

# A Review of Application of Data Mining Technique for Rainfall Prediction

Nikhil Sethi<sup>1</sup> Dr. Kanwal Garg<sup>2</sup>

<sup>1</sup>M.Tech. Scholar <sup>2</sup>Assistant Professor

<sup>1,2</sup> Department Of Computer Science and Application

<sup>1,2</sup> Kurukshetra University, Kurukshetra

**Abstract---** Globally more than 60% of population lives in monsoon regions. Agriculture is the dominating occupation in India on which most of the economy depends. Due to shortage of water supply for irrigation, agriculture works mostly depends upon other source of water like rainfall so; early prediction of rainfall is necessary for those countries where agriculture plays an important role in economy. Prediction of rainfall is a complex task as it is depends upon various factors. Various techniques are available in literature for early prediction of rainfall. Data mining can also be used for rainfall prediction

This paper gives a review of various data mining techniques for rainfall prediction. The mainly used techniques are artificial neural network, Regression, Clustering etc. All of these provide solution for early prediction of rainfall. This paper also encourages additional research and provides suggestion for further research.

**Keywords:** - Rainfall, Prediction, Neural Network, Back Propagation, Regression.

## I. INTRODUCTION

Data mining is a technique to analyze the large dataset for extracting the hidden pattern among large dataset and then use this extracted information from dataset are transformed into usable knowledge which can be used for various purposes.

There are generally two models of data mining one is the descriptive and other one is the predictive. In the descriptive model historical data are analyzed to form groups or clusters of similar items, where on other hand the predictive model are used for extracting the pattern from the dataset and then used this patterns for prediction of values of the variable.

There are two types of prediction one is forecast which is for month to years in advance where other one is short-term prediction that is just for few hours ahead or a day in advance.

Rainfall prediction is one of the most difficult challenge around the world, one of the major issue that metro logiest are facing around the world is the accurate prediction of rainfall data as it depends upon various factors which are uncertain to predict.

Presently metrological department obtain rainfall data using various means like ground observation, satellite observation, observations taken from ships, aircraft, radar etc and then metrological department arrange these obtained dataset in form of sheets, graphs, charts etc where these dataset are available for analysis purpose and using data mining technique we can find pattern among the dataset and used for prediction purpose.

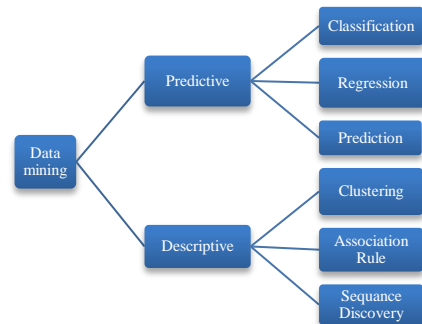


Fig. 1:

Rainfall forecast for 12-24 hours are quite accurate whereas the prediction accuracy degrades as the time period increase it is good for two three days but when the days increases beyond five days the accuracy drops rapidly.

Rainfall prediction is not only helpful in agriculture but it also helps in various other fields like industries, defense, shipping aerospace navigation etc. It is also helpful in early prediction of disaster like flood, drought etc.

## II. RELATED WORK

Godfrey C.Onwubolu et al. [1] presented the enhanced group method of data handling for mining weather data .the author used daily temperature, pressure ad monthly rainfall. The author compare the performance of e-GMDH with other method of data handling such as polynomial neural network and it's enhanced version and author concluded that E-GMDH perform better than others models.

Wint Thinda Zaw et al. [2] proposed a model for rainfall prediction using multiple polynomial regression using predictors SOI, ONI, SST etc .The author compare this model with linear regression model and concluded that result of Multiple Polynomial regression model are more accurate.

Elia Georgiana Petre [3] presented the prediction of temperature based on the decision tree build using CART algorithm .the author performed the technique using data mining WEKA tool.

Jehangir Ashraf Awar et al. [4] explored the use of artificial neural network or monsoon prediction .the author used the back propagation and learning vector quantization. The author compares the performance of this model with statistical downscaling model and demonstrated that ANN performs better. The performance parameters selected by the author are accuracy, Correlation coefficient.

Enireddy. Vamsidhar et al. [5] examined the Artificial neural network, one of the supervised technique of data mining for rainfall prediction and the author used humidity, dew point, pressure as the predictor variable for rainfall and model gave 94.28% accuracy in result.

M.Kanan et al. [6] computed the ground level rainfall using Karl Pearson correlation coefficient and used multiple linear regression approach for future years prediction and results show the approximated values.

Zhi-Liang Wang et al. [7] proposed the generalize regression neural network model for prediction of rainfall .The author compare the GRNN model with back propagation neural network and stepwise regression analysis method of prediction and concluded that GRNN is more accurate than Back propagation neural network.

Sarah N.Kohli et al. [9] discussed the case study of Gaza strip city and the author applied various data mining technique like outlier analysis, clustering, association rule mining, prediction to extract useful knowledge from large dataset and then use this extracted information in transformed into usable knowledge and then author compare the result of these technique with other prediction technique, neural network based on performance parameter like Root mean square error (RMSE) and Accuracy.

Folorunshoolaiya et al.[10] presented a model for predicting weather data based on the decision tree and artificial neural network(ANN).The author used TLFN neural network and Recurrent network architecture for developing predictive ANN model and concluded that recurrent neural network model perform better than TLFN network based on the Correlation coefficient mean square error and % error.

Meghali A. Kalyankar et al. [12] presented the overview of data mining technique for weather data and used K-mean clustering algorithm to find hidden pattern among the large dataset and these extracted information are transformed to usable knowledge which can further be used for prediction purposes.

Jyothis.Joseph et al. [13] demonstrated the clustering and classification for prediction of rainfall .The author performed the subtractive clustering and used ANN for implementing these approach .The author used Temperature, pressure, humidity, wind speed as the input variable

### III. PREDICTION TECHNIQUES

#### A. Multiple Linear Regression

Multiple linear regression is a statistical technique to predict the value of a variable depending upon the value of others predictor variables. The result of Multiple linear regression is a mathematical expression describing the relationship between the input and output variables. Input variables used here are the predictor variables that are used to determine the predicted value of a variable and output is the predicted value of a variables. The general expression of the multiple linear regressions is given below.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip}$$

here Y=output variable

X=input variable

There are some drawbacks of Multiple linear regression which are given below

- Due to complex behavior of rainfall results in
- extreme monsoon year are not satisfactory
- Normally a prediction of one month ahead is possible with a good accuracy

- Value depends upon various input parameter.

#### B. Artificial Neural Network

An artificial neural network (ANN) is a mathematical model whose structure looks similar to biological network. In the network there are various processing element known as neuron. Every neuron has its own activation function. In the network there are normally three layers of neuron one is called input layer , second one is hidden layer and last one is output layer .the neuron between different layer are linked by some weight the general diagram representing the neural network is given below.

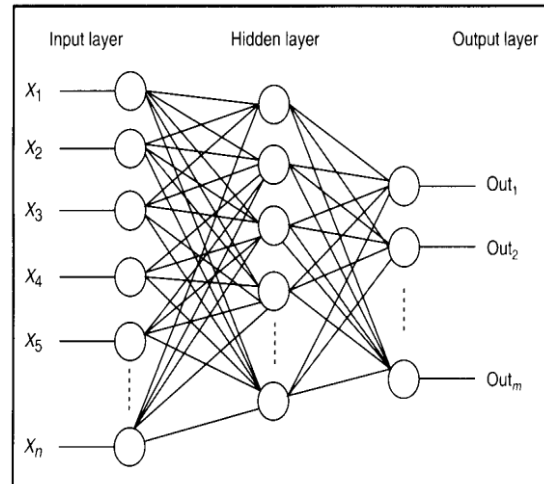


Fig. 2:

ANN is used in variety of fields like business, industry and science. While using ANN firstly we have to train the network that is learning phase. In learning phase we have to adjust the weight between the neurons .Once the network is trained to use then the network is able to find hidden and nonlinear dependencies from large dataset . normally two third of data set are used for training of network and rest one third of dataset are used to test the network.

ANN is of two types one is single layer neural network and other one is multilayer neural network. in single layer there is only one hidden layer of neuron while in the multilayer neural network there are two or more than two hidden layers of neuron.

There are two types of learning of a network one is to feed forward in which input signal is propagated to output layer from input layer via hidden layer , in this method weight values between neuron are fixed and this type of network are of limiting use . Sometimes we have to adjust the weight between neuron to solve a particular problem and obtaining an expected output from the network then we have to use Back Propagation neural network, in the back propagation error value which is the difference between the expected value and actual value in signal is sent back to input layer from output layer and weight between the neuron are adjusted according to the error value .The continuous changes between the weight made values close to actual value.

#### C. Nearest Neighbor

Nearest Neighbor is one of the prediction technique in data mining .it's basic concept is similar to clustering technique which is process for grouping related items together so that

in a single cluster all items are look similar and items in different cluster are different .the main goal of clustering is to reduce inter cluster similarity and intra cluster similarity must be high. nearest neighbor technique state that object that are near to each other will going to have same predicted value .there is some difference between the clustering and nearest neighbor technique the main difference is that clustering is a unsupervised learning technique whereas nearest neighbor is a supervised learning technique .K-nearest neighbor is the simplest and the most appropriate algorithm where Euclidian distance from the query instance to training samples to determine K-nearest neighbor

Nearest neighbor technique is used in many fields of data mining, image processing , pattern recognition .some of the application area of this technique are content retrieval ,protein- protein interaction and 3D structure prediction, recognition of handwriting etc.

#### IV. EVALUATION OF TECHNIQUES

While selecting the appropriate technique for rainfall prediction we have to consider various performance parameter depending upon which we have to select the best technique among various available techniques for prediction .the commonly used performance parameter are given below

##### A. RMSE

RMSE is root mean square error, error is the difference between the actual value and predicted value .RMSE is the square root of mean square of error values .the mathematical formula for calculating RMSE is given as below. The Desired value of RMSE is close to zero.

$$\sqrt{\frac{\sum_{t=1}^n (y_p - y_a)^2}{n}}$$

Here Yp = predicted value  
Ya = actual Value

##### B. Correlation Coefficient

Correlation coefficient is the measure of the co-relation between the actual value and predicted value .it's value lies between -1 to +1. Where +1 represent the strong positive correlation between the variables and -1 represent the strong negative correlation .when the value of correlation coefficient is 0 then there is no co-relation between the two variables and both variables are independent of each other's.

##### C. Percentage error

Percentage error represents percentage the predicted value differs from the actual value. The mathematical formula for percentage error is given below. The value of percentage error should be close to zero for better and accurate result.

$$\%Error = \frac{Actualvalue - Predictedvalue}{Actualvalue} \times 100$$

#### V. CONCLUSION

In this paper the authors conducted a review of various data mining techniques appropriate for the rainfall prediction.

The authors also discussed performance measure of the technique and concluded that there are various technique which are appropriate for rainfall prediction, mainly used technique are Artificial neural network, Regression model and K-nearest neighbor technique. Due to generalization ability of ANN model of prediction to perform better prediction than regression model of prediction .The authors believe that this paper will prove as a useful resource for anyone who is interested in the research work of rainfall prediction .In the future we can also use other technique like artificial intelligence, fuzzy set etc. for predicting rainfall.

#### REFERENCES

- [1] Godfrey C.Onwubolu Et.Al “Self-Organizing Data Mining For Weather Forecasting” In Iadis European Conference Data Ming Isbn: 978-972-8924-40-9 © 2007.
- [2] Wint Thida Zaw Et.Al “Empirical Statistical Modeling Of Rainfall Prediction Over Myanmar” In World Academy Of Science,Engineering And Technology,International Science Index 22, 2008 Waset.Org/Publication/15031.
- [3] Elia Georgiana Petre Et.Al “A Decision Tree For Weather Prediction” In Buletinul Universitaii Petrol-Gaze Din Ploiesti Vol No..Lxi 1/2009 Page 77-82.
- [4] Jehangir Ashraf Awan Et.Al “Application Of Artificial Neural Networks For Monsoon Rainfall Prediction” In 6th International Conference On Emerging Technologies (Icet) 978-1-4244-8058-6/10 2010 Ieee
- [5] Enireddy.Vamsidhar Et.Al “Prediction Of Rainfall Using Backpropagation Neural Network Model” In International Journal On Computer Science And Engineering (Ijcse) Issn : 0975-3397 Vol. 02, No. 04, 2010, Page 1119-1121.
- [6] M.Kannan Et.Al “Rainfall Forecasting Using Data Mining Technique” In International Journal Of Engineering And Technology Issn: 0975-4024 Vol.2 (6), 2010, 397-401.
- [7] Zhi-Liang Wang Et.Al “Rainfall Prediction Using Generalized Regression Neural Network: Case Study Zhengzhou” In 2010 International Conference On Computational And Information Sciences 978-0-7695-4270-6/10 Doi 10.1109/Iccis.2010.312 . 2010 Ieee.
- [8] Sarah N.Kohail Et.Al “ Implementation Of Data Mining Techniques For Metrological Data Analysis” In International Journal On Information And Communication Technology Research (Ijict) Vol. 1 No 3,, July 2011.
- [9] Sarah N.Kohail Et.Al “Implementation Of Data Mining Techniques For Meteorological Data Analysis” In International Journal Of Information And Communication Technology Research Issn-2223-4985 Volume 1 No. 3, July 2011.
- [10] Folorunsho Olaiya Et.Al “Application Of Data Mining Techniques In Weather Prediction And Climate Change Studies” In I.J. Information Engineering And Electronic Business, 2012, 1, 51-59 Http://Www.Mecs-Press.Org/.
- [11] Paras Et.Al “ A Simple Weather Forecasting Model Using Mathematical Regression ” In Indian Research

Journal Of Extension Education Special Issue (Volume I), January, 2012.

- [12] Meghali A.Kalyankar Et.Al “Data Mining Technique To Analyse The Metrological Data” In International Journal Of Advanced Research In Computer Science And Software Engineering(Ijarcse) Issn: 2277-128x WwW.Ijarcse.Com Volume 3, Issue 2, February 2013 Pp 114-118.
- [13] Jyothis Joseph Et.Al “Rainfall Prediction Using Data Mining Techniques” In International Journal Of Computer Application (Ijca) Issn: 0975-8887 Vol. 83, 8, December 2013.

