Indian Journal of Data Mining (IJDM)
ISSN: 2582-9246 (Online), Volume-3 Issue-2, November 2023

Comparison of the Proposed Rainfall Prediction Model Designed using Data Mining Techniques with the Existing Rainfall Prediction Methods

Deepak Sharma, Priti Sharma

Abstract: ***Weather prediction is a very old practice and people are doing predictions about weather much before the discovery of the weather measuring instrument. In ancient times, people give weather predictions by observing the sky for a long time and patterns of the stars at night. Things are a bit different now. People more relay on the past trends and patterns followed by the weather parameters. Data mining and machine leaning is used to analysis the historical weather trends by analyzing weather data using various Data mining techniques. In this paper three rainfall prediction model based on data mining techniques are proposed and compared with the other rainfall prediction model. The comparison has been done on the basis of accuracy, precision, Recall and RMSE. The proposed models are based on ensemble methods such as bagging, boosting, and stacking. Ensemble methods are used to enhance the overall performance and accuracy of the prediction. In both bagging and boosting based proposed rainfall prediction models, artificial neural network is used as a base leaner and daily weather data from the year 1988 to 2022 is used. In stacking based proposed rainfall prediction model, random forest, Logistic regression, and K-Nearest neighbor are used as base leaners or level -0 learners and Artificial neural network is used as Meta model.***

Keywords: ***Data Mining, Data Collection, Secondary Data Sources, Weather Data, Rainfall Prediction, Machine Learning.***

I. INTRODUCTION

Data mining is a process of knowledge discovery from large amount of data. [1] It is not a very old concept and evolved significantly from last two decades. With evaluation of hardware and introduction of integrated circuits (IC's) in the computer, processing power of the machine has increased exponentially. [2] The introduction of LSI (Large scale integration) and VLSI (Very large-scale integration) increased the ability of machine up to a level which will allow the user to process large amount of data. [3]. The process of knowledge discovery starts with collecting the related data. It can be done in two ways. Data can be either collected directly for the process called primary data or already collected data can be obtained from a data source called as secondary data. [4]

Manuscript received on 15 May 2023 | Revised Manuscript received on 22 May 2023 | Manuscript Accepted on 15 November 2023 | Manuscript published on 30 November 2023.
*Correspondence Author(s)
Deepak Sharma***,** DCSA, Maharshi Dayanand University, Rohtak (Haryana), India. E-mail: erdeepaksharmabwn@gmail.com, ORCID ID: https://orcid.org/0000-0002-7490-557X
Dr. Priti Sharma, DCSA, Maharshi Dayanand University, Rohtak (Haryana), India. E-mail: pritish80@yahoo.co.in

The collected data is preprocessed before using it for the analysis. In preprocessing of data, missing values and outlier analysis has been done. The preprocessed data has been sent for analysis using data mining techniques. [5] After the analysis the result will be in form of meaning patterns and relationship between attributes. This complete process has been shown in figure 1. [6]

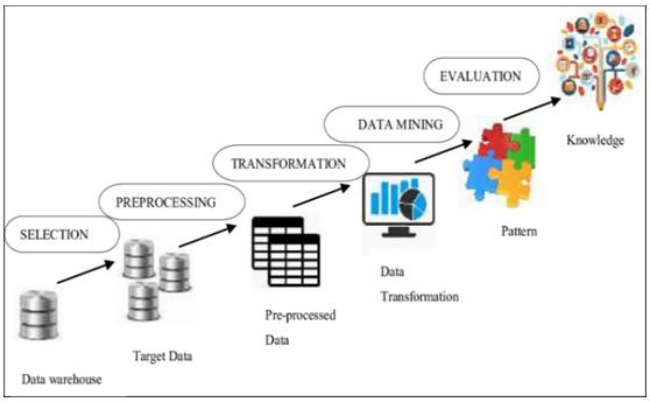


Figure 1: Process of Knowledge discovery (KDD)

II. PROPOSED RAINFALL PREDICTION MODELS

In this work, three rainfall prediction model based on ensemble-based data mining techniques are proposed. The bagging-based rainfall prediction model (RPM-BAG) has been shown in figure 2. [7] In this model, bagging ensemble data mining technique is used. Artificial neural network has been used as base learner. [8] A total of 20 ANNs are used to design the model and all are trained in parallel. [9]

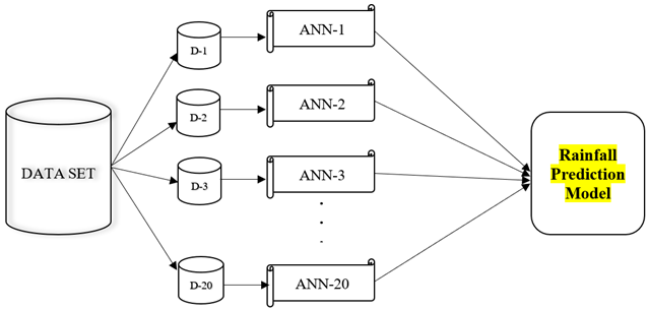


Figure 2: Rainfall prediction model-based Bagging (RPM)

The second proposed rainfall prediction model is based on boosting ensemble method which is shown in figure 3. [10] In boosting ensemble method, a base learner trained multiple times in which output of one become input for next and so on.

[11] In this way boosting method increase the overall accuracy and precision of the base learner. Here, the artificial neural network is taken as the base leaner.

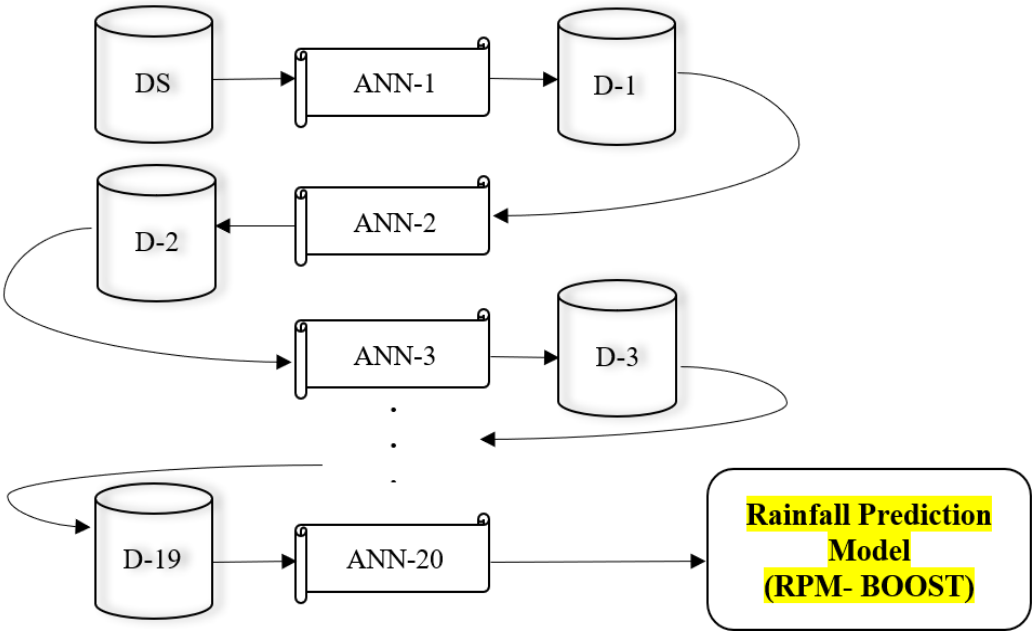


Figure 3: Rainfall prediction model-based Boosting (RPM)

The third proposed model is based on stacking ensemble data mining technique which is shown in figure 4. In stacking ensemble data mining technique, the weak learners which are also called Level-0 learners are trained using the original data set. [12] The prediction output of these level -0 learners is used as training set of level -1 learner also called as meta model. In this staking-based rainfall prediction model, random forest, logistic regression and K-nearest neighbor are used as level-0 learners and artificial neural network is used as level -1 learner or meta model. [13] These models are compared with each other and then compared with already existing models for rainfall prediction using data mining techniques. [14]

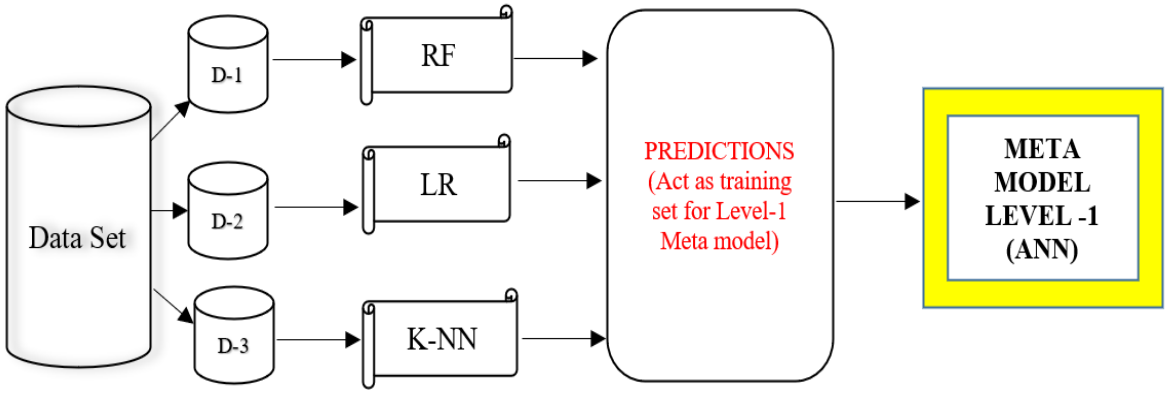


Figure 4: Rainfall prediction model-based on Stacking (RPM)

III. COMPARISON WITH THE EXISTING MODELS

The proposed models and existing models are compared based on accuracy, precision, recall and RMSE in this section. [15]

Table 1: Comparison of proposed rainfall prediction model

S. No.	PARAMETERS	RPM-BAG	RPM-BOOST	RPM-STACK
1	ACCURACY	89.14 %	88.53 %	88.82 %
2	PRECISION	79.87 %	77.9 %	78.62 %
3	RECALL	71.08 %	71.3 %	70.88 %
4	RMSE	0.284	0.315	0.295

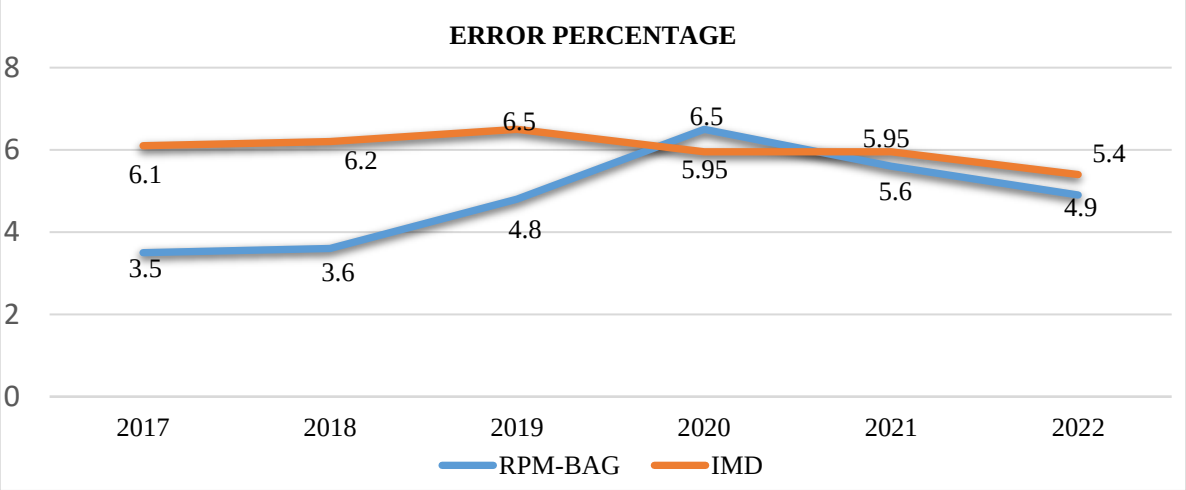
Table 2: Comparison of existing rainfall prediction model

S. No.	PARAMETERS	KC (ML-ANN) [8]	MM (SVR) [13]	DP (MLR, ANN) [14]
1	Accuracy	81.2 %	84.3 %	87.24 %
2	Precision	76.2 %	74.3 %	77.1 %
3	Recall	69.2 %	68.3 %	70.2 %
6	RMSE	0.446	0.375	0.327

Table 3: Comparison of proposed rainfall prediction model and existing models

S. No.	PARAMETERS	PROPOSED MODELS			EXISTING MODELS		
		RPM-BAG	RPM-BOOST	RPM-STACK	KC (ML-ANN) [8]	MM (SVR) [13]	DP (MLR, ANN) [14]
1	Accuracy	89.14 %	88.53 %	88.82 %	81.2 %	84.3 %	87.24 %
2	Precision	79.87 %	77.9 %	78.62 %	76.2 %	74.3 %	77.1 %
3	Recall	71.08 %	71.3 %	70.88 %	69.2 %	68.3 %	70.2 %
4	RMSE	0.284	0.315	0.295	0.446	0.375	0.327

IV. COMPARISON WITH OTHER PREDICTION METHODS



V. CONCLUSION AND FUTURE SCOPE

In this work three rainfall prediction model based on ensemble data mining technique are proposed and compared with the already existing prediction methods. [16] The comparison has been done on the basis of accuracy, precision, recall and RMSE (Root mean square error). [17] Out of three proposed rainfall prediction models bagging based prediction models performs the best with accuracy as 89.14 %, precision as 79.87 %, recall as 71.08 % and root mean square error as 0.284. [18] Also, the performance of the proposed model is better as compared to the existing model as shown in the comparative analysis section. [19]

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive.
Conflicts of Interest/ Competing Interests	The article bears No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. Tharun V.P, Ramya Prakash, S. Renuga Devi, "Prediction of Rainfall Using Data Mining Techniques", 2nd International Conference on Inventive Communication and Computational Technologies, IEEE-2018. [CrossRef]
2. Abishek.B, R.Priyatharshini, Akash Eswar M, P.Deepika, "Prediction of Effective Rainfall and Crop Water Needs using Data Mining Techniques", International Conference on Technological Innovations in ICT For Agriculture and Rural Development, IEEE-2017. [CrossRef]
3. Fahad Sheikh, S. Karthick, D. Malathi, J. S. Sudarsan, C. Arun, "Analysis of Data Mining Techniques for Weather Prediction", Indian Journal of Science and Technology, Vol 9(38), ISSN (Print): 0974-6846, IJST-2016. [CrossRef]
4. Ramsundram N, Sathya S, Karthikeyan S, "Comparison of Decision Tree Based Rainfall Prediction Model with Data Driven Model Considering Climatic Variables", Irrigation Drainage Sys Eng, an open access journal ISSN: 2168-9768, 2016.
5. Bhaskar Pratap Singh, Pravendra Kumar, Tripti Srivastava, Vijay Kumar Singh, "Estimation of Monsoon Season Rainfall and Sensitivity Analysis Using Artificial Neural Networks", Indian Journal of Ecology (2017) 44 (Special Issue-5): 317-322.
6. Nikhil Sethi, Dr.Kanwal Garg, "Exploiting Data Mining Technique for Rainfall Prediction", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 3982-3984.
7. Chandrasegar Thirumalai, M Lakshmi Deepak, K Sri Harsha, K Chaitanya Krishna, "Heuristic Prediction of Rainfall Using Machine Learning Techniques", International Conference on Trends in Electronics and Informatics - ICEI 2017. [CrossRef]

8. Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong, "Predicting Rice Crop Yield Using Bayesian Networks", Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India. [CrossRef]
9. K C Gouda, Libujashree R,Priyanka Kumari,Manisha Sharma, Ambili D Nair, "An Approach for Rainfall Prediction using Soft Computing" International Journal of Engineering Trends and Technology (IJETT) – Volume 67 Issue 3 - March 2019 ISSN: 2231-5381. [CrossRef]
10. Moulana Mohammed, Roshitha Kolapalli, Niharika Golla, Siva Sai Maturi, "Prediction Of Rainfall Using Machine Learning Techniques" International Journal of Scientific & Technology Research Volume 9, Issue 01, January 2020 ISSN: 2277-8616.
11. Deepali Patil, Shree L.R. Tiwari, Abhishek Jain, Shree L.R. Tiwari, Aniket Gupta, "Rainfall Prediction using Linear approach & Neural Networks and Crop Recommendation based on Decision Tree" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 9 Issue 04, April-2020. [CrossRef]
12. Sudha Mohankumar and Valarmathi Balasubramanian, "Identifying Effective Features and Classifiers for Short Term Rainfall Forecast Using Rough Sets Maximum Frequency Weighted Feature Reduction Technique", Journal of Computing and Information Technology, Vol. 24, No. 2, June 2016, 181–194 DOI: 10.20532/cit.2016.1002715 [CrossRef]
13. Balamurali Ananthanarayanan, Siva Balan, Anu Meera Balamurali and Karthika Balamurali, "Efficient Dissemination of Rainfall Forecasting to Safeguard Farmers from Crop Failure Using Optimized Neural Network Model" International Journal of Intelligent Engineering and Systems, Vol.10, No.1, 2017 DOI: 10.22266/ijies2017.0228.05. [CrossRef]
14. Suvidha Jambekar, Shikha Nema, Zia Saquib, "Prediction of Crop Production in India Using Data Mining Techniques", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). [CrossRef]
15. Wassamon Phusakulkajorn, Chidchanok Lursinsap, Jack Asavanant, "Wavelet-Transform Based Artificial Neural Network for Daily Rainfall Prediction in Southern Thailand", ISCIT 978-1-4244-4522-6/09 2009 IEEE.
16. N. Tyagi and A. Kumar, "Comparative analysis of backpropagation and RBF neural network on monthly rainfall prediction," Proc. Int. Conf. Inven. Comput. Technol. ICICT 2016, vol. 1, 2017.
17. N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "A Comprehensive Survey of Data Mining Techniques on Time Series Data for Rainfall Prediction," J. ICT Res. Appl., vol. 11, no. 2, p. 168, 2017. [CrossRef]
18. Deepak Sharma, Dr. Priti Sharma, "Rain Fall Prediction using Data Mining Techniques with Modernistic Schemes and Well-formed Ideas", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN (Online): 2278-3075, Volume-9 Issue-1, 2019, Page no. 258-263. [CrossRef]
19. Deepak Sharma, Dr. Priti Sharma. "Rainfall Prediction Using Classification and Clustering Complex Data Science Models with Geological Significance". International Journal of Computer Science Trends and Technology (IJCST) V8 (5): Page (39-44) Sep - Oct 2020. ISSN: 2347-8578. www.ijcstjournal.org.Published by Eighth Sense Research Group

AUTHORS PROFILE

Deepak Sharma has completed his M.tech from C-DAC: Centre for Development of Advanced Computing, Ministry of Communications and Information Technology, Government of India affiliated from Guru Gobind Singh Indraprastha University, Delhi. He is currently pursuing a Ph.D. in Computer Science at M. D. University, Rohtak. His main research areas include Data mining, Mobile Adhoc Network (MANET), wireless sensor network (WSN) and Internet of things (IoT).

Dr. Priti Sharma MCA, Ph.D. (Computer Science) is working as an Assistant Professor in the Department of Computer Science & Applications, M.D. University, Rohtak. She has published more than 50 publications in various journals/ magazines of national and international repute. She is engaged in teaching and research from the last 15 years. Her area of research includes Data mining, Big data, Software Engineering, Machine Learning.