ISSN No. 0976-5697

Volume 5, No. 2, March 2014 (Special Issue)

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Empirical Method Technique to Make Short Term Forecast of Rainfall for a Specific Region

D Ratnam
[1]Assistant Professor, Dept. of Information Technology, P V P Siddhartha Institute of Technology
[1]ratnam.dodda@gmail.com

Dr J Rajendra Prasad
[2]Professor and HOD, Dept .of Information Technology, P V P Siddhartha Institute of Technology
[2]rp.rajendra@rediffmail.com

Dr B V SubbaRao
[3]Professor, Dept. of Information Technology, P V P Siddhartha Institute of Technology
[3]bvsrau@gmail.com

Abstract: The forecasting of rainfall is a vital application in data mining techniques. In the event of planning and making decision for agricultural crop pattern and water management, the prediction of rainfall at long intervals is very helpful and useful. In fact, rainfall plays an important role for food generation, water resource management and other activity plans observed in nature. It is observed that the crop prediction is affected due to the rate arrival of monsoons as well as heavy rainfall. India is basically an agricultural based country and its economy gets affected bearing upon crop productivity. Therefore, rainfall prediction becomes an important element in deciding the destiny of countries like India that are dependent on agriculture. Rainfall forecasting has been one of the most challenging hurdles scientifically and technologically all over the world since the last century. In this paper we have presented the rainfall prediction for a specific area in the state of Andhra Pradesh using multiple linear regression technique.

Keywords: Empirical method, Rainfall forecasting, Multiple Linear Regression.

I. INTRODUCTION

In a country like India where agricultural is observed as spine of its economy and the survival of its people lings upon the monsoon rainfall during the rainy season that lasts from June to September. It is observed that there is a heavy loss in agricultural output during severe drought. India is mainly an agricultural based country and so the success or failure of the harvest and water scarcity observed in every year is always taken into account with greater concern. Different varieties of forecast methods are used in weather forecasting at regional and national levels. It is of said that there are two fundamental approaches to predict the rainfall. They are Empirical method and dynamic method. The empirical approach relies on analysis of historical data of the rainfall and its association to a variety of atmospheric and oceanic variables that are observed in different parts of the world [1]. The most widely used empirical approaches for prediction climatic changes are regression, artificial neural networks, fuzzy logic and group method of data handling. In the case of dynamical approaches predictions are generated by physical models depending upon systems of equations that foreseen the evolution of global climate system in response to initial atmospheric conditions.

The empirical approaches are implemented applying numerical rainfall forecasting method. The present paper delineates empirical method that belongs to the regression approach which attempts to assume a short –term forecast of rainfall in the specified area of Krishna district. The main objective of analysis the June month rainfall data of particular region for the successive past seven years is that the rainy season is observed during the period. Multiple linear regressions are also used to predict the rainfall following the previous year's data during the corresponding period. In section 2 we have presented the related work, section 3 data analysis, in section 4 regressions, and in section 5 we have shown the results.

II. RELATED WORK

It is observed that the accurate and timely weather forecasting is a major challenge for the scientific community. Rainfall prediction modeling comprises a combination of complete models. observation and knowledge of trends and patterns. While applying these methods reasonable accurate forecast can be estimate. Various research studies are adapted recently for the prediction of rainfall studying different weather and climate forecasting methods. Regression is statistical empirical technique and is widely used in business, the social and behavioral sciences, the biological sciences, climate predictions, and in many other areas.

In India rainfall data varies significantly from region to region. The proposed study analyzes the impact of various global atmospheric parameters such as Sea Surface Temperature (SST), Sea Level Pressure (SLP), U-wind, V-wind, U-wind stress and V-wind stress on rainfall of Krishna district Region of Andhra Pradesh. In this analysis, correlation and multiple regression approach is used for prediction of rainfall over Krishna district.M. Rajeevan[2],proposed New Models for Long Range Forecasts of Summer Monsoon Rainfall over North West and Peninsular India.The experimental results showed that the model error was 4%. N krintra Singhrattna, [3] described the development of a statistical forecasting method for SMR over Thailand using multiple linear regression and local polynomial-based nonparametric approaches. SST, sea level

CONFERENCE PAPER
Two day National Conference on Advanced Trends and Challenges in Computer Science and Applications
Organized by: Shree Vishnu Engineering College for Women, Bhimavaram A.P.
Schedule: 18-19 March 2014

pressure (SLP), wind speed, EiNino Southern Oscillation Index (ENSO), IOD was chosen as predictors. The experiments indicated that the correlation between observed and forecast rainfall Keon Tae Sohn [4], has developed a prediction model for the occurrence of heavy rain in South Korea using multiple linear and logistics regression, decision tree and artificial neural network. M. Rajeevan, D. S. Pai , Anil Kumar Rohilla [5] modeled for prediction categories of rainfall (below, above, normal) in the highlands of Eritrea. The most influential predictor of rainfall amount was the southern Indian Ocean SST.

Experimental results showed that the hit rate for the model was 70%. Hossein Hassani [6] proposed human height prediction model based on multiple polynomial regressions that was used successfully to forecast the growth potentials of height with Precision and was helpful in children growth study. Zaw, W.T.[7] proposed Modeling of Rainfall Prediction over Myanmar Using Polynomial Regression.Gouthami Bandyopadya[8],The prediction of Indian Monsoon Rainfall :A Regression Approach present paper analyses the monthly rainfall data of the Indian Summer Monsoon. Dinu John [9], Monsoon Onset Prediction using Data Mining proposed the Onset of monsoon is eagerly awaited in the Indian sub-continent as it has deep impact in the economic and social domain. Seema Mhajan[10],Modelling and Prediction of Rainfall data using Data Mining proposed long term rainfall prediction is very useful in planning and decision making of agricultural crop pattern and water management strategy.

III. DATA ANALYSIS

The rainfall forecasting is analyzed using the methods of artificial intelligence, neural networks, and data mining in some of the journals. Artificial intelligence and neural networks are more difficult when compared to data mining because artificial intelligence involves some artificial neural networks are computational models inspired by animals' central nervous systems (in particular the brain) that are capable of machine learning and pattern recognition. In Data Mining, some of the functionalities are used i.e. classification, clustering, regression or prediction, association etc. it is also classified to reason out the rainfall at the ground level. The element is grouped using clustering technique to estimate the rainfall in the particular are.

Finally, the prediction methods are taken into consideration because the occupancy levels of rainfall in done by in the specified region is done be applying the regression approach. In the process of regression, kerl Pearson correlation coefficient is used in finding the measure of rainfall in centimeters in the particular region. It is an obligation to predict the rainfall levels in the coming years, using the multiple linear regression approach. While using the data, the rainfall at the ground level is computed with the help of Pearson Correlations Coefficient. Pearson Correlations Coefficient is used to measure the strength of the association between two variables. If ranges from -1 to +1. Here, seven years data is computed applying Pearson coefficient and then it is compared with the predicted data with the help of regression approach. Generally, rain fall is measured in terms of Millimeters (mm). Measurement and years are given in the graph with x and y axis.

In order to uses the input data, the output is given as soon as possible. Therefore, the output is predicted for the future year's rainfall at the ground level. It is observed that the output data is approximate because the prediction of the rainfall for future years calculated using regression approach. In this approach, some of the predictor variables are used which is useful to predict the rainfall during June month of every year. The data used in the present study is collected from the Chief Planning Officer of Krishna district, Andhra Pradesh.

IV. REGRESSION

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent and one or more independent variables. Regression is of two methods namely Simple linear regression and multiple regression models.

a. Multiple Regressions: In multiple regressions is a technique used for numerical prediction. Regression is a statistical measure that attempts to determine the strength of the relationship between one dependent variable (i.e. the label attribute) and a series of other changing variables known as independent variables (regular attributes). Just like Classification is used for predicting categorical labels, Regression is used for predicting a continuous value.

$$Y = \alpha + \beta 1X1 + \beta 2X2 + \cdots \ldots \ldots \ldots . + \beta KXK + \varepsilon, \quad i=1,2,..N \quad (1)$$

Where β_0, $\beta 1$, βk are the unknown regression coefficients (parameters) of the model and its is assumed that the stochastic errors ε_i are independent and identically distributed with zero mean and constant variance. In this setting it is optimal (ala the Gauss-Markov theorem) to estimate these parameters by the method of least squares which minimizes the following sum of squares errors function

$$S = \sum_{i=1}^{N} \left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_K x_{iK}\right)^2 \quad (2)$$

Over choices of $\beta_0,\beta_1,\ldots,\beta_k$ the solution for this problem comes in the form of following normal equations:

$$\sum_{i=1}^{N} 2\left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_K x_{iK}\right)(-1) = 0$$

$$\sum_{i=1}^{N} 2\left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_K x_{iK}\right)(-x_{i1}) = 0$$

$$\sum_{i=1}^{N} 2\left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_K x_{iK}\right)(-x_{iK}) = 0$$

The "normal" equations represent a system of K non-homogeneous, linear equations in the K unknowns $\beta_0,\beta_1,\ldots,\beta_k$ and can easily be solved by means of matrix algebra assuming that none of the input variables can be written as a linear combination of the other input variables (i.e. the absence of perfect multi collinearity). Let us denote these solutions, called the ordinary least squares estimates, $\beta 0,\beta 1,\ldots,\beta k$

The fit of equation (1) is represented by the sum of squared errors (SSE)

CONFERENCE PAPER
Two day National Conference on Advanced Trends and Challenges in Computer Science and Applications
Organized by: Shree Vishnu Engineering College for Women, Bhimavaram A.P.
Schedule: 18-19 March 2014

$$\mathrm{SSE} = \sum_{i=1}^{N} \hat{\varepsilon}_i^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_0 x_{i1} - \cdots - \hat{\beta}_K x_{iK}\right)^2$$

Where $\varepsilon\hat{\imath}$ denotes the residual of the regression fit of the observation on the output variable for the i-th individual and the fitted value of the observation of the output of the i-th individual is represented by

$$\hat{y}_i = \hat{\beta}_0 - \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_K x_{iK} \ . \qquad (3)$$

In this setting, the ordinary least squares estimates are optimal with respect to being the best of all unbiased, linear estimators of the parameters$\widehat{\beta 0}$,β1,….,βk.A measure of the goodness-of-fit of equation (1) is represented by the so-called R-square (R^2) or coefficient of determination:

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} = \frac{\mathrm{SSR}}{\mathrm{SST}} \qquad (4)$$

Where $SST = \sum_{i=1}^{N}(y_i - \bar{y})^2$ is the total sum of squares of the deviations of the observations from their mean and $SSR = \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$ is the so-called sum of squares due to the regression.R^2 is defined in such a way to satisfy the inequality 0≤R2 ≤1 .In other words R^2 is the percentage of the variation in that is explained by the input variables

V. RESULTS

a. Rapid Miner: Rapid Miner is a software platform developed by the company of the same name that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics. It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the data mining process including results visualization, validation and optimization. In this paper we apply Rapid Miner tool for predicting the rain fall. Finally we get the same results we are calculating manually and Rapid Miner tool.

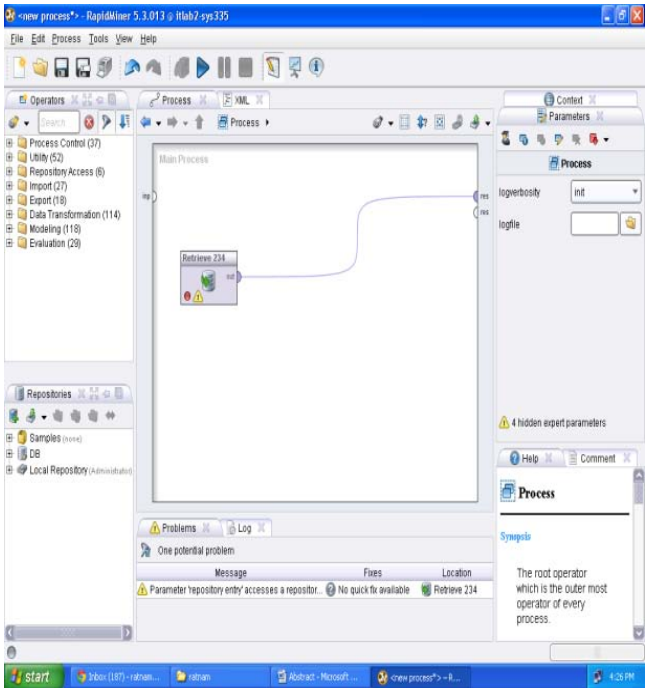

Figure 1.Importing the Rainfall data

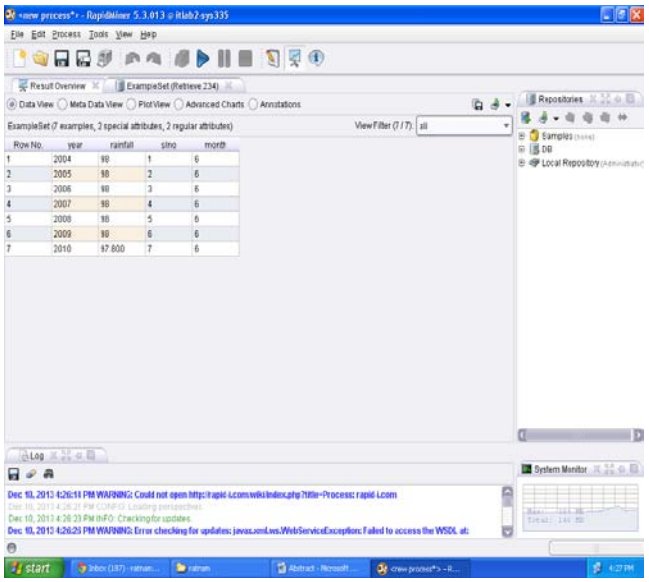

Figure 2. Rainfall data

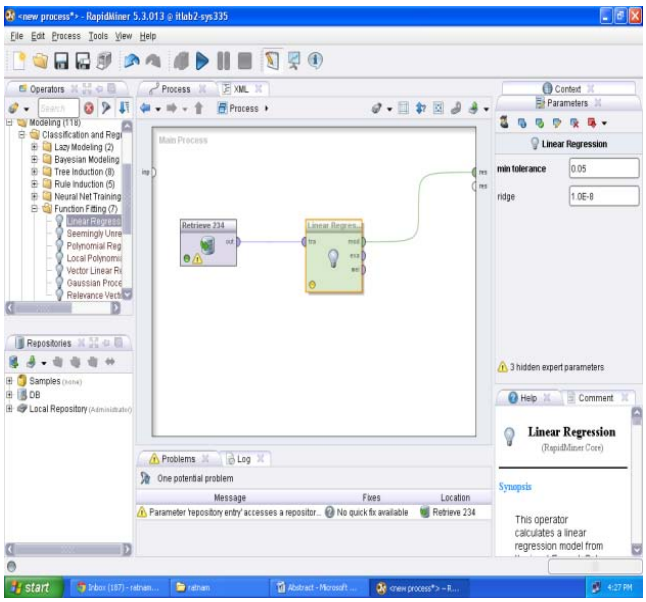

Figure 3. Multiple Linear Regression algorithms

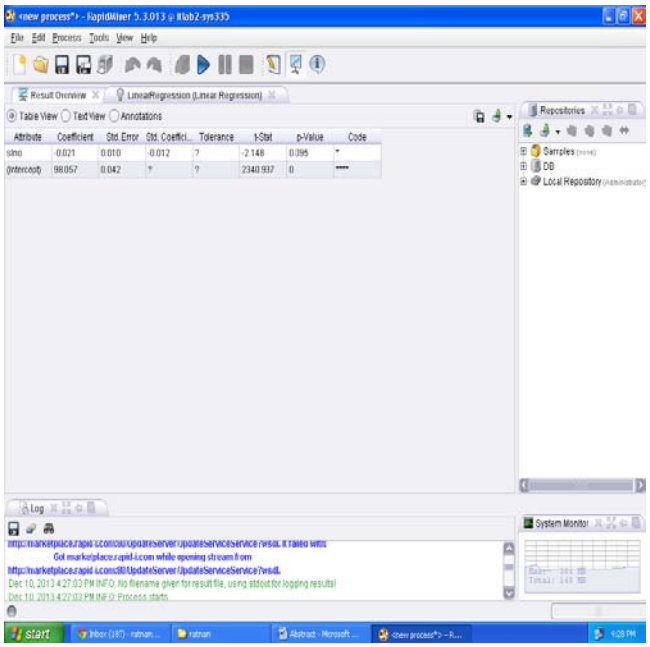

Figure 4. Result in table view

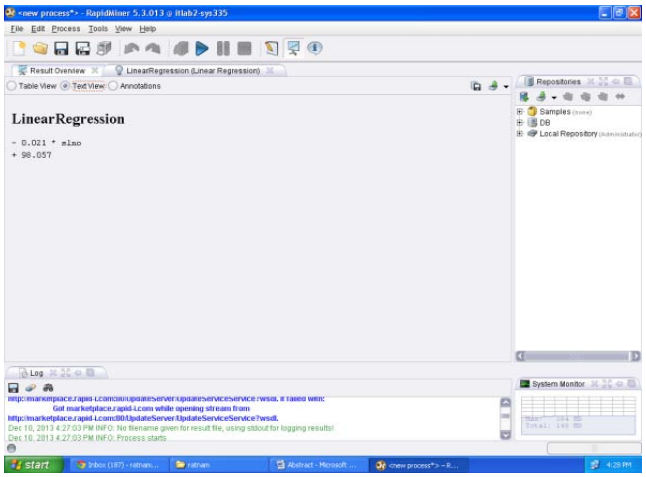

Figure 5. Result -In data view

VI. CONCLUSION

In this paper we applied Multiple Linear Regression approach to extract knowledge from Krishna District Rainfall dataset. The dataset include seven years period from 2004 to June 2010 rainfall observation. We gone through all prediction process and applied many classification and prediction techniques like simple linear regression, logistic regression, and polynomial regression. Multiple linear regressions provide a very useful and accurate knowledge in a form of rules, models, and visual graphs as shown in the figures. This knowledge can be used to obtain useful prediction and support the decision making for different sectors.

VII. ACKNOWLEDGMENT

The authors thanks to the Chief Planning Officer, Machilipatnam, Krishna District for providing the opportunity for sharing information and receiving useful comments on Rainfall data.

VIII. REFERENCES

[1]. M.Kannan, Rainfall Forecasting Using Data Mining Technique.

[2]. N krintra Singhrattna, Seasonal Forecasting of Thailand Summer Monsoon Rainfall.

[3]. M. Rajeevan,New Models for Long Range Forecasts of Summer Monsoon Rainfall over North West and Peninsular India

[4]. Keon Tae Sohn,Statistical prediction of heavy rain in South Korea.

[5]. M. Rajeevan , D. S. Pai , Anil Kumar Rohilla New Statistical Models For Long Range Forecasting of Southwest Monsoon Rainfall Over India.

[6]. Hossein Hassani, The efficiency of noise reduction for curve fitting in growth curve models.

[7]. Zaw, W.T.,Modeling of Rainfall Prediction over Myanmar Using Polynomial Regressions.

[8]. Gouthami Bandyopadya[8],The prediction of Indian Monsoon Rainfall :A Regression Approach .

[9]. Dinu John,Monsoon Onset Prediction using Data Mining.

[10]. Modelling and Prediction of Rainfall data using Data Mining.