



PREDICTING RAINFALL USING MACHINE LEARNING TECHNIQUES

BATCHANABOINA SRILATHA¹, ANNADASU SAIPRATHIMA², GUNTUPALLI SAI
SRAVANI³, MALLADI SARALA DEVI⁴, Dr. G.RAMA SWAMY⁵

^{1,2,3}UG Scholars, Department of CSE, *MALINENI LAKSHMAIAH WOMEN'S ENGINEERING COLLEGE,*
GUNTUR, India.

⁴ Professor, Department of CSE, *MALINENI LAKSHMAIAH WOMEN'S ENGINEERING COLLEGE,*
GUNTUR, India.

ABSTRACT:

Rainfall prediction is important as heavy rainfall can lead to many disasters. The prediction helps people to take preventive measures and moreover the prediction should be accurate. There are two types of prediction short term rainfall prediction and long term rainfall. Prediction mostly short term prediction can give us the accurate result. The main challenge is to build a model for long term rainfall prediction. Heavy precipitation prediction could be a major drawback for earth science department because it is closely associated with the economy and lifetime of human. It's a cause for natural disasters like flood and drought that square measure encountered by individuals across the world each year. Accuracy of rainfall statement has nice importance for countries like India whose economy is basically dependent on agriculture. The dynamic nature of atmosphere, applied mathematics techniques fail to provide sensible accuracy for precipitation statement. The prediction of precipitation using machine learning techniques may use regression. Intention of this project is to offer non-experts easy access to the techniques, approaches utilized in the sector of precipitation prediction and provide a comparative study among the various machine learning techniques.

Keywords: *Deep learning, ML, smart phone.*



1. INTRODUCTION:

In today's situation, rainfall is considered to be one of the sole responsible factors for most of the significant things across the world. In India, agriculture is considered to be one of the important factors for deciding the economy of the country and agriculture is solely dependent on rainfall. Apart from that in the coastal areas across the world, getting to know the amount of rainfall is very much necessary. In some of the areas which have water scarcity, to establish rain water harvester, prior prediction of the rainfall should be done. This project deals with the prediction of rainfall using machine learning & neural networks. The project performs the comparative study of machine learning approaches and neural network approaches then accordingly portrays the efficient approach for rainfall prediction. First of all, preprocess is performed. Preprocess is the process of representing the dataset

in the form of several graphs such as bar graph, histogram etc. When it comes to machine learning, LASSO regression is being used and for neural network, ANN (Artificial neural network) approach is being used. After calculation, types of errors, accuracy of both LASSO and ANN has been compared and accordingly conclusion has been made. To reduce the systems complexity, the prediction has been done with the approach that has better accuracy. The prediction has been done using the dataset which contains rainfall data from year 1901 to 2015 for different regions across the country. It contains month wise data as well as annual rainfall data for the same. Currently, rainfall prediction has become one of the key factors for most of the water conservation systems in and across country. One of the biggest challenges is the complexity present in rainfall data. Most of the rainfall prediction system, nowadays are unable to find the hidden layers or any non-linear



patterns present in the system. This project will assist to find all the hidden layers as well as non-linear patterns, which is useful for performing the precise prediction of rainfall [1]. Rainfall prediction is the application to predict the rainfall in a given region. It can be done in two types. The first is to analyze the physical law that affects rainfall and the second one is to make a system which will discover hidden patterns or the features that affects the physical factors and the process involved in achieving it. The second one is better because it doesn't include any type of mathematical calculations and can be useful for complex and non-linear data [2]. Due to presence of the system which doesn't find the hidden layers and nonlinear patterns accurately, the prediction results to be wrong for most of the times and that may lead to huge losses. So, the main objective for this research work is to find a system that can resolve both the issues i.e. able to find complexity as

well as hidden layers present, which will give proper and accurate prediction thereby assisting the country to develop when it comes to agriculture and economy [3].

2 RELATED STUDY

Rainfall forecasting is very important because heavy and irregular rainfall can have many impacts like destruction of crops and farms, damage of property so a better forecasting model is essential for an early warning that can minimize risks to life and property and also managing the agricultural farms in better way. This prediction mainly helps farmers and also water resources can be utilized efficiently. Rainfall prediction is a challenging task and the results should be accurate. There are many hardware devices for predicting rainfall by using the weather conditions like temperature, humidity, pressure. These traditional methods cannot work in an efficient way so by using machine learning techniques we

can produce accurate results. We can just do it by having the historical data analysis of rainfall and can predict the rainfall for future seasons. We can apply many techniques like classification, regression according to the requirements and also we can calculate the error between the actual and prediction and also the accuracy. Different techniques produce different accuracies so it is important to choose the right algorithm and model it according to the requirements. Regression analysis deals with the dependence of one variable (called as dependent variable) on one or more other variables, (called as independent variables) which are useful for estimating and/ or predicting the mean or average value of the former in terms of known or fixed values of the latter. For example, the salary of a person is based on his/her experience here; the experience attribute is independent variable salary is dependent variable. Simple linear regression defines the relationship

between a single dependent variable and a single independent variable. The below equation is the general form of regression.

3. PROPOSED SYSTEM:

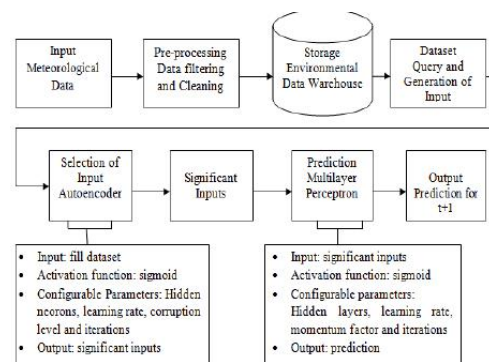
The proposed method is based on the multiple linear regression. The data for the prediction is collected from the publically available sources and the 70 percentage of the data is for training and the 30 percentage of the data is for testing. Figure describes the block diagram of the proposed methodology. Multiple regressions is used to predict the values with the help of descriptive variables and is a statistical method. It is having a linear relationship between the descriptive variable and the output values. The following is the equation for multiple linear regression:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon$$

Hyper plane- in SVM this is a basically The Separation line between the data classes also in SVR we are

going to define it is as the line that will that will help us to predict the continuous value or target value. Boundary line - the SVM plane which creates imagine the support vector can be on boundary lines or outside the boundary line separates two classes in the concept same. Vectors-these are the data points which are closest to the boundary the distance of the point is minimum. SVR performs linear regression in higher dimensional space. We can think of SVR as if each data point in the training represents its own dimension. When we evaluate kernel between a test point and a point in the training set the resulting value gives you the coordinate of your test point in that dimension. The vector we get when we evaluate the test point for all points in the training set, k is the representation of the test point in the higher dimensional space. The equation of the hyper plane is $wx+b=0$ and the two equations of boundary lines is $Wx+b=+e$, $Wx+b=-e$ Equation that satisfy our SVR is

$e \leq y - Wx - b \leq +e$ SVR has a different regression goal compared to linear regression in linear regression, we are trying to minimize the error between the prediction and data whereas in SVR a goal is to make sure that error do not exceed the threshold.

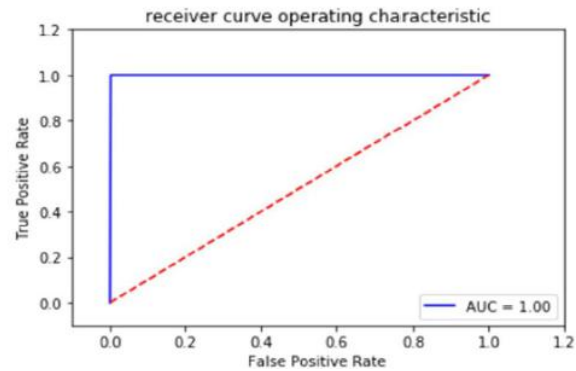
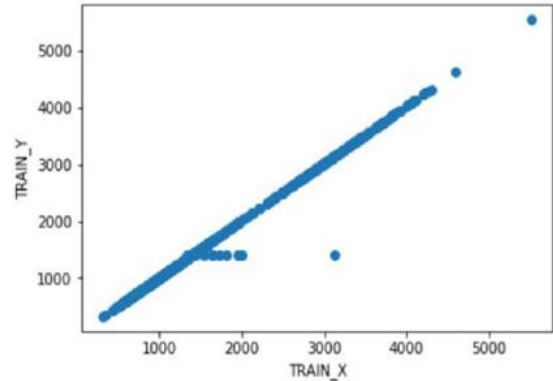


RESULTS:

Bar graph is the plotting of data in the form of bar. Mainly stacked bar graph is used in this project. Histogram is the representing of data with the help of bars of different height. For both LASSO and neural network, accuracy is calculated after algorithm is being trained. Both the algorithms are trained and then the accuracy is received for both. The accuracy is being calculated after calculating

errors. The errors calculated are MAE, MSE, R-SQUARED and RMSE. To reduce the complexity of the system, prediction will only be done for the most accurate algorithm because if both the algorithms are giving predicted value, then that can lead to unnecessary usage of data and that will increase the complexity of the system and make it slow. The result will be received in the form of graphs, Metrics and excel sheet. The graphs received are Correlation Metrics, Scatter Metrics, Max value for month by month, Subdivision mean value for month by month, Sum of every quarter of subdivision, Sum of every quarterly, Sum of year by year and Sum of month year by year. Along with this two Metrics, which will be show the accuracy of both the algorithms will be received. Three excel sheet will be received at the end of all process. Two will have the accuracy for both LASSO and neural network in tabular form and one excel sheet will have the predicted value

which will be received from the most accurate algorithm.



4. CONCLUSION:

Rain fall prediction plays the major role in agriculture production. The growth of the agricultural products is based on the rainfall amount. So it is necessary to predict the rainfall of a season to assist farmers in agriculture. The proposed method predicts the rainfall for the Indian dataset using multiple linear regression and



provides improved results in terms of accuracy, MSE and correlation.

REFERENCES:

[1] Manojit Chattopadhyay, Surajit Chattopadhyay, “Elucidating the role of topological pattern discovery and support vector machine in generating predictive models for Indian summer monsoon rainfall”, *Theoretical and Applied Climatology*, pp. 1-12, July 2015, DOI: 10.1007/s00704-015-1544-5

[2] Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar,” A Rainfall Prediction Model using Artificial Neural Network”, 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC 2012), pp. 82-87, 2012.

[3] Minghui Qiu, Peilin Zhao, Ke Zhang, Jun Huang, Xing Shi, Xiaoguang Wang, Wei Chu, “A Short-Term Rainfall Prediction Model using Multi-Task Convolutional Neural Networks”, *IEEE International Conference on Data Mining*, pp. 395-

400, 2017, DOI 10.1109/ICDM.2017.49.

[4] Aswin S, Geetha P and Vinayakumar R, “Deep Learning Models for the Prediction of Rainfall”, *International Conference on Communication and Signal Processing*, April 3-5, 2018, India, pp. 0657-0661.

[5] Xianggen Gan, Lihong Chen, Dongbao Yang, Guang Liu, “The Research Of Rainfall Prediction Models Based On Matlab Neural Network”, *Proceedings of IEEE CCIS2011*, pp. 45- 48.

[6] Sam Cramer, Michael Kampouridis, Alex A. Freitas and Antonis Alexandridis, “Predicting Rainfall in the Context of Rainfall Derivatives Using Genetic Programming”, 2015 *IEEE Symposium Series on Computational Intelligence*, pp. 711 – 718.

[7] Mohini P. Darji, Vipul K. Dabhi, Harshadkumar B.Prajapati, “Rainfall Forecasting Using Neural Network: A Survey”, 2015 *International*



Conference on Advances in Computer Engineering and Applications (ICACEA) IMS Engineering College, Ghaziabad, India, pp.706 – 713

Analytics and Cloud) (I-SMAC 2017), pp. 414 - 417

[8] Sandeep Kumar Mohapatra, Anamika Upadhyay, Channabasava Gola, “Rainfall Prediction based on 100 years of Meterological Data”,2017 International Conference on Computing and Communication Technologies for smart Nation, pp.162 – 166.

[9] Sankhadeep Chatterjee, Bimal Datta, Soumya Sen, Nilanjan Dey, “Rainfall Prediction using Hybrid Neural Network Approach”, 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), pp. 67 – 72.

[10] Mr. Sunil Navadia, Mr. Pintukumar Yadav, Mr. Jobin Thomas, Ms. Shakila Shaikh, “Weather Prediction: A novel approach for measuring and analyzing weather data”, International conference on I-SMAC (IoT in Social, Mobile,