iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VII **Month of publication:** July 2019

DOI: http://doi.org/10.22214/ijraset.2019.7080

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VII, July 2019- Available at www.ijraset.com

Different Data Mining Techniques for Rainfall Prediction - A Survey

Deeksha Tripathi[1], Rajendra Kumar Gupta[2]
[1, 2]*Department of Computer Science & Information Technology, MITS, Gwalior (M.P.) 474005, India*

Abstract: ***In this era, one of the most significant and difficult duty is prediction of rainfall. Weather and rainfall in specific are highly non-linear and complicated phenomena that require advanced computer modeling and simulation for their accurate prediction. The conduct of such nonlinear systems can be predicted with an Artificial Neural Network (ANN). For the past thirty years, ANN has been used effectively by most scientists in this sector. In this study, different data mining techniques are investigated. The aim of data mining efforts is usually to produce a descriptive model or predictive model. This article can also describe other best and coherent approaches or forecast techniques in which other study academics, researchers and scientist can conclude this.***

Keywords: ANN, Fuzzy Logic, ARIMA, ASTAR, RMSE, Data mining

I. INTRODUCTION

Data mining relates to information extracted or "mined" from big quantities of information. We understand that gold mining from rocks or sand is not called rock or sand mining, but gold mining. Data mining should therefore have been more properly described as "Knowledge extraction from Data", which is sadly rather lengthy. "Knowledge Mining" is a shorter word from big quantities of information may not indicate the emphasis on mining. Nevertheless, mining is a word that characterizes a method that discovers from a lot of raw material a tiny number of valuable nuggets. The data mining uses complex data structure, Artificial Intelligence and Algorithms. It uses complete to learn from prior understanding and to recognize hidden patterns of information and to deliver realistic outcomes along with rationalization. Knowledge Discovery in the KDD database is synonymous with data mining [1, 2]. On the attention of the scientific community, industries, governments, and risk management entities, were describe the rainfall prediction, which affects many human activities like construction power generation, forestry and tourism, agricultural production, among others. Different types of forecasting the rainfall techniques are provided in India, because India is an agricultural country and the rainfall and humidity is the main factor of agriculture [3].

II. REVIEW METHODOLOGY

Several studies are being conducted in India to determine temperature and rainfall changes and their connection with climate change. However, researchers used distinct lengths of information and now studies have been reported using information over a decade. All such studies showed a country-scale warming trend. It was better to predict rainfall estimates using long-term information sequence. Various papers described to predicting and analysis of monthly rainfall was suggesting different algorithms and methods. In this study, we have critically examined papers numerous publications and conferences preceedings. The papers evaluation criteria discussed in this paper. These are described below:

McCann (1992) [4] implemented the Artificial Neural Network model to provide 3-7 hr prediction of important thunderstorms based on surface-based lifted index and convergence of surface moisture. At the National Severe Storms Forecast Centre, Kansas City, Missouri, they produced two neural networks were operationally combined to implemented a single hourly product and discovered to improve the pattern recognition skills.

Lee et al. (1998) [5] implemented ANN for prediction of rainfall by dividing the accessible information into homogeneous sub-populations. They suggested a divide and conquer strategy, this strategy use four sub regions of the whole area and every sub region is modeled on a distinct technique. Radial basis function (RBF) networks are used to predict rainfall for two bigger fields. A “simple linear regression” model is used for another two narrower sub-regions to forecast the rainfall. A comparison between these two methods are produced by these authors and revealed that great predictions were produced by RBF networks while poor predictions were produced by linear models. For both long term management of contaminated areas and emergency circumstances, these technique was appropriate.

Wong et al. (2003) [6] created a rainfall forecasting model with ANN and Fuzzy Logic soft computing techniques. Firstly, SOM method were used for split the information into sub- population and hopefully. It decreases the insolubility of the entire dataset to

©IJRASET: All Rights are Reserved

something more homogeneous. To know the generalization characteristics, they used BPNNs after classification from the data within each cluster. For each cluster, they evaluated fuzzy rules. The base of the fuzzy rule is used to predict rainfall. They contrasted this technique with an existing technique that utilizes radial-based function networks and effect of orographic. Their outcomes showed that the suggested technique could yield comparable outcomes from the proven technique. However, they disclosed that their technique has the benefit of enabling analysts to comprehend and communicate with the model using fuzzy rules.

Chattopadhyay et al. (2007) [7] developed an Artificial Neural Network which is feed forward model to predict India's average summer monsoon rainfall. Three-layer network has been built with sigmoid non-linearity in formulating the predictive model based on ANN. The total summer rainfall, tropical rainfall indices and sea surface temperature anomalies have been regarded as predictors while producing the ANN input matrix. To the development of the predictive system, the information from the 1950-1995 were investigated. Finally, he contrasted the predictive efficiency of the neural networks with the persistence forecast and the forecasting using multiple linear regression and the supremacy of the ANN over the other process.

D. N. kumar et al. (2007) [8] submitted a monthly and seasonal time-scale Artificial Intelligence strategy to regional rainfall forecasting for the State of Orissa, India. They used an Artificial Neural Networks (ANNs) methodology to manage the extremely non-linear and complicated conduct of climate factors for rainfall prediction. They used Genetic Optimizer (GO) to optimize the architecture of ANN.

Seyed Amir Shamsnia et al. (2011) [9] developed a system for weather parameters including precipitation, monthly temperature and relative humidity using random mechanism (i.e., ARIMA). The authors have been collected 20 years long dataset which is statistics of relative humidity and monthly average temperature in Iran at Abadeh station which is used by Interactive Time Series Modelling (ITSM) analysis software. They designed ARIMA (0 0 1) (1 1 1)12 for precipitation, ARIMA(2 1 0)(2 1 0)12 for monthly average temperature and ARIMA (2 1 1) (1 1 0)12 for relative humidity.

Indrabayu et al. (2013) [10], this paper discusses the potential of statistical method in forecasting rainfall, they introduced two techniques in terms of predictive accuracy i.e., Adaptive Spline Threshold Autoregressive (ASTAR) and Autoregressive Integrated Moving Average (ARIMA). Both techniques are built in the region of Makassar, Indonesia to predict daily rainfall. They used a daily data from BMKG (Meteorology, Climatology and Geophysics) of ten years (2001-2010). LAPAN (Government Space Agent) also obtains several additional information. In Indonesia, rain issue is becoming progressively complicated owing to climatic changes resulting in high-intensity rainfall in the dry season, making it highly important in the growth of many aspects of the social economy industry. Four variables are chosen from multiple meteorological factors to predict rainfall-Temperature, humidity, wind speed and prior precipitation are based on their elevated correlation with rain event. In the ARIMA and ASTAR, These four variables are used. The prediction accuracy is evaluated based on root mean square error (RMSE). ASTAR outperformed ARIMA with less RMSE, 0.02 to 0.24.

Akash D Dubey (2015) [11] implemented various artificial neural networks, were used to predict the rainfall of Pondicherry, India's coastal region. Using three distinct training algorithms, these ANN models are implemented, namely layer recurrent algorithm, Feedforward backpropagation algorithm and Feedforward distributed time delay algorithm. In these ANN models, the number of neurons are kept at 20. For each model, the mean squared error was evaluated and the highest accuracy was acquired by a feed-forward distributed time delay algorithm with an MSE value as low as 0.0083. Total 12 ANN architectures were developed, were trained using the total no. of data points is 800, were gathered over 100 years. The neural networks were validated and tested using 200 data samples after the training process. MSE is used as to measure the performance criteria for neural networks. The MSE has to be accurate for a network.

Archana Nair et al. (2017) [12] applied a nonlinear technique i.e., Artificial Neural Network. Authors has been implemented "Global Climate Models (GCMs)". In this study, GCM are considered from the National Centre for Environmental Prediction (NCEP) and the International Research Institute (IRI). The monthly and seasonal rainfall information had been predicted over the Indian domain of different tropical region. Nowadays, the scientist of the meteorological community are being faced many challenges in the rainfall prediction concerned. In this study, two types of dataset are required: GCM predicted hindcast dataset, these type of dataset are collected from observational dataset and International Research Institute (IRI), these type of dataset are collected from Indian Meteorological Department (IMD). This analysis is developed by using double cross-validation and simple-randomization technique on dataset. The performance of coupled and uncoupled are enhanced the prediction of rainfall of the individual months using the ANN technique.

©IJRASET: All Rights are Reserved

Table 1 provides a comprehensive comparison of the methods of Data Mining.

TABLE 1
Comparison of Different Types of Data Mining Techniques for Rainfall prediction

S. No.	Author(s)	Technique used	Parameter used	Prediction
1.	Kajornrit et al.	Fuzzy-Inference System	Mean Rainfall data	Monthly Rainfall
2.	Lingzhi et al.	Wavelet Support Vector Machines Regression	Temperature, Humidity, Wind speed	Daily Rainfall
3.	Septian et al.	Hybrid Genetic Algorithms & PCFNN	Wind speed, air Pressure, Humidity, length of Sun radiation	Monthly Rainfall
4.	Fhira et al.	Fuzzy system based on Genetic Algorithm	Meteorological and Climatology parameters	Yearly Rainfall
5.	F Mekanik	Linear multiple regression and non linear ANN	Temperature, High humidity, wind speed	Monthly Rainfall
6.	Abhishek et al.	Artificial Neural Network,GPS	Rain fall, snow fall, heat around the earth, surface atmosphere	Daily Rainfall
7.	V. Dabhi et al.	Wavelet postfix-GP model, Wavelet ANN	Maximum Temperature, humidity and Minimum Temperature	Daily Rainfall
8.	M. Sharma et al.	Multiple regression model	Minimum and maximum Temperature, relative Humidity	Monthly Rainfall
9.	S Yeon et al.	Decision tree	Temperature, wind direction, speed, gust, humidity, pressure	Hourly Rainfall
10.	S Badhiye et al.	Lazy learning, Clustering	Sea level, wind speed, snow depth, dew point, rain	Inter annual climate
11.	K. Pabreja	K-mean Clustering	Humidity, Temperature	Rainfall
12.	G. J. Sawale	BPN, Hopfield networks	Wind speed, humidity, temperature	Weather prediction

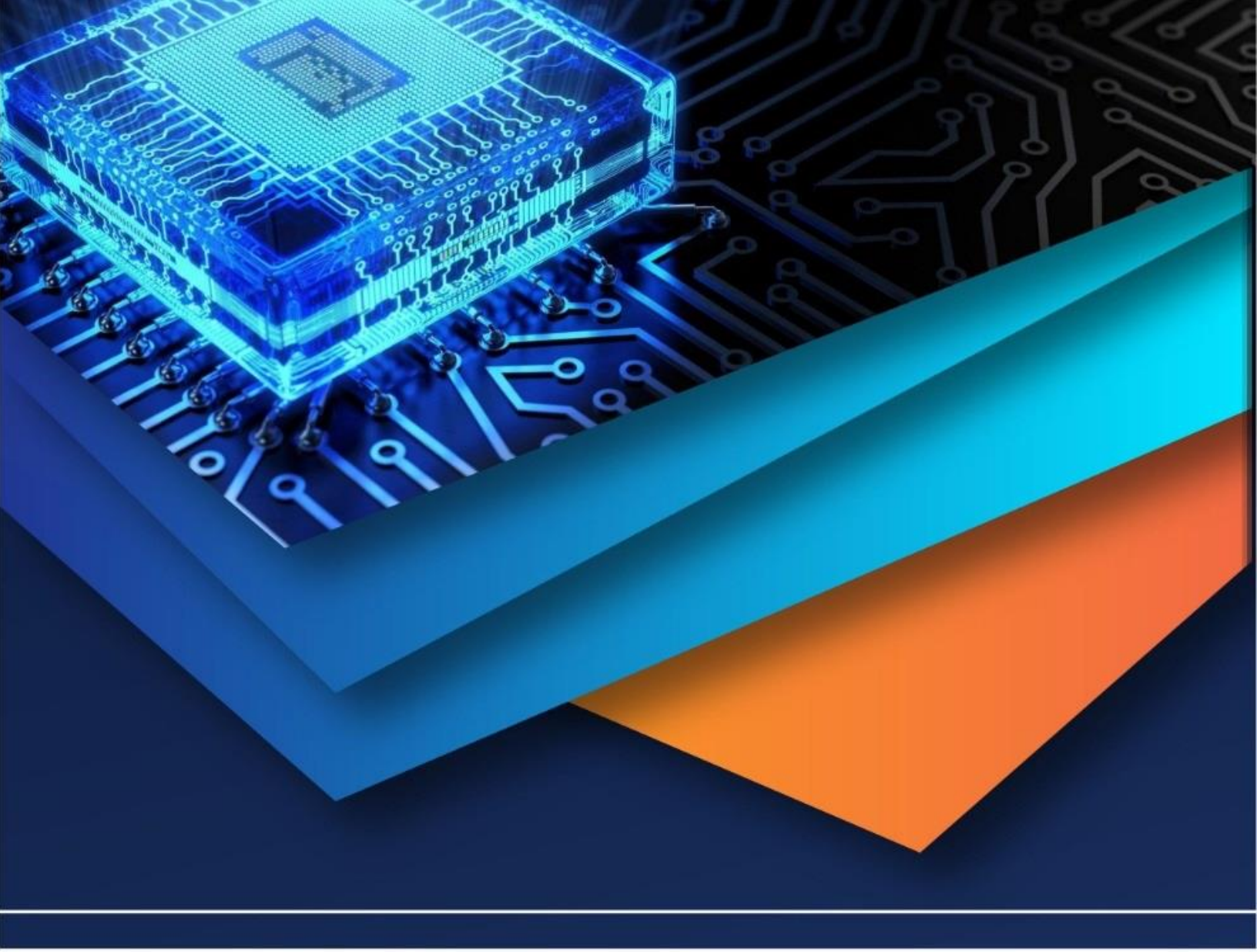
III.CONCLUSIONS

In this paper, different data mining methods are presented for analysis of predict the rainfall. During this research, discusses the ability of different algorithms to predict multiple weather phenomena such as temperature, thunderstorms, rainfall, etc. Because we understand that climate variables change owing to distinct reasons, and here we describe some best techniques for achieving forecast based on empirical studies some climate scientists and research scholars Comparison is made in the other articles, which demonstrates that decision trees and k-mean clustering are best adapted to this implementation for data mining. On the basis of survey, the authors conclude that data mining technique that can be used for forecasting the rainfall.

REFERENCES

[1] Danish Khan "CAKE – Classification, association and discovery of knowledge an approach to distributed data mining (DDM) using parallel data mining agents (PADMA)", IEEE / WIC / ACM international conference on web intelligence and intelligent agent technology, 2008.
[2] Yan Chen1, Ming Yang2, Lin Zhang3, "General Data Mining Model System Based on Sample Data Division", Second International Symposium on Knowledge Acquisition and Modeling, 2009.
[3] M. Kannan, S. Prabhakaran, P. Ramachandran "Rainfall Forecasting Using Data Mining Technique", International Journal of Engineering and Technology Vol.2 (6), pp. 397-401, 2010.
[4] D.W. McCann,"A Neural Network Short-Term Forecast of Significant Thunderstorms", Weather and Forecasting,7,525-534,doi:10.1175/1520-0434, 1992.
[5] S. Lee, S. Cho & P.M. Wong,"Rainfall prediction using artificial neural network", J. Geog. Inf. Decision Anal, 2, 233–242, 1998.
[6] K.W. Wong, P.M. Wong, T. D. Gedeon & C. C. Fung, "Rainfall Prediction Using Soft Computing Technique", Soft Computing 7, pp: 434 – 438, 2003.
[7] S Chattopadhyay and Manojit Chattopadhyay, "A Soft Computing technique in rainfall forecasting", International Conference on IT, HIT, March19-21, 2007.
[8] D.N. Kumar, M.J. Reddy, R. Maity," Regional Rainfall Forecasting using Large Scale Climate Teleconnections and Artificial Intelligence Techniques", Journal of Intelligent System, vol. 16, No. 4, 2007.
[9] Seyed, A., Shamsnia, M., Naeem, S. & Ali, L., "Modelling weather parameter using stochastic methods (ARIMA Model)(Case Study:Abadeh station, Iran)", International conference on Environment and Industrial innovation IPCBEE, 12, 2011.
[10] Indrabayu1, Nadjamuddin Harun2, M. Saleh Pallu3, Andani Achmad4, " Numerical Statistic Approach for Expert System in Rainfall Prediction Based On Data Series" in International Journal of Computational Engineering Research,Vol, 0,Issue, 4,APRIL 2013.
[11] Akash D Dubey," Artificial Neural Network Models for Rainfall Prediction in Pondicherry", International Journal of Computer Applications (0975 – 8887) Volume 120 – No.3, June 2015.
[12] Nair, Archana & Singh, Gurjeet & Mohanty, U C., "Prediction of Monthly Summer Monsoon Rainfall Using Global Climate Models Through Artificial Neural Network Technique", Pure and Applied Geophysics. 175. 10.1007/s00024-017-1652-5, 2017.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)