Hungary
- [10] Gowtham Sethupathi, Yenugudhati Sai Ganesh, Mohammad Mansoor Ali, "Efficient Rainfall Prediction and Analysis using Machine Learning Techniques", Turkish Journal of Computer and Mathematics Education

