

A Literature Review on Rainfall Prediction using different Data Mining Techniques

Ankit Rasam¹, Damini Pandare², Shruti Narkar³, Charmi Chaniyara⁴

¹B.E Student, Department of IT, Atharva College of Engineering, Maharashtra, India

²BE Student, Department of IT, Atharva College of Engineering, Maharashtra, India

³BE Student, Department of IT, Atharva College of Engineering, Maharashtra, India

⁴Professor, Department of IT, Atharva College of Engineering, Maharashtra, India

Abstract - Rainfall prediction is one of the difficult assignments in weather forecasting. Precise and ideal rainfall prediction can be extremely useful to take successful safety efforts ahead of time in regards to: continuous development projects, transportation exercises, horticultural assignments, flight tasks and flood circumstance, and so on Information mining procedures can successfully anticipate the rainfall by extracting the hidden patterns among accessible components of past climate information. This examination contributes by giving a basic investigation and survey of most recent information mining methods, utilized for precipitation expectation. Distributed papers by S. Zhang, L. Lu, J. Yu, H. Zhou, S. Zainudin, D. S. Jasim, A. A. Bakar, R. Venkata Ramana, B. Krishna, S. R. Kumar, N. G. Pandey, D. Nayak, A. Mahapatra, P. Mishra, S. Cramer, M. Kampouridis, A. A. Freitas, A. K. Alexandridis, Aftab, Shabib & Ahmad, Munir & Hameed, Noureen & Bashir, Salman & Ali, Iftikhar & Nawaz, Zahid were considered for this examination. This survey will fill in as an audit to examine the most recent work on precipitation expectation with the emphasis on information mining methods and furthermore will give a pattern to future headings and examinations.

Key Words: Rainfall prediction, data mining techniques, literature review, Wavelet Neural Network Model (WNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), Adaptive Neuro Fuzzy Inference System (ANFIS)

1. INTRODUCTION

Examination of time series information is one of the significant parts of present day research in the space of information disclosure. Time series information is gathered throughout a particular timeframe like hourly, every day, week after week, month to month, quarterly or yearly. Information mining procedures can utilize this information to foresee forthcoming circumstances in different areas, for example, environmental change, training, and money and so on These procedures can be utilized to extricate concealed information from time series information for some time later. Climate determination is extremely gainful however testing tasks. Climate information comprises of different barometrical provisions, for example, wind speed, stickiness,

tension and temperature and so forth Information mining procedures have the ability to extricate the secret examples among accessible provisions of past climate information and afterward these strategies can anticipate future climate conditions by utilizing separated examples. Precipitation is a complex climatic interaction, which relies on many climate related components. Exact and ideal precipitation expectation can be useful in numerous ways, for example, arranging the water assets the executives, issuance of early flood alerts, dealing with the flight tasks and restricting the vehicle and development exercises. Exact precipitation expectation is more mind boggling today because of environmental varieties. Specialists reliably have been attempting to anticipate precipitation with greatest precision by streamlining and coordinating information mining strategies. Information mining calculations are delegated, administered and un-regulated. Regulated strategies get prepared first with pre-characterized information (training data) and afterward arrange the info information (test information). Un-administered strategies then again don't need any preparation, rather than pre-characterized information these methods use calculations to extricate stowed away design structure non labeled information.

2. RELATED WORKS

Scientists have been attempting to work on the exactness of precipitation expectation by enhancing and incorporating information mining strategies. A portion of the chosen studies are examined in this segment. In [1], the creator played out a relative investigation of Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Adaptive Neuro Fuzzy Inference System (ANFIS) on rainfall prediction.

The creators have analyzed the forecast models in four terms:

- (I) by utilizing various slacks as demonstrating inputs
- (ii) by utilizing preparing information of substantial precipitation occasions as it were
- (iii) execution of forecasting for 1 hour to 6 hours and
- (iv) execution examination in top qualities and all qualities.

As per results in [2] ANN performed better when prepared with a dataset of heavy rainfall. For 1 to 4 hour beforehand, the past 2-hour input information was proposed for each of the three displaying strategies (ANN, SVM and ANFIS). ANFIS reflected better capacity in staying away from data commotion by utilizing various slacks of information sources. Lastly during top qualities, SVM ended up being more powerful under outrageous hurricane occasions. Analysts played out a relative examination of different information digging strategies for precipitation expectation in Malaysia, for example, Random Forest, Support Vector Machine, Naive Bayes, Neural Network, and Decision Tree. For this analysis, a dataset was obtained from different climate stations in Selangor, Malaysia. Before arrangement measure, Pre-handling undertakings were applied to manage the commotion and missing qualities in the dataset. Dataset spread over from 2002 to 2013 and comprised of 47 climate stations.

In [3], scientists anticipated the precipitation by utilizing the proposed Wavelet Neural Network Model (WNN), a joining of Wavelet Technique and Artificial Neural Network (ANN). To break down the exhibition, month to month precipitation expectation was performed with both the procedures (WNN and ANN) by utilizing a dataset of Darjeeling precipitation check station in India. Measurable strategies were utilized for execution assessment and as per results WNN performed better compared to ANN. Scientists gave a definite overview and played out a near examination of different neural organizations on precipitation determination. As per review RNN, FFNN, and TDNN are appropriate for precipitation expectation when contrasted with other factual and mathematical determining strategies. In addition TDNN, FFNN and slack FFNN performed well for yearly, month to month and week after week precipitation anticipation separately. This exploration likewise talked about the different proportions of exactness utilized by various analysts to assess the ANN's presentation.

In [4], the creator played out a study on different Neural Network models which were utilized for rainfall prediction in the most recent 25 years. The creators noted that the vast majority of the scientists got critical outcomes in rainfall prediction by utilizing Propagation Network, also the estimating methods which utilized SVM, MLP, BPN, RBFN, and SOM are more reasonable than other factual and mathematical strategies. A few constraints have likewise been featured.

The analysts in [5] thought about the prescient execution of most recent and best in class technique named "Markov chain extended with rainfall prediction" with the other generally utilized AI procedures: Support Vector Regression, Genetic Programming, M5 Rules, M5 Model trees, Radial Basis Neural Networks, and k Nearest Neighbors. Day by day rainfall datasets were gathered from 42 urban communities of two continents, with exceptionally assorted climatic provisions. 20 urban communities were chosen from around

Europe and 22 from around the USA. There were two reasons for picking two continents for information extraction, first is to perform the experiment on different climates having diverse weather and second was the geographical locations as the selected cities were exceptionally far separated from one another. A definitive objective was to not inclination the investigation to specific environment type or for specific geographic area. As indicated by results the amassing rainfall amounts can bring great outcomes as contrasted with forecast utilizing day by day stormy information. While utilizing the aggregated information, Support Vector Regression, Radial Basis Capacities, and Genetic Programming by and large performed well anyway Radial Basis Functions performed better than present day procedure of "Markov chain". For all chosen datasets, each procedure utilized similar boundaries so it was not ensured that the most ideal arrangement of boundaries was utilized for all the procedures. During the investigation, the scientists have noticed a connection between prescient precision and climatic qualities, for example, unpredictable nature of precipitation, measure of greatest precipitation and the interquartile scope of precipitation. Besides, no huge contrast was noted in calculations' expectation blunder among the urban areas of both the mainlands (USA & Europe). Issue in regards to the brokenness in precipitation information was addressed with the assistance of collected rainfall amounts.

3. RESULTS AND DISCUSSIONS

For this literature review, 5 research papers were considered and were studied thoroughly for literature analysis. This literature review did serve its purpose of answering some research questions that usually comes across every time someone tries to study about rainfall prediction using different data mining techniques or is fairly new into the field. Some of the said questions are mentioned below:

1] Which information mining strategies are utilized for rainfall prediction?

Authors in all selected papers introduced redone/coordinated/altered digging procedures for successful rainfall prediction. In each exploration, various climatic qualities/factors from past climate information were utilized as indicators with the end goal of expectation/determining. The extreme reason for each examination was to expand the accuracy rate of rainfall prediction.

2] How are the performance of prediction techniques assessed?

The selected papers have thought about the proposed procedure/model with recently distributed strategies. The execution was assessed by contrasting the anticipated

outcomes with the actual measures. Information retrieval metrics and statistical techniques were used for performance analysis of proposed techniques in comparison with other methods that were published before them.

3] Which sort of data is utilized for forecasting?

Every one of the selected papers utilized past climate information for rainfall prediction and for the training purpose used previously supervised data mining techniques. Un-supervised data mining strategies were likewise utilized in a mix of supervised techniques. Different climatic traits were utilized as indicators counting rainfall polarity, precipitation measure, least temperature, most extreme temperature, wind speed, and humidity and so forth As per specialists, utilizing more elements is not the assurance for more precision in forecast all things being equal insignificant qualities could influence the presentation. So the mix of important traits is required for exact precipitation forecasts besides these blends shifts upon case to case.

4] For which areas the Rainfall prediction is performed?

As indicated by shortlisted articles, rainfall was predicted in areas arranged in India, USA, Europe and Malaysia.

5] What are the factors affecting the results of rainfall prediction?

After the basic audit of shortlisted papers, it has been seen that following components could influence the precipitation forecast results: Past climate information: which is chosen for preparing the data mining calculation, climatic qualities: which are utilized as indicators, geographical location: for which the rainfall prediction must be performed, pre-processing techniques, general climate, surrounding environment and above all the previously used model/strategy/technique.

Limits of Research:

The limitations of this research are as follows:

1) The writing was extricated with a thorough and careful process which shows the quality and culmination of this review anyway some significant applicable work may have been missed.

2) Most of the incorporated and altered procedures were assessed by the authors themselves, so the genuine outcomes may not be as precise as clarified. This might influence the examination and results of this review.

3. CONCLUSION

Rainfall prediction is a valuable yet testing task. Data mining methods can foresee the precipitation by extricating and utilizing the stowed away information from past climate information. Somewhat recently, numerous scientists have attempted to expand the exactness of rainfall forecast by upgrading and incorporating information mining methods. Different models and strategies are accessible today for powerful precipitation expectation yet there was an absence of a conservative writing survey and methodical planning concentrating on which could mirror the current issues, proposed arrangements and the most recent patterns in this area. This exploration gave a far reaching deliberate planning just as the basic survey of the most recent examination from 2012 till 2013 in the space of precipitation expectation by zeroing in on information mining strategies. The exploration center around the space of precipitation forecast has been expanding since last decade as are the pain points. So it was reasoned that upgrades, enhancements and reconciliations of information mining strategies are essential to investigate and take care of these issues.

REFERENCES

- [1] S. Zhang, L. Lu, J. Yu, and H. Zhou, "Short-term water level prediction using different artificial intelligent models," in 2016 5th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2016, 2016
- [2] S. Zainudin, D. S. Jasim, and A. A. Bakar, "Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction," Int. J. Adv. Sci. Eng. Inf. Technol., vol. 6, no. 6, pp. 1148–1153, 2016
- [3] R. Venkata Ramana, B. Krishna, S. R. Kumar, and N. G. Pandey, "Monthly Rainfall Prediction Using Wavelet Neural Network Analysis," Water Resour. Manag., vol. 27, no. 10, pp. 3697–3711, 2013.
- [4] D. Nayak, A. Mahapatra, and P. Mishra, "A Survey on Rainfall Prediction using Artificial Neural Network," Int. J. Comput. ..., vol. 72, no. 16, pp. 32–40, 2013.
- [5] S. Cramer, M. Kampouridis, A. A. Freitas, and A. K. Alexandridis, "An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives," Expert Syst. Appl., vol. 85, pp. 169–181, 2017.
- [6] Aftab, Shabib & Ahmad, Munir & Hameed, Noureen & Bashir, Salman & Ali, Iftikhar & Nawaz, Zahid. (2018). Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review. International Journal of Advanced Computer Science and Applications. 9. 10.14569/IJACSA.2018.090518.