

Weather Prediction Using Data Mining Techniques: A Systematic Review

^{*1}Azad Abeena & ²Dixit Sarvottam.

^{*1}Mewar University, Ph.D Scholar, Department of Computer Application, Gangrar, Chittorgarh, Rajasthan (India)

²Mewar University, Professor, Department of Computer Science and Engineering, Gangrar, Chittorgarh, Rajasthan (India)

ARTICLE DETAILS

Article History

Published Online: 15 April 2019

Keywords

Data Mining, Knowledge Discovery in Databases (KDD), rainfall prediction, neural networks.

*Corresponding Author

Email: a.abeena@gmail.com

ABSTRACT

Weather prediction plays pivotal role in modern world. Predicting rainfall is vital for water resource management, crop production plan and other thing that are of greater concern. This research work is based on various existing data mining techniques that are used to predict the rainfall accurately. In this work, we have analyzed different existing data mining techniques and data sets that are used by researchers to predict rainfall.

1. Introduction

In today's challenging world organizations and scientists are exploring various methods and approaches to make this world a better place to live for current and future generations. With the evolution and revolution of information technology man has achieved a milestone in last few decades, but there is more to be done in lot of sectors. With the advent of computers, data storage and collection abilities have allowed the accumulation of enormous amounts of data. With the passage of time organizations are accumulating enormous amount of raw data exponentially which can be utilized for future prediction and decision making. Data mining approaches can be implemented in transforming these large amounts of raw data into useful information. Data mining techniques involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets [1]. These tools can include statistical models, mathematical algorithms, and machine learning methods [2]. Machine learning algorithms improve their performance automatically through learning and experience, such as neural networks. Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction. Data Mining is also known as Knowledge Discovery in Databases (KDD), or simply Knowledge Discovery process is the computational process of revealing the meaningful information or patterns that are unknown previously from a underlying data using methods, such as database systems, machine learning, artificial intelligence, and statistics. Predictive modeling is a very popular area in data mining [3]. Data mining techniques and approaches can be used in many applications and have shown significant advantages in many fields as well. Data mining is applied to a number of areas, including engineering field, customer relationship management, market analysis, medical diagnosis and weather prognosis [4]. Weather prediction used to be a challenging task for the meteorologist and scientist in recent past but with the advent of data mining techniques and algorithms, the process of weather prediction is becoming more accurate and precise. Weather prediction helps to prevent climatic hazards and enable scientist and weather department to predict the climate correctly.

Accurate and shrewd weather prediction is useful in commute and construction activities, managing the aviation operations, flood warnings, and water resources management.

Rainfall is one of the varied atmospheric processes, which is based on several weather associated features. Due to climate variations, exact rainfall prediction is more complicated to predict. Various researchers have worked on rainfall prediction using data mining and optimization techniques. Data mining algorithms are categorized into two main types, namely supervised algorithms and unsupervised algorithms. Supervised techniques uses trained with pre-classified data and classify the input data. While Un-supervised techniques do not need training; they utilized the algorithms to extract the hidden structure form unlabelled data [5].

The various methods that were used for weather prediction in general and rainfall prediction in particular are regression, genetic programming, clustering, neural networks, statistical technique of discriminate analysis, adaptive learning approach, classification, Bayesian Belief Networks etc.

2. Review Procedure

To investigate the Rainfall prediction techniques we have studied nearly 150 research papers that have explained and implemented various data mining algorithms for efficient and accurate rainfall prediction. We have identified the most relevant papers based on the following steps:

- Collected the research papers from various sources which have proposed rainfall prediction models by using actual weather data sets.
- We read the Title, Abstract, and Conclusion.
- Based on the content of papers and data set used we have selected the research work that is more relevant to our analysis.
- Out of the 100 papers investigated, 20 most relevant papers were found to suffice our study.
- Each research paper is thoroughly analyzed and

then findings have been reported.

- Finally, table is sketched out to classify the major contribution of the authors, techniques/algorithms and data sets used in various studies.

- RQ1: To identify various data mining techniques used in the rainfall prediction?
- RQ2: To identify the experiments, observations and findings which were done to predict rainfall
- RQ3: To identify various data sets used for the experimentation?

3. Research Questions

**Table-1
Analysis of Research Papers Related To Various Data Mining Techniques Used In Rainfall Prediction**

S.No	Title	Authors	Year of Publication & Source	Algorithm/ Technique Used	Findings & Contribution
1	An improved naive Bayesian classifier technique coupled with a novel input solution method [rainfall prediction].	Liu, James NK, Bavy NL Li, and Tharam S. Dillon	IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) (2001).	Naïve Bayes classifier & Genetic Algorithm.	Proposed An Improved Naïve Bayes Classifier (INCB) Technique For Rainfall Prediction.
2	Rainfall forecasting using data mining technique	Kannan, M., S. Prabhakaran, and P. Ramachandran	International Journal of Engineering and Technology Vol.2 (6), 2010, 397-401 (2010).	Multiple Linear Regression Method	predicted rainfall for future years in ground level
3	Weather forecasting using ANFIS and ARIMA models.	Tektaş, Mehmet	Environmental Research, Engineering and Management 51.1 (2010): 5-10.	Statistical And Neuro-Fuzzy Network	Proposed rainfall prediction model which shows that ANFIS has better predictive capability than ARIMA.
4	Nonlinear ensemble prediction of chaotic daily rainfall	Dhanya, C.T., and Kumar, D.N	Advances in Water resources, vol.33, no.3, pp.327-347, 2010.	Auto-correlation , nearest neighbor algorithm	Developed model maximizes the prediction of rainfall over large area.
5	An efficient weather forecasting system using radial basis function neural network.	Santhanam, Tiruvenkadam, and A. C. Subhajini.	Journal of Computer Science 7.7 (2011):	Back propagation neural network (BPN) and radical basis functioned neural network (RBF)	Accuracy of RBF was better which makes it more useful for fast real time weather forecasting.
6	Modeling rainfall prediction using data mining method: A Bayesian approach	Nikam, Valmik B., and B. B. Meshram	CIMSim, 2013 Fifth International Conference IEEE, 2013.	Bayesian approach	Proposed rainfall prediction model with good accuracy rate and takes moderate compute resources to predict the rainfall.
7	A rainfall forecasting using fuzzy system based on genetic algorithm.	Nhita, Fhira	ICoICT), 2013 International Conference of. IEEE, 2013	using fuzzy system and genetic algorithm (GA).	proposed rainfall forecasting system with combination of GA and Fuzzy System.
8	Rainfall Prediction using Neural Net based Frequency Analysis Approach.	Mahajan, Seema, and Himanshu Mazumdar	International Journal of Computer Applications 84.9 (2013).	artificial neural network (ANN)	proposed prediction model is able to predicted the quantity of rainfall before 1 year.
9	A modular technique for monthly rainfall time series prediction.	Kaiornrit J, Wong, K.W and Fung, C.C	IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environment, pp. 76-83, April 2013.	Mamdani Type Fuzzy Inference System, Non-linear programming, Bayesian Learning	Obtained weights in a sequential fashion for accurate rainfall prediction
10	Data mining for meteorological applications: Decision trees for modeling rainfall prediction.	Geetha, A., and G. M. Nasira	(ICCCIC), 2014 IEEE International Conference.	Decision Trees (DT)	Proposed rainfall prediction model, also highlighted the used of DT to predict other weather phenomena.
11	Prediction of rainfall using data mining technique over Assam	Dutta, Pinky Saikia, and Hitesh Tahbilder	IJCSE 5.2 (2014): 85-90.	Multiple Linear Regression	proposed rainfall prediction model to predict monthly rainfall.

12	Effective Features and Hybrid Classifier for Rainfall Prediction.	KavithaRani, B., and A. Govardhan	International Journal of Computational Intelligence Systems 7.5 (2014): 937-951.	Artificial bee colony algorithm & genetic algorithm	Rainfall prediction model, a hybrid classifier, which has better accuracy.
13	Bayesian network model for monthly rainfall forecast.	Sharma, Ashutosh, and Manish Kumar Goyal	(ICRCICN), 2015 IEEE International Conference	Bayesian network	Proposed Bayesian network model for mean monthly rainfall prediction of 21 stations in Assam, India.
14	Artificial neural network models for rainfall prediction in Pondicherry	Dubey, Akash D	International Journal of Computer Applications (2015)	artificial neural networks (ANN)	Proposed rainfall prediction model with good accuracy rate.
15	Prediction of daily rainfall by a hybrid wavelet-season-neuro technique	Altunkaynak, A. and Nigussie, T.A.	Journal of Hydrology, vol.529, pp.287-301, 2015.	Multilayer Perceptrons	Proposed model showed lead time of prediction of daily rainfall upto 5 days with enhanced accuracy.
16	A Support Vector Regression Model For Forecasting Rainfall.	Hasan, N., Nath, N.C. and Rasel, R.I	2 nd International Conference On Electrical Information And Communication Technology, Pp. 554-559, December 2015.	Regression Based	Developed a robust rainfall prediction model.
17	Prediction Of Meiyu Rainfall In Taiwan By Multi-Lead Physical-Empirical Models.	Yim, S.Y., Wang, B., Xing, W And Lu, M.M	Climate Dynamics, Vol.44, No.11, Pp.3033-3042, 2015.	Dynamical Models	Proposed model predicts the amount of rainfall for water resource management.
18	Daily Rainfall Prediction Using Generalized Linear Bivariate Model-A Case Study.	George, J., Letha, J. And Jairaj, P.G	Procedia Technology, Vol.24, Pp.31-38, 2016.	Logic Regression And Gamma Distributions.	Developed a system that is utilized as a weather generator for simulating the future rainfall.
19	Gamma Distribution Linear Modeling With Statistical Downscaling To Predict Extreme Monthly Rainfall in Indramayu.	Soleh, A.M., Wigena, A.H., Djuraidah, A. And Saefuddin, A	12th International Conference On Mathematics, Statistics, And Their Applications, Pp. 134-138, October 2016.	Linear Model And Gamma Distribution	The Proposed Model Predicts Extreme Monthly Rainfall., method failed to produce the smallest RMSE values.
20	Performance Improvement of a Rainfall Prediction Model using Particle Swarm Optimization	Islam, S and Talukdar, B	International Journal of Computational Engineering Research, vol.6, no.7, pp.39-42, 2016	PSO, Autoregression model.	Proposed model for annual rainfall prediction shows good performance as compared to the statistical ARIMA Model.

4. Related Work

In [6], Liu et al, proposed a rainfall prediction technique that implements an improved Naïve Bayes classifier (INCB) in combination with genetic algorithms (GAs) which is used for feature selection. A comparison of genetic algorithms, C4.5, and INBC with relative frequency or initial probability density is carried out on real weather data of Hong Kong and results revealed that the proposed model has better accuracy rate of 90% for classification problems. Moreover the results also showed that INBC has appreciable predictive ability to predict rainfall range levels with 65%-70% of accuracy rate.

In [7], Kannan et al, proposed a rainfall forecast model by using regression methods. Five years of historical

weather data from Statistical department of Tamil Nadu, Chennai, India is collected by the authors for this work. The authors have used Karl Pearson correlation to calculate values for rainfall in the ground level which is then implemented for rainfall prediction for future years in ground level by using multiple linear regression technique.

In [8], Mehmet Tektaş, collected nine years weather data of Göztepe, İstanbul, Turkey from 2000-2008. The author performed a comparative analysis of statistical and neuro-fuzzy network models for weather forecasting by employing Adaptive Network Based Fuzzy Inference System (ANFIS) and Auto Regressive Moving Average (ARIMA) models on the collected data. The experimental results illustrate that ANFIS has better predictive capability than ARIMA.

In [9], Dhanya, C.T., and Kumar, D.N, performed an experiment for identifying the chaos utilizing different techniques. Authors have used nonlinear ensemble prediction of chaotic daily rainfall by using real weather data. Auto-correlation and mutual information methods are utilized for determining the delay time. The authors also showed that the spatial averaging over a large area maximizes predictability, so the shifting of the system is a non-chaotic one.

In [10], Santhanam et al, collected 10 year weather data from meteorological department, Kanyakumari, Tamil Nadu, India. The authors proposed rainfall prediction model based on Neural networks. Implementation and comparison of performances of Back propagation neural network (BPN) and radical basis functioned neural network (RBF) is done on collected weather data. The experimental results revealed that radical basis functioned neural network (RBF) has better accuracy. The predictive accuracy of RBF was found to be 88.49% which is appreciable for weather forecasting.

In [11], Valmik et al, used historical weather data set of Indian Meteorological Department (IMD) Pune, India, to propose a rainfall prediction model based on Bayesian network approach which can be deployed on commodity hardware without the need of high-performance computers. The experimental results show that the Bayesian approach used for this research has appreciable level in predicting rainfall for future.

In [12], Nhita et al, collected and pre-processed weather data from Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG) for Kemayoran area, Jakarta. The authors proposed rainfall prediction system based on fuzzy system and genetic algorithm (GA). The experiments carried out on the data set revealed that by applying combination of GA and Fuzzy System techniques for Kemayoran weather data, a prediction model produced has more than 90% accuracy and has better rainfall prediction capability.

In [13], Mahajan et al, proposed a rainfall prediction model based on artificial neural network and frequency analysis approach using Fast Fourier Transform. The authors collected the weather data of 30 sub divisions from IITM, Pune, India. The proposed rainfall prediction model can predict the amount of rainfall before 1 year which can be helpful for farmers and government to issues alerts and warnings.

In [14], Jesada Kajornrit et al, designed a prediction model based on a modular technique which implements Mamdani Type Fuzzy Inference System, Non-linear programming and Bayesian Learning approaches for monthly time series rainfall prediction. The authors have used uses eight patterns of monthly rainfall data from northeast Thailand. The experimental result shows that the sequential method shows better rainfall prediction accuracy than the non-sequential method.

In [15], Geetha et al, weather dataset is collected from Trivandrum, India. The authors proposed rainfall prediction

model based on decision tree. The designed model predicts weather phenomena like fog, rainfall, cyclones and thunderstorms. Experimental results show that success rate is 80.67%, when compared with the target data.

In [16], Dutta et al, collected and pre-processed weather data of six years from 2007-2012 from Regional Meteorological Center, Guwahati, Assam, India. The authors designed a rainfall prediction model based on Multiple Linear Regression technique that allows predicting monthly rainfall. The experiments performed shows that the performance of proposed model measured by R-squared is quite acceptable and have good accuracy in rainfall prediction.

In [17], KavithaRani et al, collected and cleaned weather data from Rayalaseema, Aandhra and Telangana regions of India. The authors designed rainfall prediction model by implanting amalgam classifier that used artificial bee colony algorithm is association with the genetic algorithm. The hybrid classifier developed is used to predict rainfall. The experimental result shows that the proposed prediction model (hybrid classifier) has better performance and predictive ability than Artificial Bee Colony algorithm.

In [18], Sharma et al, gathered and pre-processed monthly weather data of 20 years from 1981 to 2000 of 21 stations in Assam, India. The authors proposed a Bayesian network based model for monthly rainfall prediction. The authors used 5 different atmospheric parameters namely Temperature, Cloud cover, Relative humidity, Wind speed and Southern Oscillation Index (SOI). The experimental results shows that temperature is found most efficient and wind speed least. It was found that some station showed efficiency above 95% whereas other stations showed less.

In [19], Akash D Dubey, collected weather data set of Pondicherry, India for this work. The authors designed a model based on artificial neural networks (ANN) to predict rainfall. The ANN model is developed with 20 neurons by using three different training algorithms namely feed-forward back propagation algorithm, layer recurrent algorithm and feed-forward distributed time delay algorithm. Experiment results showed that that feed-forward distributed time delay algorithm has best accuracy in predicting rainfall.

In [20], Altunkaynak, A. and Nigussie, T.A., designed two rainfall prediction model to predict rainfall and extending the lead time of prediction of daily rainfall up to 5 days. The authors used weather data of two stations in Turkey. In this work season multilayer perceptrons (SAS-MP) and hybrid wavelet-season-multilayer perceptrons (W SAS-MP) were developed to enhance rainfall prediction. The performance measures used were mean squared error, the coefficient of determination and coefficient of efficiency.

In [21], Nasimul Hasan, accrued and pre-processed rainfall data of Bangladesh. The authors designed the robust rainfall prediction model by implementing Support Vector Regression (SVR) technique. The experimentation result shows appraisable level of accuracy of proposed rainfall

model. However, more data processing and cleaning is needed to refine the data for better results and accuracy.

In [22], Yim, S.Y et al., developed rainfall prediction based on a physical–empirical model. Weather data of Taiwan Meiyu is collected. The model predicts the amount of Meiyu rainfall which helps in water resource management. The dynamic physical methods are used to develop the prediction model which has appreciable prediction capability.

In [23], George, J et al., developed a rainfall prediction model that was implemented on the weather dataset collected from Kerala, India. The model is named as Bivariate Generalized Linear Model which simulating daily rainfall time series data using atmospheric predictors and historical data by using bivariate Generalized Linear Model. Gamma distributions are used for rainfall modeling on a rainy day. The proposed models are used as weather generators for simulating and predicting future rainfall.

In [24], Soleh, A.M. et al., used rainfall data of Indramayu, India at 11 rain stations. The authors designed prediction method for predicting extreme monthly rainfall. The experimentation result shows that 11 regularization technique with principal component analysis implemented and compared to predict the response. The technique was not able to get the minimum RMSE value.

In [25], Islam, S. and Talukdar, B., developed a rainfall prediction model that predicts time series rainfall. The authors implemented Particle Swarm Optimization (PSO)

technique for producing reliable results. Experimental work shows that when model is used for annual rainfall prediction, it provided good performance and accuracy as compared to the statistical ARIMA model. However more research and use of hybrid models can increase the accuracy rate of rainfall prediction.

5. Conclusion and Future Scope

In this research review work, we have tried to identify and explain the questions RQ1, RQ2 and RQ3 thoroughly. The main objective of this systematic review is to identify various data mining prediction techniques like regression, genetic programming, clustering, neural network, statistical technique, optimization techniques, hybrid approach, physical modeling approaches, classification, Bayesian Belief Networks etc that are used for rainfall prediction. In this work, we have seen the most of the weather data sets used by the researchers for prediction are real and taken from actual geographical location. Based on our study, further research can be extended to propose new method for better prediction of rainfall.

Acknowledgement

We would like to express our gratitude to Mewar University, Chittorgarh (Rajasthan), India, for granting the valuable opportunities and facilities for research. We are also highly grateful to our Director and HOD, Computer Science and Engineering for providing their effective feedback and constant encouragement for the research.

References

- Pieter Adriaans and Dolf Zantinge., "Two Crows Corporation, Introduction to Data Mining and Knowledge Discovery", Third Edition (Potomac, MD: Two Crows Corporation, 1999); Data Mining (New York: Addison Wesley, 1996).
- Seifert, Jeffrey W. "Data mining and homeland security: An overview." LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE, 2008.
- Laura Patterson, "Nine Common Types of Data Mining Techniques Used in Predictive Analytics", <http://www.visionedgemarketing.com/>.
- Mourad LASFAR, and Halima BOUDEN, "A method of data mining using Hidden Markov Models (HMMs for protein secondary structure prediction)", *Procedia Computer Science*, vol.127, pp.42-51, 2018.
- Shabib Aftab, Munir Ahmad, Noreen Hameed, Muhammad Salman Bashir, Iftikhar Ali, and Zahid Nawaz, "Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp. 143-150, 2018.
- Liu, James NK, Bavy NL Li, and Tharam S. Dillon. "An improved naive Bayesian classifier technique coupled with a novel input solution method [rainfall prediction]." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 31.2 (2001): 249-256.
- Kannan, M., S. Prabhakaran, and P. Ramachandran. "Rainfall forecasting using data mining technique." *International Journal of Engineering and Technology* Vol.2 (6), 2010, 397-401 (2010).
- Tektaş, Mehmet. "Weather forecasting using ANFIS and ARIMA models." *Environmental Research, Engineering and Management* 51.1 (2010): 5-10.
- Dhanya, C.T. and Kumar, D.N, "Nonlinear ensemble prediction of chaotic daily rainfall," *Advances in Water resources*, vol.33, no.3, pp.327-347, 2010.
- Santhanam, Tiruvenkadam, and A. C. Subhajini. "An efficient weather forecasting system using radial basis function neural network." *Journal of Computer Science* 7.7 (2011): 962.
- Nikam, Valmik B., and B. B. Meshram. "Modeling rainfall prediction using data mining method: A Bayesian approach." *Computational Intelligence, Modelling and Simulation (CIMSIm)*, 2013 Fifth International Conference on. IEEE, 2013.
- Nhita, Fhira. "A rainfall forecasting using fuzzy system based on genetic algorithm." *Information and Communication Technology (ICoICT)*, 2013 International Conference of. IEEE, 2013.
- Mahajan, Seema, and Himanshu Mazumdar. "Rainfall Prediction using Neural Net based Frequency Analysis Approach." *International Journal of Computer Applications* 84.9 (2013).
- Kaiornrit J, Wong, K.W and Fung, C.C, "A modular technique for monthly rainfall time series prediction," In proceedings of IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environment ,pp. 76-83, April 2013.

15. Geetha, A., and G. M. Nasira. "Data mining for meteorological applications: Decision trees for modeling rainfall prediction." *Computational Intelligence and Computing Research (ICCIC)*, 2014 IEEE International Conference on. IEEE, 2014.
16. Dutta, Pinky Saikia, and Hitesh Tahbilder. "Prediction of rainfall using data mining technique over Assam." *IJCSE* 5.2 (2014): 85-90.
17. KavithaRani, B., and A. Govardhan. "Effective Features and Hybrid Classifier for Rainfall Prediction." *International Journal of Computational Intelligence Systems* 7.5 (2014): 937-951.
18. Sharma, Ashutosh, and Manish Kumar Goyal. "Bayesian network model for monthly rainfall forecast." *Research in Computational Intelligence and Communication Networks (ICRCICN)*, 2015 IEEE International Conference on. IEEE, 2015.
19. Dubey, Akash D. "Artificial neural network models for rainfall prediction in Pondicherry." *International Journal of Computer Applications* 120.3 (2015).
20. Altunkaynak, A. and Nigussie, T.A., "Prediction of daily rainfall by a hybrid wavelet-season-neuro technique," *Journal of Hydrology*, vol.529, pp.287-301, 2015.
21. Hasan, N., Nath, N.C. and Rasel, R.I., "A support vector regression model for forecasting rainfall," In proceedings of 2nd International Conference on Electrical Information and Communication Technology, pp. 554-559, December 2015.
22. Yim, S.Y., Wang, B., Xing, W and Lu, M.M., "Prediction of Meiyu rainfall in Taiwan by multi-lead physical-empirical models," *Climate Dynamics*, vol.44, no.11, pp.3033-3042, 2015.
23. George, J., Letha, J. and Jairaj, P.G, "Daily rainfall prediction using generalized linear bivariate model-A case study," *Procedia Technology*, vol.24, pp.31-38, 2016.
24. Soleh, A.M., Wigena, A.H., Djuraidah, A. and Saefuddin, A., "Gamma distribution linear modeling with statistical downscaling to predict extreme monthly rainfall in Indramayu," In proceedings of 12th International Conference on Mathematics, Statistics, and Their Applications, pp. 134-138, October 2016.
25. Islam, S and Talukdar, B, "Performance Improvement of a Rainfall Prediction Model using Particle Swarm Optimization," *International Journal of Computational Engineering Research*, vol.6, no.7, pp.39-42, 2016.