

# Rainfall Prediction using Apriori Algorithm

J.Refonaa, M. Lakshmi, Allu Chaya Satya kiran , Anantha Ravi Teja

**ABSTRACT---** *The growth of machine learning techniques has paved way for predicting various areas and weather prediction is one of the most commonly used application. One of the most interesting part of research is rainfall prediction Rainfall prediction is done by various researchers as it is one of the prime aspects in detecting the rainfall and measuring its intensity beforehand to reduce the number of loss of lives and properties. Nowadays, severe rainfall is one of the major reason is most of disasters such as floods and landslides. Predicting its occurrence well before could be very much helpful in further disaster management operations. In this paper, we have predicted rainfall by making use of various machine learning techniques and is been deployed by using it in predicting Chennai rainfall dataset. The research work makes use of Apriori Algorithm to demonstrate the feature set and to look out for the possibility of rainfall. Preprocessing of the dataset is done in order to extract the features of the dataset and then the Apriori Algorithm is used. The designed model is proposed to produce about 95% of accuracy, potency and lesser time when compared to other models.*

**Keywords—**Apriori Algorithm, Feature Extraction, Rainfall Dataset, Rainfall Prediction, Rule Set, Accuracy

## I. INTRODUCTION

Nowadays we have been seeing an intense rainfall and is considered to be one of the deadliest natural hazards that becomes the root cause for many other natural disasters such as floods and landslides. These natural hazards then tend to risk the lives of many people and also severe damages to the properties. Rescue operations at that time is a serious threat as most of the locations are not accessible by a human. Many damages occur due to lack of proper rescue operations. Rainfall prediction is one of the popular research areas that are done by many researchers using various technologies to predict the rainfall extent well before so that the people of that particular areas as could be well notified before and evacuation process can be done well before the rainfall actually occurs.

An immense role is played by the rainfall as it comes to damaging the natural life and other surrounding areas [18,19]. When it comes to agriculture rainfall is one of the most primary aspects but intense rainfall could also damage all the agricultural lands. Numerous research works are going to accurately predict the rainfall in a particular region [20]. Numerous studies say that the actual number of rainfall in a particular region purely depends on various weather-related parameters such as humidity, atmospheric pressure, and temperature.

**Revised Manuscript Received on July 10, 2019.**

**J.Refonaa**, Assistant Professor, School of Computing, Sathyabama Institute of Science and Technology. Chennai, T.N, India.

**M. Lakshmi**, Principal, Professor, Sri Krishna College of Technology, Coimbatore, T.N, India.

**Allu Chaya Satya kiran**, Student, School of Computing, Sathyabama Institute of Science and Technology. Chennai, T.N, India.

**Anantha Ravi Teja, India**

Climatic changes are also responsible for severe rainfall. Formation of cyclones in the sea beds also tends to severe rainfall at a particular region. Many models are designed that are dependent on mathematical calculations for the prediction purposes but tend to fail when it comes to accuracy at a given small geographic area. In order to eradicate this challenge we have proposed a model that makes use of Apriori Algorithm for predicting the rainfall in Chennai district. The model is designed in such a way that the features of the dataset is extracted and then is used to train the rule set used in the framework.

The rest of the section is as follows: Section II consists of Literature Survey, section III consists of the methodology used in the paper and section III consists of various results obtained. The paper is concluded in the last by mentioning the relevant future works that could be applied or added to the proposed work.

## II. RELATED WORK

Numerous researchers have been working on predicting the rainfall. Deepak Ranjan Nayak [1], has surveyed on various prediction models that have been used by previous researchers. The researchers tend to make use of the Artificial Neural Networks. The analysis on various earlier research works has discovered about FLANN being primarily used for rainfall prediction and provides more better accuracy. The information from the model is generated on an hourly basis in Asian countries [2]. The research was done on various other weather parameters that considered some of the parameters such as humidity, pressure, and also the wet bulb temperature. The cloudiness present in the sky was also used to predict the occurrence of rain before one to three hours ahead. The study has given a report that wet bulb temperature is considered as one amongst the key features that are used for such short term predictions that makes use of victimization feature choice technique. Akash D Dubey [3], researched on predicting the rainfall in Pondicherry district using Artificial Neural Networks. A literature survey has encountered that [4,5,6,7] Artificial Neural Networks provides a better efficiency when compared to other traditional classifiers for predicting the data. ANN is also used immensely for the prediction of various tasks as it provides good accuracy. In spite of Brodningnagian being a success of ANN is primarily based on modeling and predicting various weather parameters, and it also suffers from one of the biggest downsides that majorly concerns on the training various algorithms [8].

Some of the traditionally used algorithms are supported with the use of a gradient descent technique. These

algorithms try to realize the optimum weight values of a particular ANN by totally minimizing the slip operations that works as a calculator and also performs as same as an ANN. However, these algorithms have a partial improvement in the way the algorithms are structured and are unwell fitted to train the ANNs [9,10,11]. These algorithms don't guarantee optimum weights for the ANN. Besides, recent studies have discovered that the performance of ANN is best than ancient classifiers in many engineering and science applications [12]. Consequently, this article proposes Hybrid Neural primarily based on machine learning concept for rain prediction supported weather information consisting of options specifically vapor content, ratio, air pressure, and temperature. the information is gathered by dumdum bullet earth science center located at state [13,14,15,16,17].

III. PROPOSED SYSTEM

The proposed model made use of Apriori Algorithm to extract the features of the dataset that was gathered. In Fig. 1 we have depicted the block diagram of the entire proposed model. The model consists of a Chennai rainfall Dataset that is been used for the prediction of future rainfall in Chennai District. The dataset is used to select the features where the features are pre-processed using an array queue. The array is split and then is given for training. While training the dataset, Apriori Algorithm is been used for detecting the frequency of the rule set. A rule set is created such that whenever the rules are satisfied, there is a chance of rainfall in the region. According to the highest prediction values, the accuracy is calculated for splitting the future prediction of rainfall.

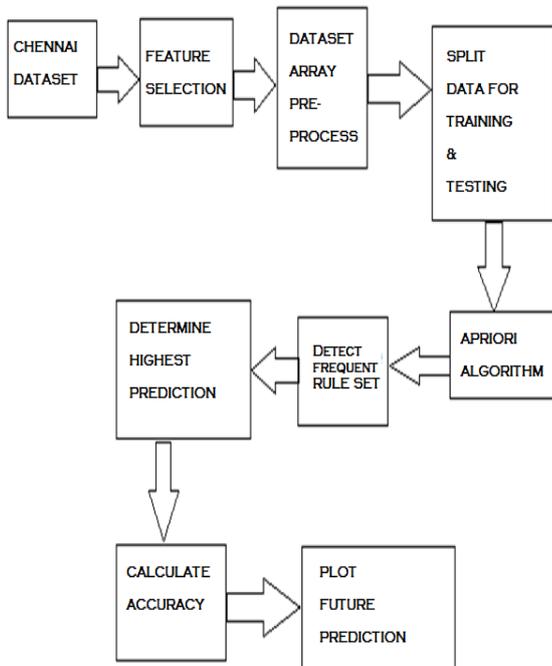


Fig. 1 Block Diagram of Proposed System

While making use of the Apriori Algorithm, the frequent item set is noted such that the temperature and the climatic changes are notes. Some of the climatic parameters that are used in predicting the rainfall is temperature, humidity and

rainfall. Once the information is gathered it is then used for the overall prediction. In Fig. 2, the architecture of the proposed framework is depicted. The diagram shows that as the features are extracted, it is then trained using the Apriori Algorithm in order to obtain the frequency of the highest item set obtained from the dataset.

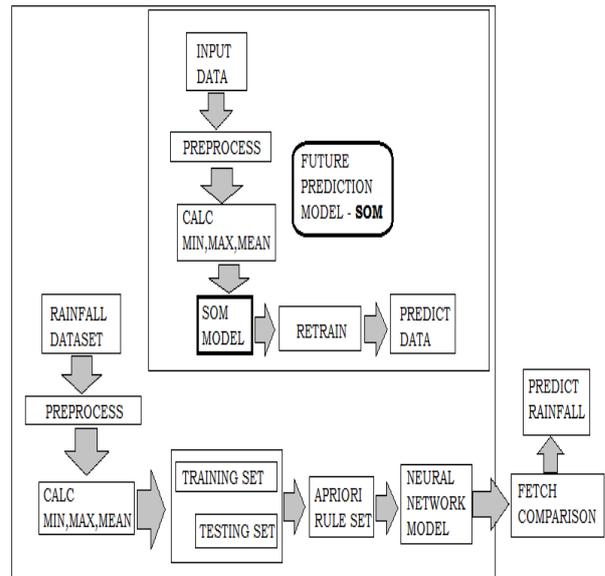


Fig. 2 Architecture of Proposed System

Apriori Algorithm

This algorithm is used for filtering the items from the database. It is used for seeing the accuracy values for the calculation. This algorithm will set the rules for our proposed system. There are three types of calculations we have, namely support, confidence, lift and conviction. Example, onion, milk, bread and butter is the item set i, all this item set is stored in database D. By using this algorithm, we can use the probability (i) to find the item from the selected database or area.

Support

The support of an itemset X,  $supp(X)$  is the proportion of transaction in the database in which the item X appears. It signifies the popularity of an itemset.

$$supp(X) = \frac{\text{Number of transaction in which X appears}}{\text{Total number of transactions}}$$

In the example above,  $supp(Onion) = \frac{4}{6} = 0.66667$ .

(i)

Here, number of transactions can be divided by the total transactions, for example, If onion is selected item from the database. Then Supp (onion) will be processed.

## Confidence

Confidence of a rule is defined as follows:

$$conf(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} \quad (ii)$$

For finding the accuracy value, for example, X denotes onion, butter and Y denotes burger. Then Supp. of X and Y will be union (Adding), then value of supp (X) is calculated using the previous mathematical expression. Here we can see the accurate value for the system

## Lift

The lift of a rule is defined as:

$$lift(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X) * supp(Y)} \quad (iii)$$

Rules setup, here X and Y will be multiplied for seeing the inaccurate values inside the item set. This kind of operations may apply to the particular area in the rainfall prediction area for seeing the inaccurate values. Using this expression, we can satisfy the rainfall prediction values from the datasets.

## Conviction

The conviction of a rule can be defined as:

$$conv(X \rightarrow Y) = \frac{1 - supp(Y)}{1 - conf(X \rightarrow Y)} \quad (iv)$$

This expression defines the difference of the particular value. For example, we have see the onion set difference in the database, then it shows the related data when comparing. Also this type of operations are using in the online shopping to see the difference and comparison for the product as well as for the values. Expression is used to see the conviction values for predicting the rainfall.

By satisfying all expression, we have used the apriori algorithm for predicting the rainfall values from the dataset. There are number of rainfall datasets, by filtering the values from the database is not comfortable to the developers. Some kind of research says for filtering the data, we can use the SQL engine but finding the random values for the rainfall prediction, we cannot satisfy the area as good. So we have used this algorithm for predicting the rainfall values as accurate. We have also satisfied the speed of prediction in this scheme.

## IV. EXPERIMENTAL RESULTS

The experimental analysis for the proposed model was performed efficiently using various parameters. The results were obtained using the MatLab software and are depicted in Fig. 3. The figure shows the number of times the

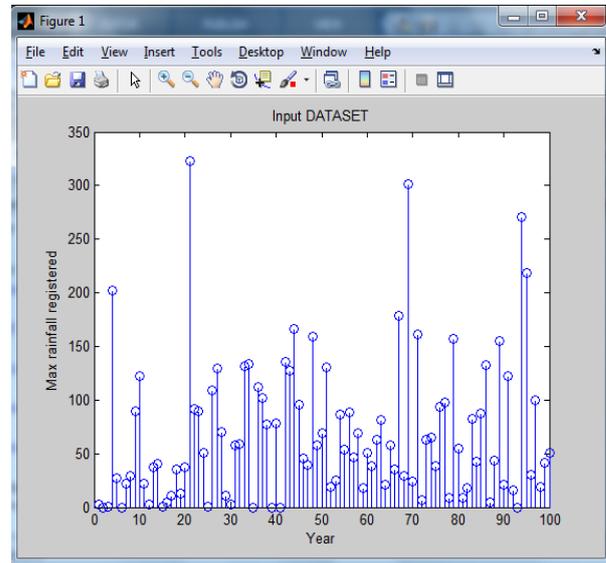


Fig. 3 Rainfall Detection in Input Dataset

Rainfall has been predicted in the last few years of the Chennai dataset. While observing at Fig. 4, it is noted that the rainfall of the Chennai district is predicted using the dataset and the Apriori Algorithm. The results shows the number of rainfalls that are likely to occur by mentioning the rainfall amount in millimeters.

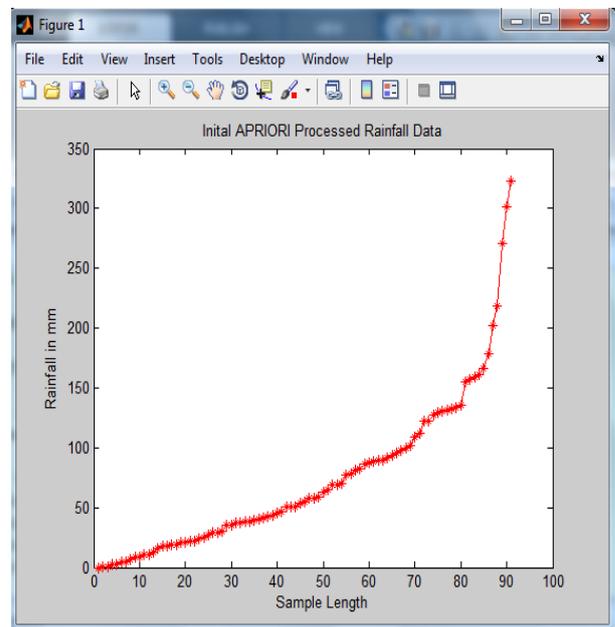


Fig. 4 Accuracy of the Proposed Model

The dataset used is directly trained Apriori Algorithm until the entire training is done. The experimental results are used to depict the feature that is chosen for training and is able to predict the rainfall with high accuracy. The comparative analysis of comparing the current classifier with other classifier has proved that training the classifier using the Apriori Algorithm provides a better performance when compared to other training models.

V. CONCLUSION

When it comes to loss of human lives by stuck in natural disaster, rainfall is considered as one of the most deadliest hazard that causes death to numerous lives. Rainfall also plays a vital role in causing numerous other disasters such as flood, landslides and many more that also generates an immense effect on the loss of human lives and property damage. Some of the applications such as irrigation and agriculture purely depend on the amount of rainfall it gets while irrigation. But heavy rainfall could make damage to the entire farming land leaving nothing for the farmers to cultivate. Rainfall prediction tends to predict the amount of rainfall that could be expected in a very short time in any particular area. In this paper, we have proposed a model for predicting the rainfall using Apriori Algorithm. It is designed in such a way that it accurately predicts the rainfall at any given region. In this paper we have used the previous Chennai dataset. The proposed model is compared with other existing models and tends to provide better accuracy when compared to other traditional algorithms.

REFERENCES

1. Nayak, D. R., Mahapatra, A., & Mishra, P. (2013). A survey on rainfall prediction using artificial neural network. *International Journal of Computer Applications*, 72(16).
2. Nayak, D. R., Mahapatra, A., & Mishra, P. (2013). A survey on rainfall prediction using an artificial neural network. *International Journal of Computer Applications*, 72(16).
3. Dubey, A. D. (2015). Artificial neural network models for rainfall prediction in Pondicherry. *International Journal of Computer Applications*, 120(3).
4. Dubey, A. D. (2015). Artificial neural network models for rainfall prediction in Pondicherry. *International Journal of Computer Applications*, 120(3).
5. Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A. S., & Balas, V. E. (2016). Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. *Neural Computing and Applications*, 1-12.
6. Chatterjee, S., Ghosh, S., Dawn, S., Hore, S., & Dey, N. (2016). Forest Type Classification: A Hybrid NN-GA Model-Based Approach. In *Information Systems Design and Intelligent Applications* (pp. 227-236). Springer India.
7. Knezevic, M., Cvetkovska, M., & Trombeva-Gavriloska, A. (2014). APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN CIVIL ENGINEERING. *Tehnickivjesnik/Technical Gazette*, 21(6).
8. Socha, K., & Blum, C. (2007). An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training. *Neural Computing and Applications*, 16(3), 235-247.
9. Dehuri S, Cho SB (2010) A hybrid genetic-based functional link artificial neural network with a statistical comparison of classifiers over multiple datasets. *Neural Comput Appl* 19(2):317–328
10. MacIntyre J (2013) Applications of neural computing in the twenty-first century and 21 years of neural computing and applications. *Neural Comput Appl* 23(3–4):657–665
11. Azar AT, El-Said SA, Balas VE, Olariu T (2013) Linguistic hedges fuzzy feature selection for differential diagnosis of Erythematous-Squamous diseases. *Soft Comput Appl AISC* 195:487–500
12. Awan SM, Aslam M, Khan ZA, Saeed H (2014) An efficient model based on artificial bee colony optimization algorithm with neural networks for electric load forecasting. *Neural Comput Appl* 25(7–8):1967–1978
13. Siddiquee MSA, Hossain MMA (2015) Development of a sequential artificial neural network for predicting river water levels based on Brahmaputra and Ganges water levels. *Neural Comput Appl* 26(8):1979–1990
14. Cao Z, Cheng L, Zhou C, Gu N, Wang X, Tan M (2015) Spiking neural network-based target tracking control for autonomous mobile robots. *Neural Comput Appl* 26(8):1839–1847
15. Gao S, Ning B, Dong H (2015) Adaptive neural control with intercepted adaptation for time-delay saturated nonlinear systems. *Neural Comput Appl* 26(8):1849–1857
16. Kamala, V. R., and L. Mary Gladence. "An optimal approach for social data analysis in Big Data." In 2015 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), pp. 0205-0208. IEEE, 2015.
17. Sanjanna, Y. Bevis J. (2015), "An Approach on Automated Rescue System with Intelligent Traffic Lights for Emergency Services", International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS 2015), pp. 1-5, IEEE
18. S. Ancy, Kumar. R, Asokan. R, Subhashini. R, "Prediction of Onset of South West Monsoon using Multiple Regression" On "Image Processing" at the "2nd IEEE- International Conference on Computer Communication and Systems" (ICCCS 2014) on at Saveetha Engineering College, Chennai on 20th & 21st February, 2014. ISBN No: 978-1-4799-3671-7
19. Saravanan, M., Nithiya, P., Recapitulation of coding to enable data integrity protection in cloud storage, *Global Journal of Pure Mathematics*, No7, Vol 15, 2015
20. Sethuraman, R., Chiranjeevi Saagar, K.B "Importing adaptive database schema design" *International Journal of Applied Engineering Research* Volume 10, Number 20 (2015) pp.18490-18493