

# Meteorological Data Analysis using Artificial Neural Networks

Prajwala T R, D.Ramesh, H Venugopal

**Abstract:** This paper focuses on weather data analysis for Bangalore urban region(Karnataka,India) over a span of 30 years. The 30 years data is preprocessed to have average monthly temperature, vapor pressure, PET (Potential-Evapo Transpiration), cloud cover, rainfall. These features are considered as factors affecting the rainfall. The correlation between the above mentioned parameters with the monthly rainfall are found using spearman correlation. Artificial Neural Networks (ANN) is used to classify instances as less rain, medium and heavy rain. The results of accuracy, confusion matrix is tabulated. Also the optimal number epochs, number of neurons and number of hidden layers is also identified for the data. The graph of actual output and predicted output is plotted.

**Keywords :** Spearman coefficient, Vapour pressure, PET (Potential-Evapo Transpiration), MultiLayerPercepton, confusion matrix, precision, Recall

## I. INTRODUCTION

Data mining, or knowledge discovery, is the computer-assisted process of analyzing enormous sets of data and then extracting the meaning of the data. Meteorological data mining is a form of data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge.

This paper focuses on data analysis of Bangalore urban region for a time span of 30 years from 1975 to 2002 for Bangalore, Karnataka, India region. The data is collected from IMD (Indian Meteorological Department) and preprocessed to have monthly data. There are many factors affecting the rainfall for example temperature, humidity, wind speed, air pressure etc. But this paper focuses on factors like temperature, Vapour Pressure and PET (Potential-Evapo Transpiration). Vapour Pressure is pressure exerted by water droplets at a given temperature in closed environment. It is one of the method to measure humidity in air. PET represents loss of water through process of transpiration by plants and evaporation. This is influenced by wind, sunlight, humidity and temperature. The spearman coefficient is used to find correlation between these parameters and rainfall. ANN is used to classify the rainfall as below average, average and heavy rainfall.

**Revised Manuscript Received on December 22, 2019.**

**Prajwala T Rm**, Professor, PESIT, Bangalore. Emailid: [prajwalatr@gmail.com](mailto:prajwalatr@gmail.com)

**Dr. D Ramesh**, Professor and Head, Master of Computer Application(MCA), SSIT, Tumkur, Karnataka, India.

**Dr.H.Venugopal**, Professor, Department of Computer Science and Engineering [CSE], SSIT, Tumkur, Karnataka, India.

## II. LITERATURE SURVEY

Data mining is technique of discovering useful information from large dataset. It helps in extraction of hidden predictive information from large databases. Weather is one of the meteorological data that is rich by important knowledge. his paper focuses on data analysis of Bangalore urban region for a time span of 30 years from 1975 to 2002. The data is collected from IMD (Indian Meteorological Department) and preprocessed to have monthly data.

Factors like temperature, Vapour Pressure and PET (Potential-Evapo Transpiration) are affecting the amount of rainfall. Vapour Pressure is pressure exerted by water droplets at a given temperature in closed environment. At a given temperature, an increase of water vapour in the air corresponds to an increase in the humidity of the air. PET represents loss of water through process of transpiration by plants and evaporation. This is influenced by wind, sunlight, humidity and temperature.

Coefficient of correlation is one of the statistical method to find the mutual relationship between the parameters. For example in finding relationship between stock sales and advertisement. Correlation is useful metric in finding which parameter affects the response variable the most. There is many correlation coefficient like pearson to measure linear association, spearman which is not restricted to linear relationship and Kendall coefficient which identifies difference between rank. Spearman correlation coefficient is considered since the data doesn't have linear relationship. It measures monotonic relationship. It considers relative order of the values. The spearman correlation coefficient is given by the formula

$$R_s = \text{COV}(R_x, R_y) / \tau_{R_x} \tau_{R_y}$$

Where

$\text{COV}(R_x, R_y)$  covariance of Rank variables.

$\tau_{R_x} \tau_{R_y}$  are Standard deviations of Rank variables.

The spearman correlation identifies the factor influencing the rainfall most by having high value of  $R_s$ . A positive value means positive relationship. Increase in value of independent variable increase the effect of response variable. A negative value means positive relationship. Increase in value of independent variable decreases the effect of response variable. Classification is a supervised learning technique which uses labeled data for training. Using the learned method classifies the test data into one of the categories of target class. Artificial Neural Network(ANN) is a classification algorithm that works similar to human brain. The components of artificial neural network are neurons which respond to stimulus,

hidden layers which traverses the output of one layer to another using activation function or a sigmoid function and an output layer. ANN can be feed forward and backpropagation where the output is propagated to previous layer. The Input: set of features are temperature, Vapor pressure, PET) the output is Class label is rainfall-average, below average or above average

III. RESULTS AND DISCUSSION

The data for Bangalore urban region (Karnataka, India) is considered for a span of 30 years from 1975 to 2002 collected from Indian Metrological Department (IMD). The features considered are average monthly temperature, vapor pressure, PET (Potential Evapo Transpiration) and rainfall. The raw data from IMD was preprocessed by filling the missing data using mean of previous data. A total of 670 samples were considered. Spearman correlation coefficient was used to find relation between temperature, vapor pressure, PET and rainfall. The result is as shown in Table 1.

	Tempe rature	Vapor Pressure	PET
Rainfa ll	0.2003	0.720040	-0.202

Table 1: Results of Spearman Correlation Coefficient

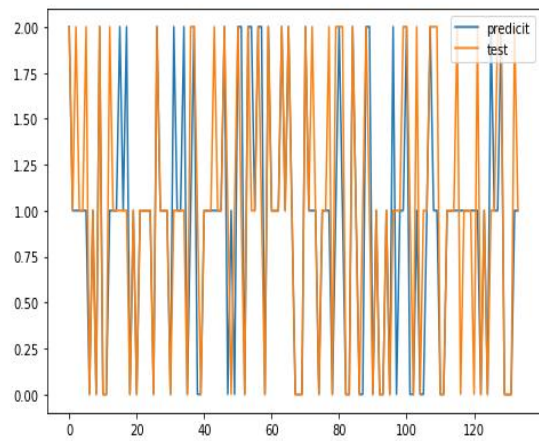
From the Table 1 it is clear that there is a positive correlation between rainfall and temperature, positive correlation between rainfall and vapour pressure, negative correlation between rainfall and PET. The parameter vapour pressure is first factor that is highly influencing the amount of rainfall then the temperature. Lastly PET is affecting negatively for the amount of rainfall.

The classification is done using Artificial Neural Networks. The input variables are temperature, vapor pressure and PET. The response variable or target class is to predict rainfall is below average (less than 10mm), average (between 10mm and 100mm), lastly above average (greater than 100mm). The input variable data is collected for 12 months for period of over 30 years. There are 4 hidden layers each having 15 neurons. The maximum epochs is 1000 which gives accuracy of 75.37%. The precision and recall for the 3 classes are as shown in Table 2.

	Precisio n	Recall	F1 score
Class 0 (<10mm rainfall)	0.82	0.92	0.87
Class 1 (>10mm & <100mm )	0.70	0.73	0.71
Class 2 (>100mm)	0.64	0.51	0.57

Table 2: Results of ANN algorithm.

A graph is plotted for predicted and tested values as shown in figure 1. The accuracy of the ANN algorithm for this dataset is around 76%. The weighted average is 0.72%



The graph depicts the difference in tested and predicted values for class0, class1 and class2 with respect to rainfall.

IV. CONCLUSION

Weather datamining is one of the most challenging problems. Datamining helps in identifying the hidden pattern. Artificial neural networks is one of the supervised learning methodology to identify the target class for the test dataset. Spearman coefficient is used to find correlation in non linear data. This paper uses 30 years data from IMD to perform weather data analysis. Using spearman coefficient we can say that rainfall is positively dependent on vapor pressure and temperature, negatively dependent on PET. ANN uses 4 hidden layers, each consisting of 15 neurons and max epoches of 1000. This results in average precision of 0.80 and average recall of 0.78. The average accuracy is 77%. The future scope is to use the data and perform classification and prediction using RNN-Recurring Neural Network to get more accurate results.

REFERENCES

1. "Rainfall prediction based on 100 year meteorological data", sandeep kumar et.al, 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), feb 2018
2. "Analyze the Rainfall of land slide on Apache Spark" Chou-yann-lee et.al IEEE conference on advanced computer Intelligence, march 2018.
3. "Rainfall prediction of a maritime state (Kerala), India using SLFN and ELM techniques", Yajnaseni Dash, et.al, : 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT) IEEE transactions, April 2018
4. "Early Prediction System Using Neural Network in Kelantan River, Malaysia" Mohd Azrol Syafiee Anuar\* et.al, IEEE conference, 2017
5. "precipitation prediction in shenzhen city based on wnn" sizhe lin1, jianping li1, lingling zhang2, yanan lu1, IEEE 2017
6. "Evaluation of recent prediction models using a long-term database of rainfall rate data", Pedro Garcia-del-Pino et.al, 2017 11th European Conference on Antennas and Propagation (EUCAP) IEEE, may 2017
7. "Rainfall Rate Prediction for Propagation Applications: Model Performance at Regional Level Over Ireland", Loreno luni et.al, IEEE transactions on Antennas and Propagation, September 2017
8. "Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan", Mislana, et.al, Elsevier 2015
9. Mahapatra et al., "A Survey on rainfall prediction using Artificial Neural Network," Int. J. of Comput. Applicat., vol. 72, no. 16, pp. 32-40, 2013.
10. G.B. Huang, M.B. Li, L. Chen, C.K. Siew, "Incremental extreme learning machine with fully complex hidden nodes," Neurocomputing, vol. 71(x), pp. 576-583, 2008.



## AUTHORS PROFILE



**Prajwala T R**, Research scholar in CMRIT Bangalore. Currently working as Professor in PESIT Bangalore. emailid:prajwalatr@gmail.com



**Dr. D Ramesh**, Prof. and Head of Master of Computer Application(MCA), SSIT, Tumkur.



**Dr. H. Venugopal**, Prof. in Department of Computer Science and Engineering [CSE] SSIT Tumkur